

PROTAGONIST SCIENCE EXCLUSIVE

# TREACHEROUS ANCESTRY

AN EXTRAORDINARY HUNT FOR THE  
GHOSTS OF SARS-COV-2

# Table of Contents

<b>Prelude: The ongoing confusions</b>	4
<b>Chapter 1: Unraveling a mysterious chimera</b>	
I) An uncanny genome	6
II) On recombination versus artificial assembly	7
III) Recombination patterns are not random	10
IV) Of locks, keys, and the door to human infection	12
V) Defusing the myth of the furin cleavage site	14
VI) Putting the “gain of function” narrative to bed	19
<b>Chapter 2: Chasing the ghosts of SARS-CoV-2</b>	
Prelude	22
I) The cave(s) where it happens	24
II) Undersampled karst in the Wild Wild East	28
III) Ghost hunting with phylogenetic inference	32
IV) Reconstructing the phylogeographic histories	30
V) Smoking guns pointing at the wildlife industry	34
<b>Conclusion: Don’t bet against scientific inquiry</b>	39
<b>Bonus material and references</b>	
A word of caution	40
Motivation statement and copyright	41
References	42

# About Protagonist Science

The scientific method is one of the most important processes to understand in today's technological age.

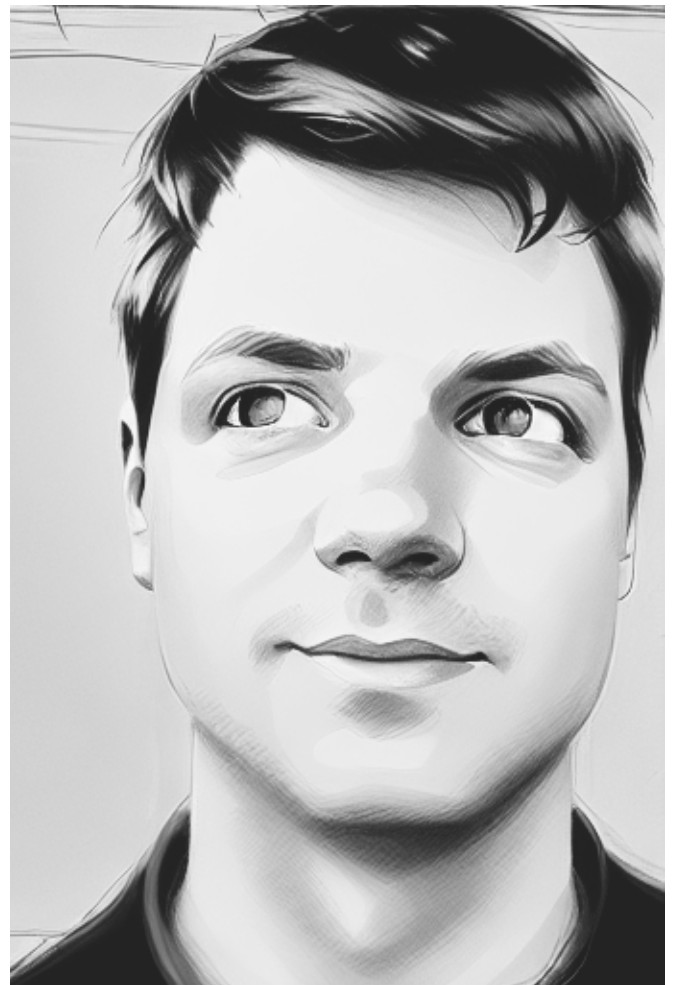
I believe that science is the hidden protagonist in any story that truly matters. Currently, *Protagonist Science* found its role as an outlet for independent science journalism and debunking of SARS-CoV-2 conspiracy myths.

In the end, this science communication endeavor is a long and eternally incomplete love letter about finding the human element in our technological future.

Sometimes written, sometimes spoken, sometimes visualized. Never, ever not worth your time

## About the author

Philipp Markolin is a Switzerland-based science communicator and writer. He holds a Bachelor degree in Chemistry and a Master degree in biochemistry from the Technical University of Graz, Austria and a doctorate degree in translational biomedicine from the Swiss Federal Institute of Technology. After a postdoctoral position as bioinformatic analyst in a machine learning lab, he decided to apply his hybrid background to understand viral phenomena in our modern information ecosystems, and how they shape discourse and public perceptions of science and society.



# Prelude: The ongoing Confusions

## 1) Virology under suspicion

Why are US scientists dragged in front of Congress in Benghazi-style show trials? How come highly rated research programs get quietly shut down? Is it justified to treat virologists with suspicion, consider them guilty of recklessness or dishonesty, even leak their subpoenaed private conversations to professional smear artists for cherry-picking?

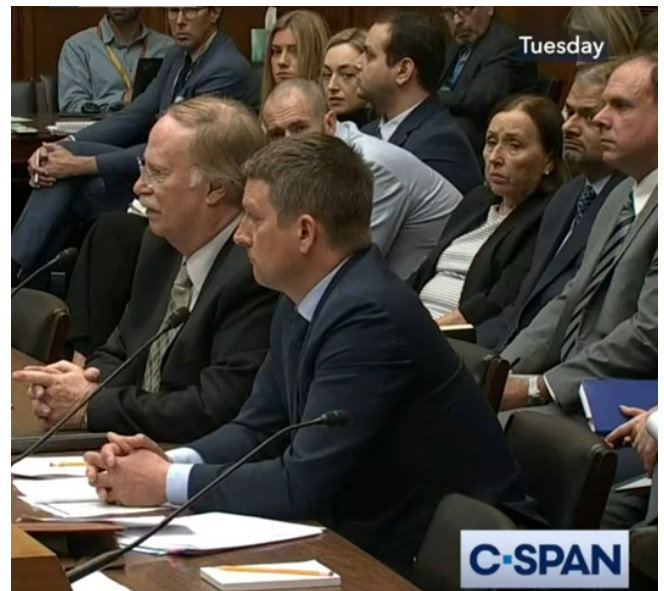
**The origin controversy is political and polarized.** Myths that COVID-19 was somehow a manmade pandemic are impactful, whether they are true or not. Polls have shown that 2 out of 3 US citizens believe that SARS-CoV-2, the virus that started the COVID-19 pandemic, came out of a laboratory rather than nature.

**Scientists worldwide vehemently disagree.** The emerging scientific consensus among domain experts is that SARS-CoV-2 is a natural virus that entered humanity via zoonotic spillover (more importantly, there is a consensus stemming from the body of evidence that is entirely unequivocal).

Yet fewer and fewer citizens seem to hear or care for their evidence-based voices. Domain experts are distrusted, but how much of that is self-inflicted, and how much is sown by anti-science actors? We can observe scientists are increasingly bullied out of public discourse by information combatants or targeted by harassers on social media. Most simply have their knowledge drowned out by an information ecosystem that favors emotional falsehoods over complicated facts. It is fair to say that science has come under pressure.

On top of that, various conspiracy myths about a "lab leak" continue to garner attention in the media, international forums, and the halls of US Congress, all while elected representatives, senators, anti-science activists, and propagandists seek to gain popularity, profit, or power by instigating a **"gain-of-function research" moral panic.**

*"We keep changing our minds about things based on data - that is science, not fraud." - R. Garry*



Dr. Kristian Andersen and Dr. Robert Garry called to testify on the origins of COVID-19 before the House Oversight Select Subcommittee on the Coronavirus Pandemic. [Image credit: C-SPAN]

### **On moral panics:**

*"As a distinct species of collective behavior, moral panics represent contentious and intensely affective campaigns to police the parameters of public knowledge and morality. [...] seeking to actuate alarm by influencing the imagery and representations of the mainstream press".*

*-Walsh JP, International Journal of Cultural Studies, 2020*

## II) A scientific question still needs a scientific answer

How many concerns about GoF research are justified, and how many just make for a good story? Media manipulators are skilled at fabricating false uncertainty and inventing sensationalist narratives out of whole cloth.

*When it comes to supporting evidence, the 'gain-of-function' emperor is embarrassingly naked. But it makes for a fantastic story.*



The Wuhan Institute of Virology is often center of speculations about "gain-of-function" research that supposedly created SARS-CoV-2. No evidence of any lab involvement has yet emerged.

All scientists can hope for is that eventually, citizens will get tired enough of the spectacle and make-belief to take a closer look at what we already know about the topic.

### **An asymmetric controversy**

*"It takes more effort to debunk a conspiracy theory than to create one. It took only one podcast to make me think the lab leak theory was possible, but it took months of research to understand why it's not"*

—Peter Miller, who recently won \$100,000 in an 18h origin debate against a lab leak truther

**Outside the heated spotlight**, multiple teams of international scientists have quietly honed in on a set of new clues to the origin of COVID-19.

Clues that might not only **hold the definitive answer to whether SARS-CoV-2 came from nature or the lab**, but also how to best anticipate and mitigate the risk of another SARS-related coronavirus pandemic. So where is the real birthplace of SARS-CoV-2? How exactly did this chimeric virus come about? Which trajectory did it take before the first outbreak emerged at the Huanan market in Wuhan? Will we ever find out?

Given current geopolitics, deliberate sabotage of scientific processes, and Beijing's obfuscation of the matter, **many say the true origins of COVID-19 will never be solved**. Some governments explicitly root for that outcome, preferring to be stuck in perpetual uncertainty forever. I would not take that bet against scientific inquiry — *and after reading this deep dive on the origin controversy* — I believe neither will you.

**Scientific literacy is a superpower** in the (dis)information age. It enables citizens to cut through the noise, intelligently engage with a topic of global controversy, and come out wiser. Scientifically literate citizens however need relevant evidence and the reasoning of scientists to be made accessible to form their opinion. On top of that, they will probably need months to familiarize themselves with the complexity of the origin discussion, and honestly, who has time for that? Having done the (sometimes painful) legwork, here is where I believe I can offer a helping hand. If you gift me your curiosity and time, this article will be a guide to understand some of the most compelling evidence for a natural origin of this virus.

*Tune out the politics, the broken media, and the bad actors. Let's return to the process of scientific inquiry.*

It is time to put the false "gain-of-function research" myth to bed once and for all, and learn about the **cutting-edge science on the hunt for the true origins** of SARS-CoV-2.

Buckle up, or fetch a large cup of tea, this is **the origin story told through the lens of viral recombination**.

# Chapter 1: Unraveling a mysterious chimera

## 1) An uncanny genome

When SARS-CoV-2 emerged, it was a confusing virus for a lot of reasons. First, it was very infectious to humans for a novel virus, spreading effectively between them via the respiratory route. Second, it did not cause severe disease in many patients, was sometimes asymptomatic, and subsequently hard to track. Third, side-by-side comparisons to known SARS-related viruses seemed to show that the novel virus was a chimera. It had parts of very high genetic similarity, and other parts of low genetic similarity, on top of a few other oddities that gave researchers initially a hard time to wrap their heads around. Even seasoned virologists were quoted wondering how *"this gets accomplished in nature"* in early February 2020.

As the pandemic turned into full swing, multiple "man-made" theories of varied quality were advanced on how the SARS-COV-2—and its odd genome—came to be. From bioweapon development to gain-of-function research, from de-novo genetic engineering to the alleged introduction of HIV sequences, from serial passage through human cells or humanized mice to arcane vaccine experiments, many asserted that some type of human manipulation was necessary to explain how this dangerous patchwork virus of high and low sequence similarities to other coronaviruses came about.

Only coronavirologists with decades of experience with that particular viral family would disagree, they certainly saw nothing unheard of in the genome. But the biggest problem in debunking the plethora of false notions was the lack of reference points; namely viral cousins of SARS-CoV-2. Those only gradually came trickling in once the severity of the outbreak turned into a global pandemic and jolted more and more scientists into urgent action. Rather than speculate on inconclusive data, many researchers set out to find related coronaviruses, either by discovering neglected genomes in large biomedical databases or by direct sampling of bats in nature.

They quickly realized that SARS-CoV-2-related viruses all looked a bit weird and stitched together; for example, a pangolin virus had an ACE2 receptor binding domain very closely related and able to bind and infect human cells. Another bat virus discovered in Mengla country in China had an insertion reminiscent of the furin cleavage site at the S1/S2 boundary, as well as what looked like an ancestral genome to SC2, at least for about the first 2/3rds of its full span.

How could these diverse animal viruses seem so closely related to SARS-CoV-2 in one part, and so distant in another?

*"Everything that seems counterintuitive with CoVs is actually logical when you work with them a few times",*

Jasnah Kholin  
(pseudonym)  
a coronavirus researcher  
from Hong Kong

## II) On recombination versus artificial assembly

Almost from day one, coronavirus veterans had been quick to educate their collaborators that CoV genomes **might just seem unintuitive** because they are shaped by a process called **recombination**.

Recombination is a mechanism for genetic exchange between two different parental viruses that creates a new viral genome sharing genetic information of both. It requires two (*usually distinct*) viruses to be present in the same host cell.

*As a useful but imprecise abstraction, one might think of recombination as a form of “virus sex” that produces unique offspring sharing a mix of parental genomic regions*

Offspring that came about by recombination is in many cases unproductive, meaning it can either not fulfill all essential functions necessary for the virus to replicate and infect new hosts; or it can do so but worse than the parental lineages, thus getting quickly outcompeted by them and vanish into nothingness. A virus needs to constantly spread and adapt to persist, after all.

However, sometimes this “virus sex” brings forth recombinant offspring that is in some aspect better—*meaning more fit in its current or new environment*—than the parental lineages, ergo it will spread and gain ground in the host population or particular environmental niche, possibly establishing itself for years to come.

**Recombination frequencies vary dramatically between viral families, from the promiscuous to the prudent.** For example, recombination does not seem to play an important role in some Filoviridae (Ebola, Zika) and Paramyxoviridae (Hendra, Nipah). At the same time, segmented viruses such as influenza use a different mechanism of genetic exchange altogether called “re-assortment”.

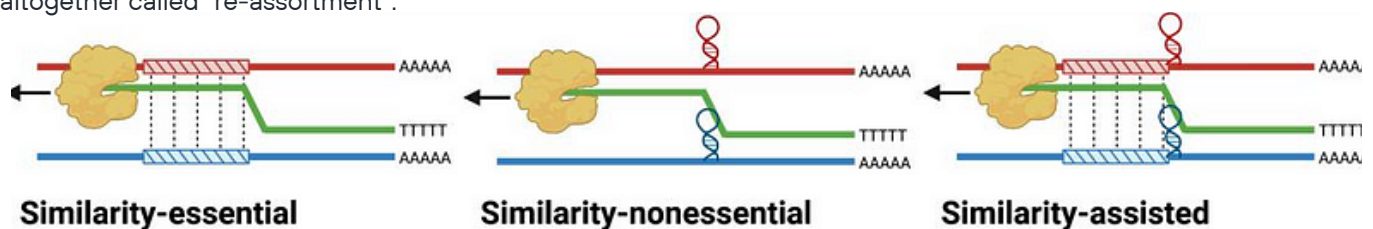
The coronavirus family of negative-strand RNA viruses reportedly recombines frequently. Recombination in CoVs is thought to be facilitated by a molecular mechanism called RNA-dependent RNA polymerase (RdRp) template switching leading to **copy-choice recombination (see figure below)**.

*“Replication of the coronavirus genome requires continuous RNA synthesis, whereas transcription is a discontinuous process unique among RNA viruses. Transcription includes a template switch during the synthesis of subgenomic negative-strand RNAs to add a copy of the leader sequence.”*

*—Sola I. et al., Annu Rev Virol., 2015*

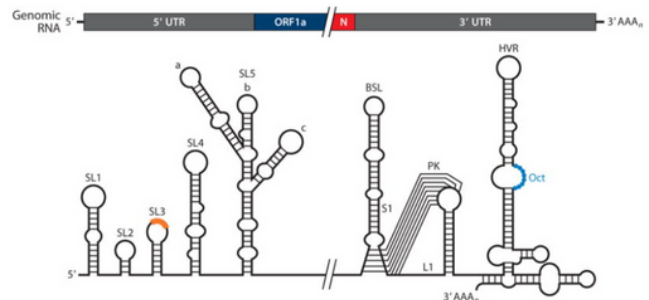
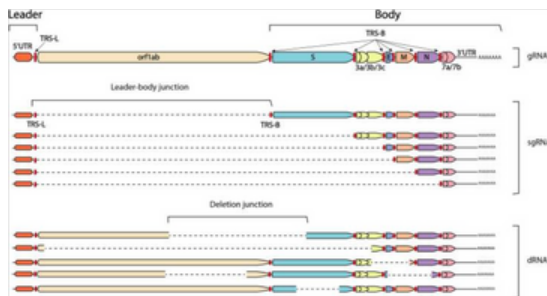
What this means is that the RNA production machinery (**RdRp**) can jump between instruction templates at any point if more than one is available, producing a hybrid sequence (chimeric genome).

I sketched out the mechanistic interplay between replication and viral gene transcription (*see next page*) to illustrate how template switching is not a freak event but an essential step in the viral life cycle.



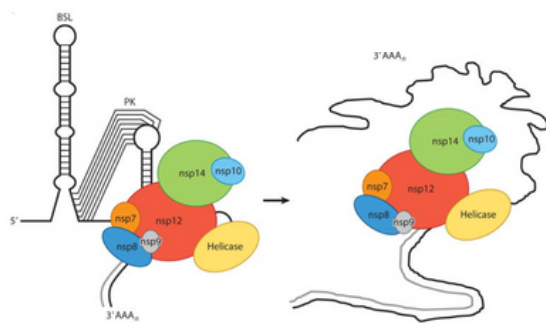
Three proposed molecular mechanisms for copy-choice recombination in CoVs utilizing RdRp template switching. (Wells H. et al., Cell Host Microbe, 2023)

## CoV's versatile RdRp usage facilitates template-switching and recombination susceptibility



1) Genome replication is a continuous linear process, but **transcription is discontinuous** (making the RdRp jump back-and-forth between elements)

2) RNA is not a straight line. Precise RdRp **jumping** (template switching) in transcription is **influenced by higher-order RNA structure**



3) Replication and transcription happen simultaneously and a single **complex 3D protein/RNA machinery** guides the RdRp meticulously between continuous synthesis and jumpy transcription

4) RdRp jumping can get messy (template switching during replication) when **similar RNA sequences from different viruses** get mixed together, creating a new recombinant CoV genome

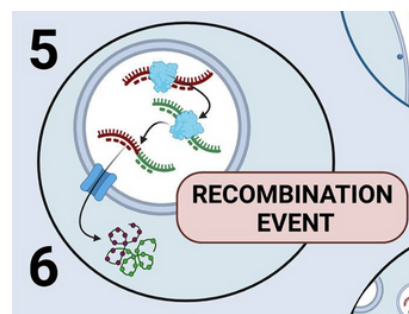


Illustration of recombination-prone genome replication in negative-stranded coronaviruses. When two RNA viruses find themselves in the same host cell, recombination is not an unusual event but common. (Individual figures from Sola I. et al., Annu Rev Virol., 2015 and Wells H. et al., Cell Host Microbe, 2023)

An RdRp that needs to jump back and forth between genetic elements constantly might easily jump to another different genome when thrown into the mix, thus providing ample mechanistic opportunity for recombination. (*Where CoVs get physical opportunity to have so much virus sex and what we can learn from it we will look at in Chapter 2*)

Again, thinking about this promiscuous mingling is merely a useful abstraction, because in reality, each individual virus genome is just one of many millions copies in a single infected person. Out of this complexity, recombination can result in ever new chimeric offspring that—if successful—will carry parental genetic segments forth from the generations that came before it. **However, in the early days with less information, SARS-CoV-2 did not seem very recombinant.** But maybe something changed with more bat cousins being found since?

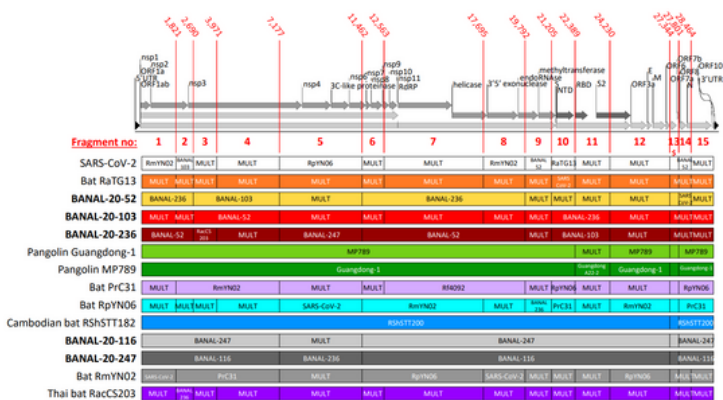
This is of course not a secret to experts in the field. **That coronaviruses fuck around** – sorry recombine constantly – has been **scientifically established** for many years before COVID-19.

*If SARS-CoV-2 and related sarbecoviruses indeed came about by frequent recombination, we would expect a colorful mix (a mosaic) of genome segments shared between them and their closest cousins.*

For now, what is important is that **the promiscuous sarbecovirus (SARS-related beta-coronavirus) family gets a lot of recombinant progeny**; and over time, established offspring lineages can themselves engage in more virus sex (= undergo recombination) with other circulating viruses. From long-estranged cousins and viral strangers to incestuous siblings and even their own offspring, anything seemingly goes (*hey, not our place to judge!*).

Let's take a look (*next page*).





Representation of the 15 recombinant fragments of relevant Sarbecovirus genomes compared to the SARS-CoV-2 human prototype strain. Where possible, the closest viral sequence is indicated for each fragment. In other cases, MULT indicates a group of multiple sequences. (Figure from: Temmam S. et al., Nature, 2022)

For SARS-CoV-2, the above figure shows that it shares regions of high sequence similarity with a bunch of different bat viruses, namely RmYN02, Banal-103, RpYN06, Banal-52, and RaTG13. Close cousins. However, for each SARS-CoV-2 to bat virus comparison individually, there might be some regions of very low sequence similarity (usually around the spike) where they are not similar at all.

The observation of this mosaic genome makes clear that **SARS-CoV-2's overall genome could not have been designed or derived from any known single progenitor virus** (for example RaTG13 or a BANAL-like virus, as lab leak activists like to argue), but rather that **it is a genetic chimera** containing bits and pieces from multiple related viruses. We know this technically because **no engineering or culturing approach could have magically created the "ancestral" sequence in hundreds of positions** any time a cousin diverged from SARS-COV-2.

Given this reality slowly breaking through, some lab leak activists (not prone to hold any coherent hypotheses on this issue anyway) have since moved on to fantasize that SARS-CoV-2 must then have been "stitched together" in a lab from a set of (multiple known and unknown) progenitor viruses secretly sampled by the Wuhan Institute of Virology. ("What can be asserted without evidence can also be discarded without evidence", I would say about the never-ending epicycles of conspiratorial ideation) In reality, all of the proposed **"stitched together" speculations** are of course **contradicted by evidence and epistemically risky at face value.**

These ideas also fail to account for the fact that all other SARS-CoV-2 relatives found in the wild look equally stitched together from multiple viruses.

### How come every random viral cousin found in the wild looks equally "stitched together"?

Recombination is the only scientific explanation that can account for the observation of these natural chimeras. (There are many more technical details as to why scientists know that SARS-CoV-2's mosaic genome was not assembled and stitched together in a lab from a sequence, but you get the main point)

On top of that, the observed recombination patterns left behind by **past "sexual" encounters between parental lineages can not be reproduced, simulated, or faked in a laboratory.** Pause for a second and read that again.

The relevant "sexual history" that played out over the last five decades or so needed to be lived in the wild, it **could only have happened in nature** where all these viruses meet.

All scientific evidence and prudence suggest therefore that **a naturally evolved, immediate bat ancestor to SARS-CoV-2 must have existed at one point.** It did not spring from a computer sequence, was not dreamt up by a mad scientist, or derived and recklessly assembled from disparate parts. SARS-CoV-2's bat ancestor came about as naturally as all the other recombinant children of parental sarbecoviruses that we have since discovered in bats.

### This is what scientists mean when they say the "backbone" of SARS-CoV-2 is natural.

On the larger genetic makeup, this ancestral bat version of SARS-CoV-2 already looked very much like the one that first surfaced in Wuhan. But this alone is not sufficient to exclude that such a recombinant bat ancestor was found by researchers and then "tinkered with" in the lab. **Recombination has a low resolution,** it can not track targeted single mutations. So *genetic tinkering remains possible and plausible, right?*

How else can we explain that a bat virus seems so damn good at infecting human cells?

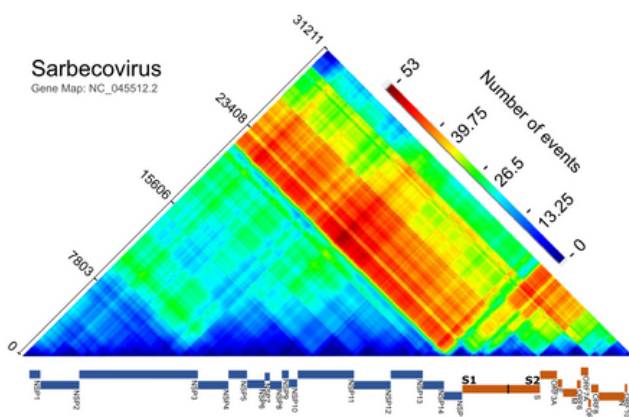
### III) Recombination patterns are not random

Another common line of erroneous argumentation by lab leak proponents concerns the idea that because SARS-CoV-2 was capable of infecting humans very well, it must have been somehow pre-adapted or optimized to do so via serial passaging in human cells, or been given this remarkable ability by thoughtful engineers. How else could a bat virus circulating in bats develop this human affinity just by chance?

For me, this argumentation from supposed "optimization" always has a bit of a fallacious character – "How can something marvelous such as the eye have come about by chance alone?" – seen many times in creationist arguments. **Evolution is not the same as chance.** We humans tend to underestimate the diversity of nature and the power of evolutionary selection at our peril.

*So how do we know that nature has optimized the human affinity of SARS-CoV-2 all by itself?*

I am glad you asked. CoV virologists have studied over the years how the S gene—encoding the spike protein—seems to be one of the biggest factors influencing cellular and species tropism (*the ability to successfully infect different hosts*). When scientists plot the frequency of recombination events along the Sarbecovirus genome, they found that there was not an even distribution, but rather "hot spots" and "cold spots" for genetic exchanges, with the spike region shining bright red.



Recombination region count matrices indicating genome regions that are most and least commonly transferred during detectable coronavirus recombination events (Klerk A. et al, Virus Evol., 2022)

What does this mean? Well, let us remember that recombination almost exclusively causes fucked up offspring genomes that are unable to procreate or establish themselves. **All scientists ever get to observe are the survivors**, those one-in-a-billion viral children who came out fitter than their parents.

*A recombination event that adds to niche fitness gets to stay, and whatever impairs fitness gets selected out.*

*[...] non-random and mostly conserved recombination patterns that we and others have detected in various coronavirus subgenera are likely shaped both by evolutionarily conserved variations in the mechanistic predispositions of different genome regions to recombination and by shared selective processes disfavoring the survival of recombinants that express improperly folded proteins."*

*—Klerk A. et al, Virus Evol., 2022*

The figure to the left basically shows that **within and surrounding the spike** and some accessory protein genes (orange), **observing new productive recombinant segments** (and with it, a higher sequence diversity) **is much more likely** than anywhere else along Sarbecoviruses' genomes.

So finding that SARS-CoV-2 has an "unusual" spike gene sequence compared to many of its relatives is actually not unusual at all. It is the norm among members sarbecovirus family, they all have unusually diverse sequences there.

The question is why?

In general, high sequence diversity enables viruses to stumble upon new functionalities, from evading immune systems to changing their tropism (what type of cells they can infect), facilitating more efficient cellular entry, or infecting new host organisms.

However, developing and maintaining sequence diversity over a viral population is not always easy to come by.

- Single point mutations arising from replication errors rarely get the job done, and face obstacles such as purifying selection.
- On top of that, RNA viruses with large genomes require a strong proofreading ability to correct replication errors that give rise to deleterious mutations.
- That is why for some RNA viruses, such as HIV, it is thought that sequence diversity more frequently arises through recombination than point mutation.

In coronaviruses, recombination is also assumed as an efficient mechanism for how these long, proofreading RNA viruses can rapidly create **genetic diversity from an existing pool of sequences**; presumably to cope with quickly changing niche environmental conditions. *(More on that in Chapter 2)*

*[...] recombination also has the potential to act as an evolutionary "fast-forward" by quickly shuffling genetic material between vastly different viruses.*

*For the same result to be produced by mutation alone, long spans of time would be needed for selective forces to shape such extensive nucleotide changes, especially considering the high proofreading capacity of coronaviruses.*

— Wells H. et al., Cell Host Microbe, 2023

## Gain-of-function experiments, but different?

When we think of general genetic diversity as a "potential for new functionalities" and recombination as a way to "shuffle varied genome segments" around, some alarm bells should be ringing.

With chimeric sarbecoviruses genomes, we are in essence **observing direct evidence** of countless potential **"gain-of-function" experiments** conducted on a scale hardly imaginable that must have happened in the past. *(More on that in Chapter 2)*

*The existence of untold numbers of such recombinant cousins in the wild implies that chimeric sarbecoviruses must have been birthed in some vast, natural "gain-of-function" laboratory.*

A laboratory where gain-of-function experiments with promiscuous parental lineages and distinct genetic functionalities relentlessly produce chimeric offspring that might have tricks from both parents.

Over time, the **constant combinatoric mixing** of varied elements and functionalities will **on occasion produce some very elaborate traits**, such as the ability to infect multiple different hosts (broad species tropism), including us humans.

I think it is worth looking at that in detail.

## IV) Of locks, keys, and the door to human infection

Broken down to a structural level, the major determinant of CoVs to infect various host species has to do with the makeup of the receptor-binding domain (RBD), a part of the 3D organization of the larger viral spike protein that has to fit the 3D scaffold of host receptor (ACE2 in humans) exposed on host cells. This is a bit of the lock-and-key principle.

- The higher the **tolerated genetic diversity of the RBM** (the genetic sequences coding for the RBD in the spike protein), the bigger the repertoire of potential keys available to CoVs.

Viral recombination in CoV spike genes is considered a major evolutionary mechanism that drives new adaptation processes, such as viral host switching. (If you already have a key factory within the family, and promiscuous virus sex where successful keys get passed around like a hot potato, all new chimeras need is the physical opportunity to try them on previously locked doors).

Here is an interesting little sidebar:

01 Side bar

•••

### COULD THE RBD HAVE BEEN ENGINEERED FOR HUMAN ACE2?

Engineering *de novo* protein-protein surface binding is extremely difficult and scientists have no reliable tools to design for these complex 3D interactions.

SARS-CoV-1 and SARS-CoV-2 both use the ACE2 host receptor for infecting humans; but they have found different "keys" to unlock that particular evolutionary door. If human engineers in 2019 were indeed keen on equipping a bat virus with this trait, why go through the Herculean task of reinventing the wheel, forging an hitherto unknown key, instead of just using the readily-available one nature gave them 16 years ago with SARS?

In any case, **we do not need to speculate about** whether it is possible or what potential ambitions genome engineers might have had or not.

Since 2021, researchers have not only found more cousins of SARS-CoV-2, but they also discovered that some of these cousins had keys very much identical to it, proving that at least for now, **nature is and remains the ultimate key master** for this particular set of keys, so to speak.

*Sarbecoviruses exhibit extensive genetic diversity in RBM, likely arising from frequent recombination and the high selective pressure associated with inter-species host jumping.*

—[Si J. et al., biorxiv, 2024](#)

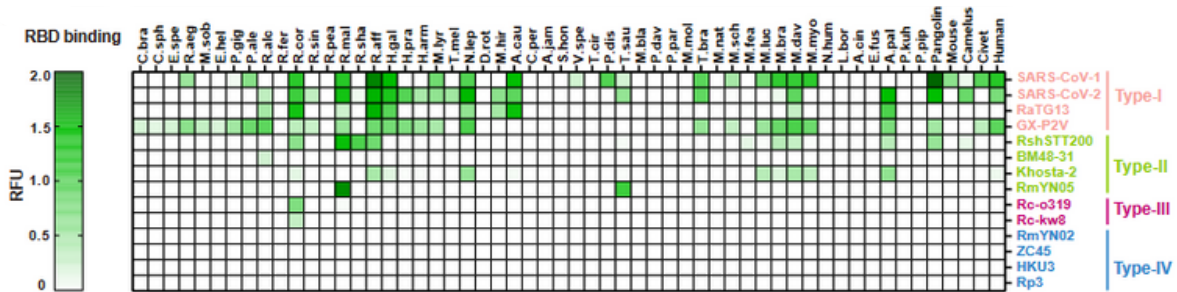
A very recent **illuminating preprint** from Chinese scientists— including **Zhengli Shi**, *the unjustly blamed "Batwoman" herself*— took a deep dive to understand how exactly RBM motifs (key shapes) relate to broad ACE2 tropism.

Experimentally, they produced 56 individual cell lines expressing ACE2 orthologues (locks) from bats and selected mammals and then tried 14 different RBDs (keys) from Sarbecoviruses.

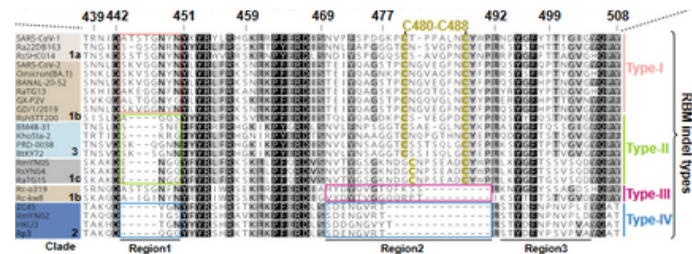
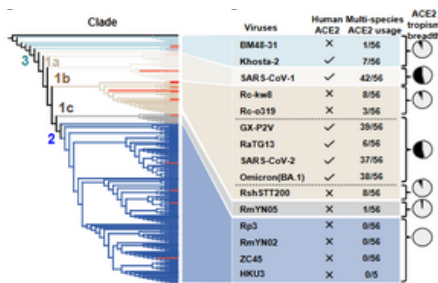
*What they discovered was that some keys could open almost all locks, whereas others could open only one, or even none of the locks presented.*

Much of this is **not unique, but based on specific deletions in specific positions of the RBM** (see *next page*).

## Many Sarbecoviruses show broad species tropism and can use hACE2 as a host receptor from the get go



1) Binding assay of 14 sarbecovirus RBD against cell cultures expressing ACE2 receptor protein of 56 species show broad binding affinity (incl. multiple bat species, but also pangolin, mouse, civet, camel and human)



2) The broader the binding affinity, the more likely that an RBD will also be able to infect human cells via hACE2

3) Sarbecoviruses RBM motives are highly diverse, with the full-length RBMs typically conferring broad species tropism and double region deletions resulting in loss of ACE2 entirely

Sarbecovirus RBMs exhibit a non-random pattern for narrowing or widening ACE2 affinity and species tropism. (Figures from Si J. et al., biorxiv, 2024)

*"We revealed that most sarbecoviruses with longer RBMs (type-I), present broad ACE2 tropism, whereas viruses with single deletions in Region 1 (type-II) or Region 2 (type-III) generally exhibit narrow ACE2 tropism, typically favoring their hosts' ACE2. Sarbecoviruses with double region deletions (type-IV) exhibit a complete loss of ACE2 usage"*

—Si J. et al., biorxiv, 2024.

The RBD in SARS-CoV-2 has had a very broad binding affinity from the start, which is why it is not only great at infecting humans, but minks, deer, house- and zoo animals as well.

*The supposedly "uncanny" ability of SARS-CoV-2 to infect human cells is neither uncanny nor especially unique. Some bat CoVs can just infect humans from the get-go. No pre-adaption, no magic human hand, no designer, no arcane lab experiments needed.*

On top of this, unique recombination patterns in the spike gene and the discovery of identical RBMs in nature make it unequivocally clear that SARS-CoV-2's RBD was not "designed", "created" or "swapped in from an unknown virus" by researchers in any kind of "gain-of-function" setup.

It acquired this receptor binding domain in nature.

## V) Defusing the myth of the Furin cleavage site

*But what about the furin cleavage site? Project DEFUSE?*

You might hear lab leak proponents lament. **Recombination does not have the granularity to rule out that this small sequence motif was artificially introduced.** The idea gained traction after a rejected research proposal from 2018 was played up by credulous amplifiers in the press.

In the “man-made” mythology, the furin cleavage site (FCS)—a polybasic cleavage site recognized by furin-like proteases—is a dramatic functional element and **allegedly the secret sauce and trigger** whose artificial insertion turned an ordinary bat virus into the pandemic blight pathogen we have today. For lab leak believers, the DEFUSE proposal mentioning the possibility of introducing FCS coupled with the **“suspicious” occurrence of an FCS** in SARS-CoV-2 is all but proof that the virus was engineered or tinkered with. In other words, researchers wanted to create a pandemic virus, or had reckless disregard if their work would do so.

Another quick sidebar:

02

Side bar

---

**PROJECT DEFUSE**

A 2018 research proposal from US scientists, together with the Wuhan Institute of Virology, included the suggestion of screening bat sarbecoviruses for potential polybasic cleavage sites and potentially introduce them into “appropriate high abundant, low risk parental strains” to “check growth potential in Vero cell and HAE cultures”

The proposal was rejected and the work never conducted. But even if this type of work would have been performed -- with a very generous interpretation of its scope -- it could not have created SARS-CoV-2 nor was this type of work a blueprint for such a virus.

Despite the DEFUSE proposal being irrelevant, I think it is worth discussing the **two associated insinuations based on scientific ignorance** a bit deeper.

**First, there is the common confusion** that the introduction of a single genetic element has the power to make a pandemic pathogen.

**Second, there is deliberate deception** about how likely it is that nature or engineering came up with the FCS insertion in SARS-CoV-2.

***The FCS is probably one of the most misunderstood elements in the history of the SARS-CoV-2 origin controversy.***

So let us first talk about why the Spike protein needs to be cleaved, where it needs to be cleaved, and how the FCS might help. (*The next points are going to be highly technical, so I made a visual summary to help on the next page*)

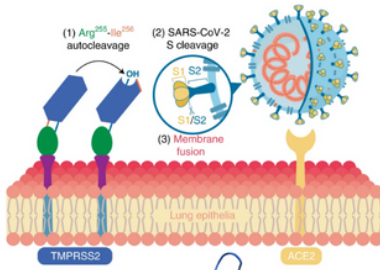
- The Spike protein in SARS-COV-2 needs to be cut during viral processing, first in the S1/S2 region, and second at the S2' site
- TMPRSS2 extracellular cleavage facilitates entry at or near the cell surface (cell fusion), as opposed to viral entry through the endosome pathway and late cleavage by cathepsins
- Extracellular cleavage allows the virus to avoid the potent endosomal/ endolysosomal restriction factors— the IFITM proteins— which inhibit viral membrane fusion and can stall virus replication
- TMPRSS2-expression is particularly high for cells in the upper respiratory tract, so viruses that can use this route more efficiently will have a selective advantage in respiratory transmission

### **Furin pre-cleavage aids TMPRSS2-mediated cell fusion**

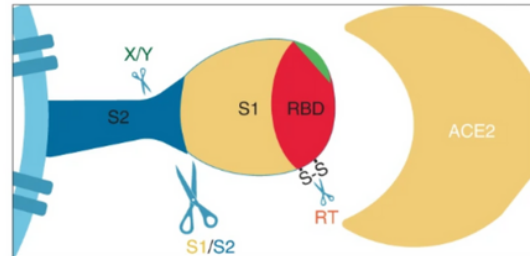
- The FCS motif allows spike proteins to be pre-cleaved before they leave their host cell; but also makes them less stable
- Spike protein egress that was pre-cleaved by the furin-like proteases in S1/S2 seem to take away some of the processing work of TMPRSS2, bind better to ACE2, and thus make viral entry via this membrane-fusion route more efficient for the virus

*continued on next page*

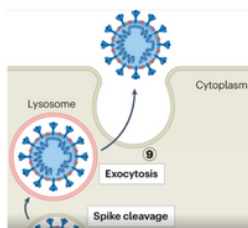
## Pre-cleavage of SARS-CoV-2 spike by furin-like proteases is an efficiency hack, but only in specific contexts



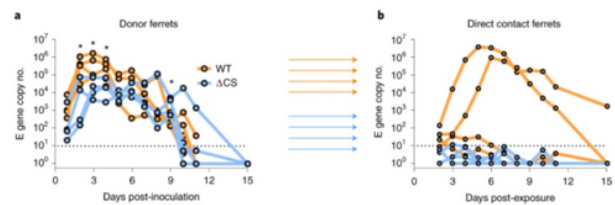
1) Efficient SARS-CoV-2 spike processing requires auxiliary host proteases such as the transmembrane serine protease TMPRSS2



2) TMPRSS2 needs to cut at S1/S2 and S2' to enable membrane fusion of the SARS-CoV-2 virion with the ACE2-carrying host cell



3) The **FCS motif is not required** for extracellular processing and membrane fusion. Furin-like proteases **pre-cleave spikes at S1/S2** before they leave infected host cells



4) Pre-cleavage of the spike during viral egress enhances entry of progeny virions into TMPRSS2-expressing cells such as those abundant in respiratory tissue

Cleavage at S1/S2 and S2' viral processivity studies highlight the function of the FCS and role in egress priming, respiratory tropism as well as transmission and infectivity. (Figures from: [Lavie M. et al., J Virol, 2022](#), [Jackson JB. et al., Nat Rev Mol Cell Biol, 2022](#), [Peacock T. et al., Nature Microbiology, 2022](#), [Steiner S. et al., Nature Reviews Microbiology, 2024](#))

continued from previous page:

- Studies have shown that TMPRSS2-mediated entry was particularly potent for virus particles that had FCS-containing spike compared with the non-furin-cleaved mutants
- In ferrets it was shown that virions with FCS-containing SC2 spike protein could spread to new hosts, but did not observe this for virions with FCS-mutated spikes
- In circulating lineages, the FCS cleavage is highly conserved
- Conservation of the FCS since pandemic start argues for its selective advantage in human-to-human transmission as well

**Therefore, we know that the FCS in SARS-CoV-2 is critically involved in transmission and pathogenicity.**

That an FCS acquisition can aid a respiratory virus is not exactly unobserved in nature either; for example, the transition from low pathogenic influenza strains towards high pathogenic versions often happens by the acquisition of a polybasic cleavage site which can be recognized by furin-like proteases.

**However, context is critical:**

- It was shown that even furin-deficient cells can still cleave SARS-CoV-2 spike protein at S1/S2
- The furin-cleavage site in SARS-CoV-2 is quickly lost by normal cell culture techniques because of slower kinetics, so it impedes the virus in this laboratory "serial passage" context
- The requirement of FCS and basic residues at S2' for S-mediated cell fusion is entirely cell type dependent
- an FCS does not make (or necessarily break) an pandemic virus, the whole viral-host context matters

## Necessary, but not sufficient

For a subset of highly pathogenic viruses infecting humans, an FCS might be a **necessary, but not sufficient element** to efficiently infect humans or sustain respiratory transmissions.

*However, merely adding an FCS can not magically turn any bat sarbecoviruses into a pandemic pathogen*

For example, MERS-CoV and SARS-CoV-1 **both caused epidemics, one with and one without an FCS**. In the case of SARS, some experiments have shown that the artificial addition of an FCS does not increase infectivity or pathogenicity. Additionally, SARS-CoV-1 without an FCS is no less infectious in ferrets compared to FCS-containing SARS-CoV-2. **Some of our endemic human CoVs have an FCS** (hCoV-HKU1) **but do not use ACE2** as an entry receptor, **others have no FCS but still use ACE2** (hCoV-NL63)

*I think we have to make peace with the idea that viruses are complex, highly optimized biological machines working in complex environments. Their molecular details, host context and environmental circumstances matter way more than single genetic elements.*

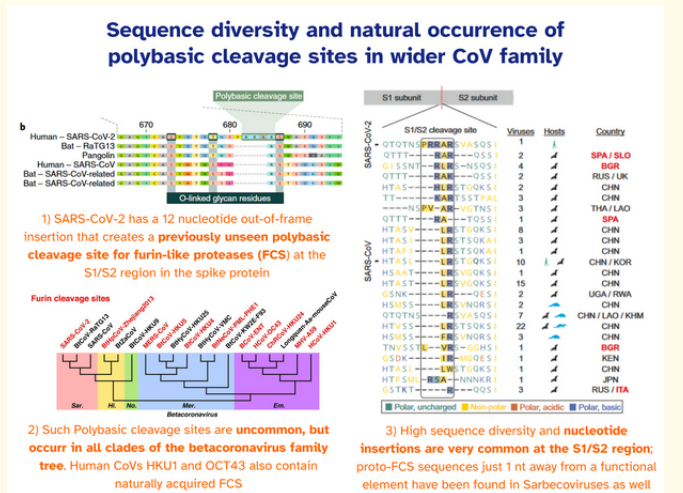
**In nature, no genetic element acts alone**

For a pandemic virus, the right genetic elements, host and environmental conditions have to fall together to create a perfect storm. We also would expect for any natural virus capable to start a pandemic to have some unusual genetic tricks up its sleeve. **Otherwise, every ordinary virus would cause a pandemic.**

## So why are many suspicious about SARS-CoV-2 having an FCS?

The main reason why the short FCS sequence in SARS-CoV-2 (12 nucleotide insertion) has been hotly discussed is **because it stood out like a sore thumb in side-by-side sequence comparisons** to known coronaviruses in early 2020. (Since then, scientists have found plenty of naturally occurring FCS in the wider CoV family, but not in the sarbecovirus clade specifically)

Just **because something is rare or even unique** among currently known sarbecovirus family members is however **not good evidence for artificial introduction given the high sequence diversity** at S1/S2 (see below). On top of that, there is some evidence to suggest the FCS region itself underwent recombination with a related sarbecovirus. Again, we underestimate the diversity of nature at our own peril.



Polybasic cleavage sites recognized by furin-like proteases are spread all over the wider beta-CoV family tree; including human CoVs HKU1 and OCT43. S1/S2 boundary region is highly sequence-diverse and susceptible to nucleotide insertions. These are the ingredients necessary to produce an FCS, irrespective of whether it is common or well-maintained in the bat sarbecovirus family specifically. (Figures from: Andersen KG., Nature Medicine, 2020, Wu Y. et al, Stem Cell Research, 2021, Sander AL. et al., Communications Biology, 2022)



## Is the FCS engineered?

Let's discard for a second our experience with vast and crafty nature—and the naïve presumption that an FCS can turn any virus into a pandemic threat—and take the idea of a “man-made” FCS introduction as a null hypothesis.

*Is there some evidence that could help us confirm or reject the engineering hypothesis?*

For an engineered FCS, the observed sequence motif in SARS-CoV-2 is odd, to say the least (see figure below for a visual summary)

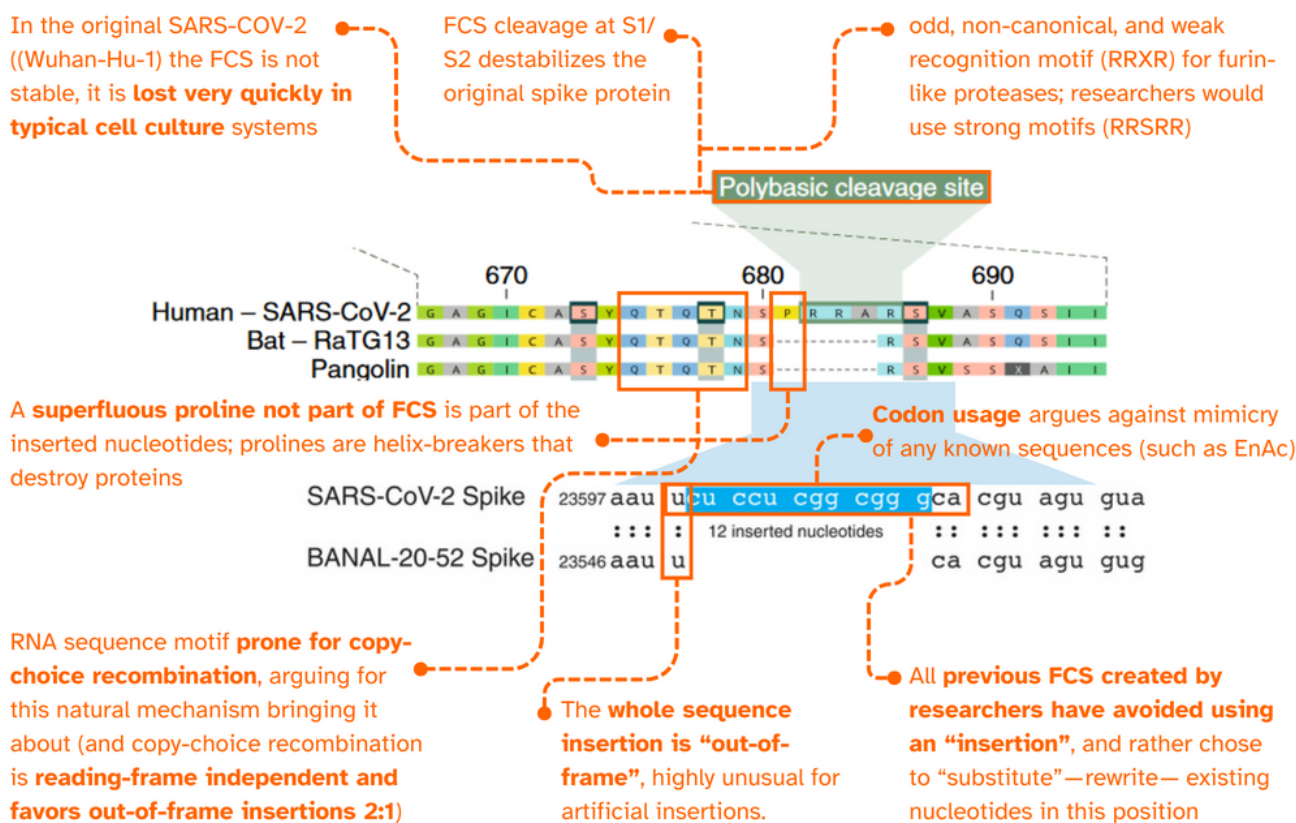
From an engineering perspective, **none of these odd, suboptimal, and self-defeating FCS design choices make any sense** and certainly contradict the assumption of “rational design” as a whole. They also serve to rule out engineering or mimicry of existing furin cleavage sites.

But that alone is not proof. Who knows what weird and arcane experiments those Chinese researchers would cook up in a lab, right? (*Some lab leak truthers with no molecular biology experience would say. For actual genetic engineers, the situation is already clear as day*)

Fortunately, we do not need to rely on experience or speculations about any motivations or experimental procedures Chinese scientists might have followed.

Because nature always has an ace up it's sleeve...

## The specific FCS sequence and introduction in SARS-CoV-2 argues strongly against being designed



Looking at the specifics of the observed FCS sequence motif, as well as how and where it is placed within the S1/S2 site, makes the hypothesis of an artificial insertion via genetic engineering extremely unlikely. (Figure based on Andersen KG., Nature Medicine, 2020, Garry R., PNAS, 2022)

## Hidden synergy

After intensive mechanistic studies of the FCS in SARS-CoV-2, researchers discovered a **hitherto unknown synergistic interplay** between the odd, suboptimal FCS and the genetic backbone of the virus (*which we by now know is natural because of recombination*).

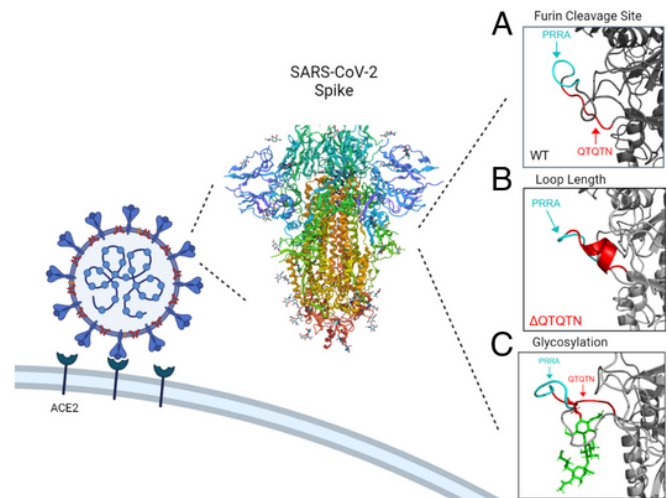
- We have learned that pre-cleavage of SARS-CoV-2 S1/S2 by furin is what boost TMPRSS2-mediated cell fusion and thus impacts cellular tropism and transmission dynamics (ergo infectivity)

*But it turns out, that was not the full mechanistic story of the FCS. In nature, no genetic element acts alone.*

- The viral **QTQTN amino acid motif** is an uncommon natural sequence in several sarbecovirus spike proteins directly upstream of the FCS sequence
- In cell culture experiments, QTQTN is not stable and repeatedly lost, similar to the FCS. Loss results in impaired viral replication.
- QTQTN determines how tightly the loop harboring the FCS is bound to the spike protein, thus it regulates spike protein stability and shapes how well the Golgi-bound furin-like proteases have access to recognize the FCS motif and pre-cleave at S1/S2
- The QTQTN motif is also glycosylated and loss of that glycosylation impairs viral replication as well.

Researchers showed beautifully (see figure top left) that there is an intricate functional interplay between the FCS, the loop length, and glycosylation: These separate but co-dependent elements work together synergistically and all elements are ultimately required for efficient FCS pre-cleavage (and with it, impact viral replication and pathogenesis).

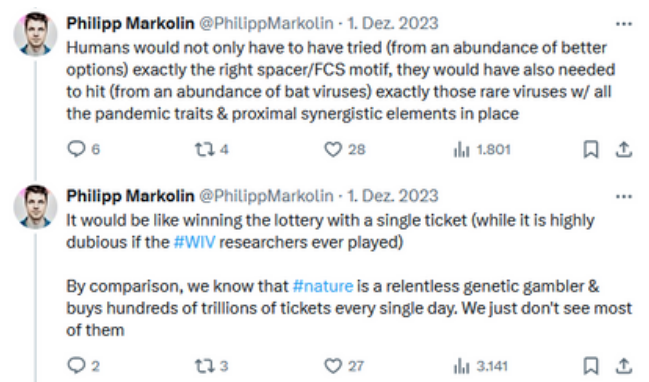
Co-dependency and synergy between genetic **elements are hallmarks of evolutionary history and selection.**



Synergistic interaction between the FCS, proximal sequence motif (QTQTN) and glycosylation (Figure: [Vu MN. et al., PNAS, 2022](#))

These type of complex interactions are extremely hard to design for even with perfect knowledge of all structural components, which nobody had in 2019. In the case of SARS-CoV-2's FCS, there is no conceivable way how engineers could have designed the observed – but hitherto undiscovered – mechanistic synergy purposefully.

Some might still contest that engineers could have stumbled upon this interaction by sheer “lucky accident”, like winning the lottery. I'd say this:



To just give you the odds of a “lucky engineering accident” scenario, scientists would have to have tried an unholy combination of FCS sequence/random spacer motifs against the backdrop of total viral diversity with the right proximal sequence elements available from rare viruses nobody previously described but that are capable of causing a pandemic in the first place.

Even a special pleading for a “miracle accident” evaporates in comparison to what we know about nature's relentless efforts and trillionic opportunities to bring forth such elements.

## We underestimate the viruses nature can come up with

Mechanistic details and probabilistic observations are key to wrapping our heads around the FCS in SARS-CoV-2. If we understand that:

- *an FCS alone can not turn a random bat virus into a weapon of mass disruption*
- *the FCS has many hallmarks speaking against design or engineering but strongly arguing in favor of evolutionary mechanisms*
- *a previous unknown synergistic interaction could not have come about by engineering or stumbled upon by chance with any known laboratory experimental setup*

...only then can we fully appreciate the scope and intricacies of gain-of-function experiments nature runs every day.

They are of a sophistication and breadth human engineers yet can not hope to match, nor fully fathom.

***“I always say that making a recombinant virus is easy. But to make a virus like SARS-CoV-2, before nature came up with it? Impossible.”***

*Prof. Linfa Wang, Duke-NUS  
Singapore*

## Bonus: Zoonoses - not research how to stop them - is what creates biosecurity threats

The roles of *virus discovery* or *gain-of-function investigations* for pandemic prevention are often deliberately misrepresented as cautionary tales against dual-use research for bioweapons. The contextual and mechanistic intricacies of the FCS should expose the naivete of such shallow arguments.

**In nature, no genetic element acts alone** and viruses are freaking complex molecular machines that we humans have no idea how to bend to our will. While naive suppositions about the FCS get all the undue spotlight, there are other single mutations how **nature “weaponized” SARS-CoV-2 in ways no engineer could have ever figured out**. Take the role of amino acid 37 in the nsp6 protein. In bats, the “ancestral” version has a valine at this position (V37), whereas SARS-CoV-2 has a mixture of L37 and F37 that seems to play a critical role in asymptomatic spread. We do not even have a concept of how one would ever figure this out starting with a bat virus, nor experimentally test for it in any research setting.

**Weaponization is generally much harder than people realize.** Even extremely well funded programs in the past were not successful for multiple complex reasons beyond technical capabilities. For example, receptor binding is not the sole determinant of species tropism, and for a lot of viruses, it has nothing to do with pathogenicity. Additionally, viral fitness in a permissive host cell or model says nothing about in vivo fitness or transmissibility. Weaponizing transmissibility or pathogenicity would need to be tested in human cohorts, not animal models.

In my opinion, it is not *virus discovery* or *gain-of-function research* that create an **“information hazard”** by *“increasing the stockpile of select agents and threat of biowarfare”*. Naive nonsense.

It is **letting zoonotic spillovers happen where nature figures out all the tough parts of viral engineering**. This collective neglect forces the world to deal with ever-new and possibly effective organisms added to the select agents list. A novel pathogen that bad actors can then attempt to deploy by simply modifying key residues to avoid prior immunity.

Activists, spooks and politicians concerned with bioweapons, biosecurity and national security should probably worry about nature’s bioweapon R&D program a lot more than they currently do.

## VI) Putting the gain-of-function narrative to bed

Speculations that the “odd and unusual” genetic features of SARS-CoV-2 came about by gain-of-function research, genetic engineering, design, or any other type of human tinkering—*while hard to disprove in 2020*—are today contradicted by available evidence and published scientific literature. Science did not stand still in the last four years, researchers learned an incredible amount about the intricacies of the virus that disrupted the world.

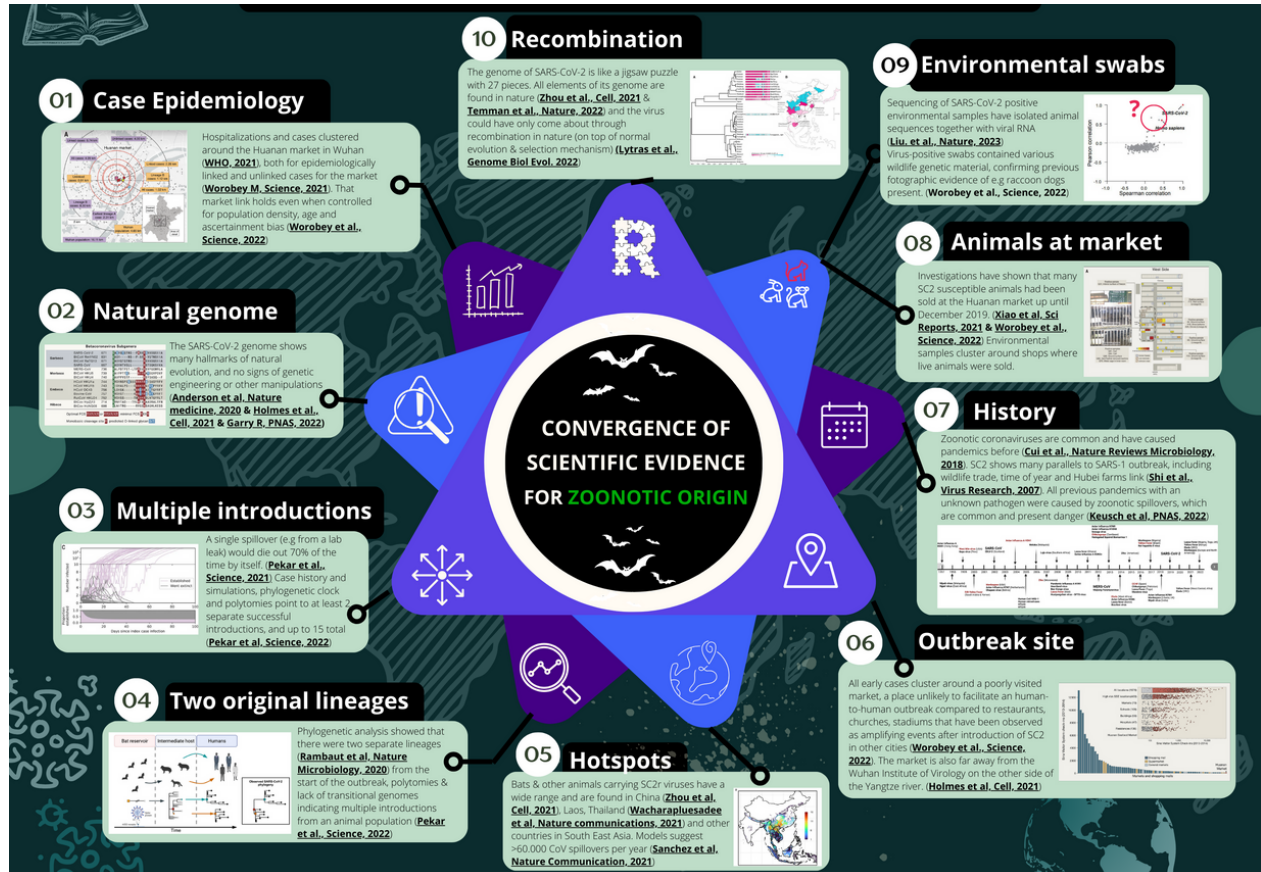
Taking together various recombination analyses, sequence diversity in nature, and hallmarks of evolution in the recombinant backbone, RBM and FCS leave only one parsimonious conclusion:

*SARS-CoV-2 is a fully natural virus, its specific genomic features could not have come about by any known or unknown laboratory experiment.*

**Note: A natural virus by itself does not disprove a lab leak**, nor all conceivable research-related origin scenarios. But it certainly disproves all notions of “gain-of-function” research causing the pandemic.

**Science works with evidence**, so it is often impossible to prove a negative—that something did not happen—scientifically, as there would be no evidence of something “not occurring”. Conspiracy myths are adaptive and built to sustain the belief that the evidence-based “mainstream” explanation is wrong. This often makes their scientific refutations difficult especially when the goalposts of how an event supposedly happened are constantly shifting, including central elements such as which lab is supposed to be the culprit can shift at a moment’s notice.

All I am cautioning about is that such cover-up **conspiracy myths filled with extraordinary and magical assumptions are epistemologically irresponsible**, contradicted by multiple other strong lines of scientific evidence, and can not even explain most of the body of evidence we observe. (see below)



There is a large and detailed body of evidence that any scientific hypothesis needs to be able to explain. The zoonotic spillover hypothesis is not only consistent with but powerfully explains every piece of evidence we have. No such consistent hypothesis currently exists for the speculations surrounding a research-related accident. (Figure source: here)

## Isn't stopping gain-of-function research causing pandemics a worthy cause?

*"Who even cares about the true origin of this virus? Stopping gain-of-function research causing pandemics is still a worthy cause, and if it took a little mythmaking to get the ball rolling, fine with me"*

You might have seen versions of this "greater good" justification for keeping the false "gain-of-function lab accident" narrative alive. I certainly have.

**There is some logic behind it.** Even if this pandemic virus was not created by researchers, future ones conceivably might be. The reality is that gain-of-function research has potential risks and thus needs to be regulated, possibly more tightly than it currently is. Biosafety, biosecurity, and dual-use research in labs also need societal oversight (& *should never have been self-regulated by reckless scientific panels or agencies, some gleefully suggest*). For many biosafety experts and existential risk think tankers, the story that SARS-CoV-2 was created by gain-of-function research finally shocked the world into action on a long-overdue publicly neglected topic, so overall it served a greater good irrespective of veracity. Fair enough.

**Just two little caveats:** First, the topic was not neglected and regulations were in place. Second, nobody of the dozens of virologists I talked to is or has ever been opposed to regulation and oversight. Without exception, virologists welcome and are supportive of efforts aimed at constantly improving existing frameworks. This is because—*in case you have not thought this through*—their bodies are first-in-line for deadly lab escapes. Clear, actionable, and effective regulations are welcome and desired from all sides here, and are subjected to ongoing discussions about scope, practicality, enforceability, financing, and a million other nitty-gritty details between different domains, stakeholders, and the public.

**So where is the problem?** The false "gain-of-function lab accident" myth has created a destructive asymmetry in these discussions, with existential risk types and biosafety fearmongers getting to have their cake and eat it, subverting the need for any scientific discussion, negotiation, or compromise between different possible implementations and future visions.

Currently, biosafety advocates (*and activists who pretend to be*) use the false narrative to paint all virologists as hopelessly plagued by conflicts of interest; that they can not be trusted to have any say or input in regulations shaping their everyday work. Some are true believers, others do it for self-serving or petty reasons. Some hope to gain political sway or direct more financial resources to their research departments and agendas. Some even demand a total ban on gain-of-function research on viruses. However, these **dramatic societal decisions should not happen under the exclusion of virologists** and domain experts **who know the most about the topic**. That is not just bad policy, but recipe for disaster.

For example, knee-jerk regulation currently being considered has already cast a chill over virological research even before being implemented. Sweeping bans on gain-of-function research come with unbearable risks, or more provocatively stated, a predictable death toll. Lives not saved because medical interventions (e.g. oncolytic virotherapy, vaccines, gene therapy) were slowed down, research was abandoned, or made inaccessible are still unnecessary deaths; and that is not even considering the much contested utility of GoF for pandemic prevention.

*"Even without implementation, the NSABB recommendations are already having a chilling effect on virology research. The uncertainty generated by these proposed rules is causing scientists and their institutions to avoid research that is likely to be affected, including the development of vaccines and therapeutics."*

Rasmussen AL. et al., Journal of Virology, 2024

## Why false myths are dangerous

**No matter where you stand on the gain-of-function discussion in general**, I believe it is fair to say that the continued push for a “gain-of-function lab accident” narrative in politics and media is not driven by evidence, not justified by remaining scientific uncertainties, and can be safely discarded.

We have seen again and again that no amount of hard evidence for any scientific theory has the power to prevent certain activists, politicians, influencers, and other manipulators from trying to sell citizens a different—and usually emotionally more satisfying or engaging—story. Climate change denial, anti-vaccine activism, homeopathy, various alleged alternative medicines or wonderdrugs; none have any scientific leg to stand on, but are **profitable narratives to use for popularity, persuasion, or power**. Those who have something to gain from pushing false myths will continue to mislead no matter the evidence. We just do not have to buy into their debunked falsehoods.

*The constant performative pearl-clutching and fearmongering about gain-of-function research or lab biosafety in the media should not distract citizens from the factual reality that SARS-CoV-2 is a natural virus.*

My larger point is that biosafety is an important, complex, and omnipresent topic that virologists take seriously. But do its most fervent advocates in politics, media, and other halls of power?

If you observe these activists, politicians, contrarians, or influencers who claim gain-of-function research — or any type of human genetic engineering technology — created the virus, **you'd be wise to question not only their motifs but their leadership.** (top right)

Unfortunately, getting the origin of SARS-COV-2 right is not just an academic exercise, or avoiding being fooled. It is about the actions we collectively fail to take when falsehoods dominate the discourse.

### **A dangerous blindspot**

*The relentless mythmaking about a “gain-of-function” virus and alleged lab biosafety breaches have exposed a dangerous blind spot in its most fervent purveyors:*

*The inability to appreciate the vastness of viral diversity and the “gain-of-function” ingenuity of nature, both of which continue to pose a predictable and outsized threat to human health and prosperity.*

**False narratives and beliefs in a “gain-of-function lab accident” are not without consequences** for society, and neither is scientific illiteracy or dishonesty in elected representatives and social elites. Are we truly okay with too many leaders chasing false myths while sleepwalking into the next natural pandemic they could not be bothered to understand? Whatever your risk-to-benefit perception about gain-of-function research in labs, ignorance about the scope and predictable danger of gain-of-function dynamics in the wild is a surefire recipe for disaster.

*I hope citizens can at least agree that if you care about “gain-of-function viruses”, you should care about them anywhere in the world.*

**Any biosafety advocacy that magically stops at the lab door** — that does not consider nor care for the quantifiably orders-of-magnitude larger natural risks — **is not a good basis for collective decision-making** in an interconnected world. The threat we neglect will be the one we will keep facing.

So let me help focus our attention and understanding of **where the true origins of this chimeric gain-of-function virus** — and its future unwelcome pandemic cousins — actually are to be found.

# Prelude: A river in the sky

**A river of black drew its line in the darkening sky.** Above bat ornaments on crimson roofs from the pagodas at Wat Khao Chong Pran, the river flow turned Southeast tonight, towards the crop fields.

*Rationally, I knew that the cave housed around two and a half million horseshoe bats, but observing a seemingly never-ending flood of hectic creatures – the fly-out of 2,5 million bats took more than forty-five minutes – I realized that I never truly appreciated how many bats share the world with us.*

There are around 1500 described bat species that have emerged from their last common ancestor over 60 million years ago. Because they are the only flying mammals, we conceptualize them all together under an umbrella term called “bats” like we do with “fish” in the sea. **But based on genetic diversity, that simplification is rarely adequate.** It feels equivalent to lumping together giraffes and cows with dolphins and whales, all of which diverged from a shared common Artiodactyla – even-toed ungulate – about fifty million years ago. Hardly justified we think of them as one and the same, so why do we do it for bats? It is no exaggeration to claim that bats come in almost all sizes, shapes, and forms; from the thumb-sized *Craseonycteris thonglongyai* – also known as bumblebee bat weighing just 1,5 grams and holding the title of smallest mammal on earth – to various majestic flying foxes with wing spans of over 6 feet, close to two meters. Some bats look almost like dog puppies you’d want to cuddle and take home, others might appear like they’ve escaped from a horror-movie production.

Maybe we humans just tend to be afraid of what we do not understand.

**The horseshoe bats flying here were small insectivores**—insect-eating bats—who got their name from the weird horseshoe-shaped disfigurement where their nose should be. Intuitively, we humans find them rather ugly – I was no exception – at least at first. I think this is partly because we assume their faces are weirdly deformed, like a fully cleft palate, rather than what they actually are: optimized.

Horseshoe bats belong to a group of bats that echolocate—sending out and receiving sonar waves—primarily through their nostrils. Other bats mostly use their mouths. These varied shapes and forms in the middle of their faces however have intricate functionality for shaping their calls, impacting not only orientation but for their feeding and social lives too. Given their enormous diversity, bats exist in almost all variations of social structures, from eremites that don’t want to bother with others, to small family groups, to villages, to multicultural megacities. Some like to mingle with other bat species, others are territorial and of the “get off my lawn” persuasion, with threatening grunts and fletching teeth and all. Some horseshoe bat species are not only considered hypersocial cosmopolitans, they are also the most prominent **host reservoir to SARS-related bat coronaviruses**; close viral cousins of both SARS-COV-1 and SARS-CoV-2 that have since caused havoc in our human world.

**I believe to truly understand where these dangerous viruses come from, we have to first take a look into the lively homes of their hosts.**

# Chapter 2: Chasing the ghosts of SARS-CoV-2

## I) The cave(s) where it happens

*"No one else was in the room where it happened. The room where it happened. The room where it happened. No one really knows how the game is played, the art of the trade, how the sausage gets made. We just assume that it happens, but no one else is in the room where it happens"*

- Hamilton (Musical), the Room Where It Happens.

**Physiologically, bats are extraordinary.** They can speed up their metabolism 16 times, creating immense heat that would denature our proteins and fry our cells. A bat's heart can beat up to 1,000 beats per minute and its body temperature can rise to 42 degrees Celsius [107.6 F] during flight at night. Even if we could withstand the initial stress of flight, our inflammatory response after would render us sick. That is not true for bats. Many species have unique immune systems that do not overreact to stressors; including viral infections. There are also no reliable markers of aging after they reach adulthood, bats seem young until they are dead. Some species tend to live up to 40 years in vast, dense, and diverse colonies. All of the above seem to make bats uniquely suited as reservoirs to almost all viral families that befall mammals. But for sarbecoviruses specifically, that is not the whole story.

*Virus outbreaks are always social phenomena as well, and that rarely is limited to humans.*

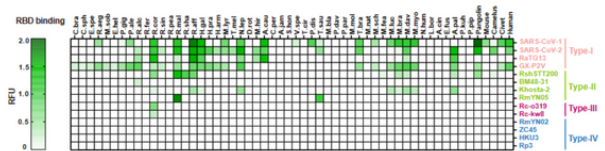
There is emergent evidence that the lives of their primary hosts, the Rhinolophids—horseshoe bats with odd noses—are not only a lot more crowded and cosmopolitan than the average bat, but also a lot more socially intricate. Horseshoe bats vary in nose shapes because that allows them to specialize in echolocation at very wide ranges of frequencies.

While not fully understood today, researchers believe that bat echolocation is not only for orientation and hunting but also for communication and vocal learning. Some bat's extensive vocal repertoire is needed to create specialized social calls used either for parent-offspring reunions, territorial defense, or maintaining group integration. Some speculate that the evolution and differentiation of vocal frequencies within a species leads to mating preferences and exclusions, driving niche formation and ultimate species separation into species complexes.

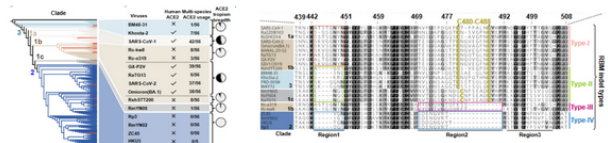
A species complex is interesting from the perspective of viruses. The genetic stratification of their host reservoir means that some viruses will specialize on a small niche or subgroup of the species complex, whereas others might try to develop or maintain broad affinity to multiple species that might still inhabit the same spaces, but not mate with each other.

Let's recall **an observation from Chapter 1**: the RBDs of sarbecoviruses can be very broad (even opening the door to non-bat ACE2 receptors), or very niche.

### Many Sarbecoviruses show broad species tropism and can use hACE2 as a host receptor from the get go



1) Binding assay of 14 sarbecovirus RBD against cell cultures expressing ACE2 receptor protein of 56 species show broad binding affinity (incl. multiple bat species, but also pangolin, mouse, civet, camel and human)



2) The broader the binding affinity, the more likely that an RBD will also be able to infect human cells via hACE2  
3) Sarbecoviruses RBM motives are highly diverse, with the full-length RBMs typically conferring broad species tropism and double region deletions resulting in loss of ACE2 entirely

Sarbecovirus RBMs and ACE2 tropism are likely a reflection of the population structure and intricate social lives of their host species (Figures from Si J. et al., biorxiv, 2024)



## Rhinolophids are gregarious neighbors

It is reasonable to assume that the intricate population structures of Rhinolophids create evolutionary pressures for viruses to constantly adapt to new niches. Recombination as an **"evolutionary fast-forward"** is the mechanism of how sarbecoviruses survive in such an environment, which is why this is such a prominent and critical feature in their evolution.

*"Recombinant viruses that are most likely to succeed under competitive circumstances with small population sizes are those for which the recombination event has conferred a strong selective advantage over both parental viruses."*

—[Wells H. et al., Cell Host Microbe, 2023](#)

While various horseshoe bat species were identified as the most prominent hosts and natural reservoirs of sarbecoviruses in the wild, they are certainly not the only bat family involved in creating overall sarbecoviral diversity.

**Viral recombination needs physical opportunity for other viruses to infect the same cell;** parental viruses can not create offspring if they never meet, but their offspring might also not be as diverse if they never meet very distinct others. So how do our hypersocial rhinolophids do on that "meeting new people" front?



Some bat species have been observed to regularly co-mingle with other species, especially in cooler caves or during torpor (short-term hibernation). Image credit: Prof. Alice C Hughed, University of Hong Kong.

**Beyond mere cuddling with strangers, rhinolophids are known to switch roosts constantly.** There are maternity-only roosts, summer and winter locations, and roosts that are calm or have special geographic features. Some bats like to stay as permanent residents in a single cave, others cycle outside but come back periodically over the year—*maybe kids going to college would be a social analogy, if only we could tell bat age, that is*—and many bats do food tourism with the seasons. All of this leads to constant turnover, and various bat species mixing and mingling with each other.

*Researchers observe complex social, seasonal, and geographic mixing of various bat species even when just studying single locations, like a specific cave or a forest.*



Rhinolophus bat species are very diverse and their nostril shapes have been optimized for echolocation at a large range of frequencies. Shown: R. rex, R. stheno, R. l. yunanensis, R. sinicus, R. malayanus (from left to right) Image credit: Prof. Alice Hughes, University of Hong Kong

## Virus-sharing networks stir the genetic cauldron

**Some of this mingling and movement was always a natural part of bat lives.** More recently, however, human encroachment on bat territory—urbanization, deforestation, hunting, tourism, mining, etc—is considered the biggest factor in displacing bats from their traditional roost sites and forcing species together that might not have ordinarily met.

“Some of those different species will roost together, especially when it is cooler or during hibernation. So when we have done work in various cooler caves [...], we will see clusters. *Rhinolophus* [horseshoe bats] cuddled up to *Miniopterus* [long-winged bats], and the next one is *Myotis* [mouse-eared bats]. Now these are lineages that diverged 50 million years ago.”

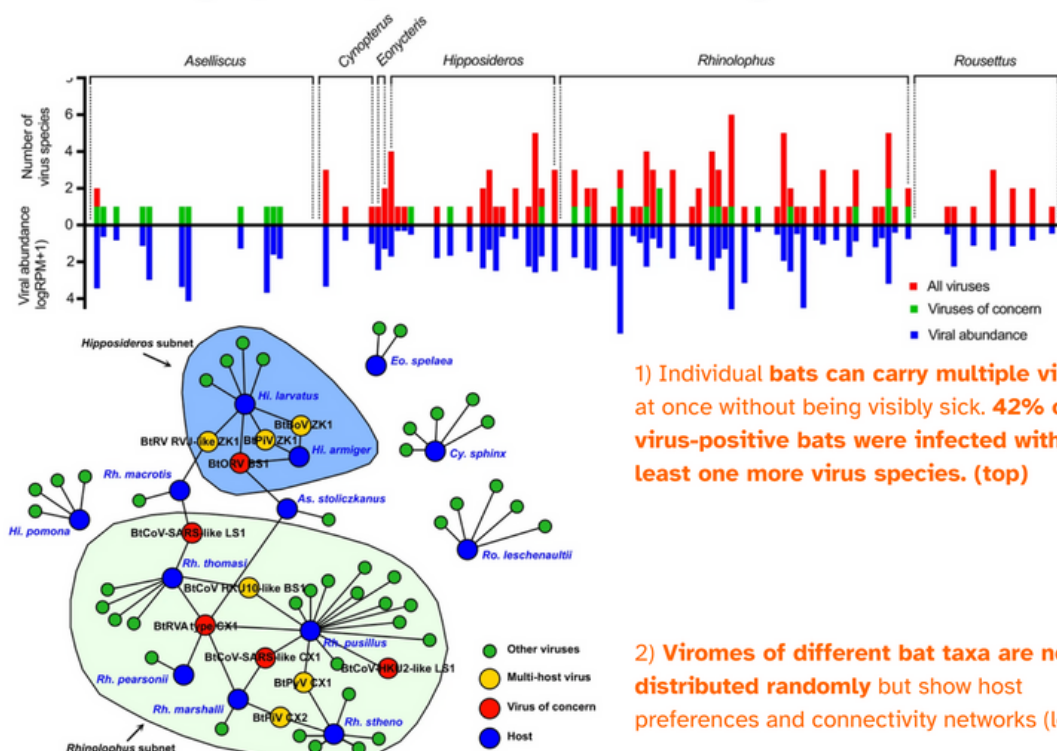
—Bat ecologist Prof. Alice Hughes

If some bat species that diverged often millions of years ago tend to huddle together—and we already learned that the ACE2 tropism of various sarbecovirus RBDs can be very broad—do the diverse viruses they carry infect each other and influence recombination patterns as well?

*This is one question researchers were able to answer recently with a resounding yes.*

**By using a meta-transcriptomics approach**—taking a sample and sequencing every single piece of genetic information in it—they discovered that many of the bats they sampled carried more than one virus. On top of that, the study found that some of the discovered viruses were shared between different bat species, suggesting frequent spillover events distributed the virome over multiple hosts. Dramatic opportunity for virus sex (recombination) and diversifying the viral gene pool with new genetic elements.

## Meta-transcriptomics show that sarbecoviruses exist in highly-competitive virus-sharing networks



1) Individual bats can carry multiple viruses at once without being visibly sick. 42% of virus-positive bats were infected with at least one more virus species. (top)

2) Viromes of different bat taxa are not distributed randomly but show host preferences and connectivity networks (left)

Co-infection with multiple viruses is a prerequisite for virus recombination. Diverse virus-sharing network reveals connectivity among viromes of different bat taxa. Viruses of concern and putative cross-species transmissions are shown in different colors. (Wang J. et al, Nature communications, 2023)

## Nature's gain-of-function laboratory

“The frequent virus spillover among phylogenetically related or spatially co-located bats provides an opportunity for viromes of different bat species to exchange, further expanding genetic diversity of circulating viruses”

—[Wang J. et al, Nature communications, 2023](#)

“The phylogeographic structure in bat hosts (and their diverse immune strategies) [...] creates a landscape of selective pressure; the trajectory of viruses' coevolutionary response is, in turn, constrained by their opportunities for either specialization or diversification through host jumps and recombination”

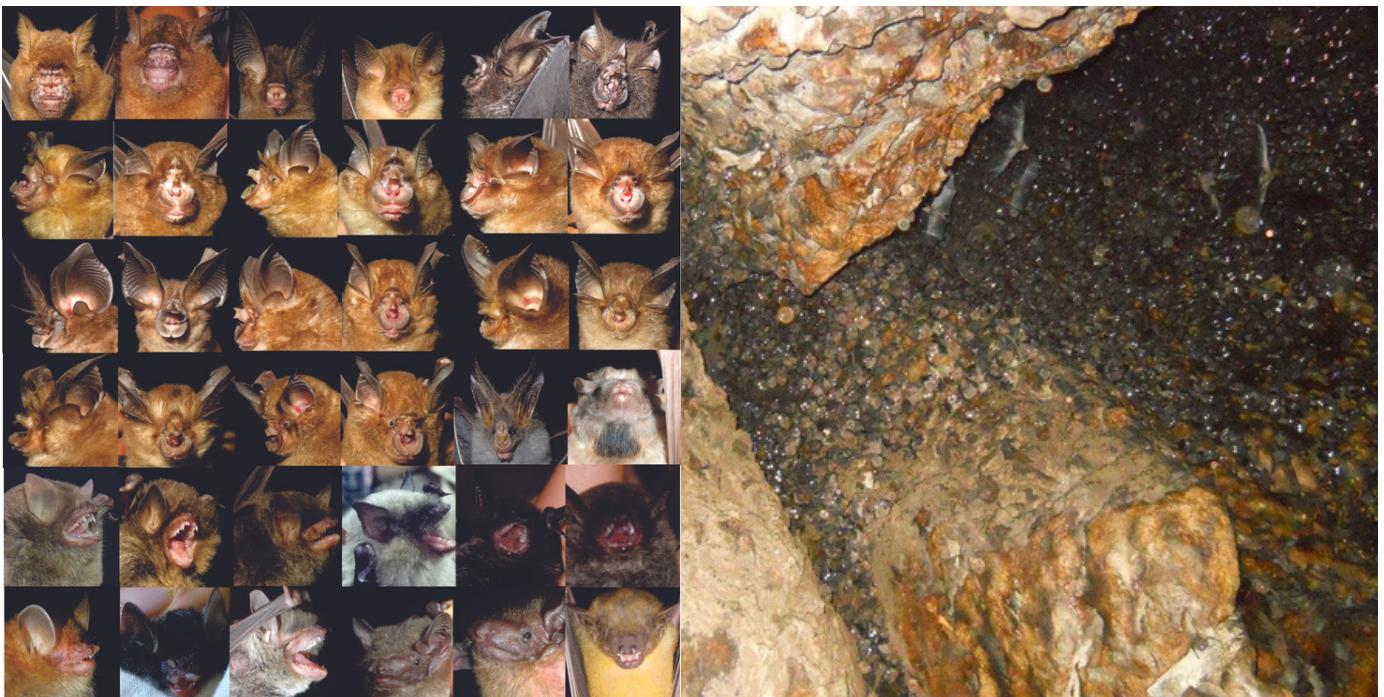
—[Forero-Muñoz NR. et al., Virus evolution, 2024](#)

I think these findings are important to consider. Whatever innate mechanistic ability for recombination sarbecovirus genomes might have, **it is largely the social lives of their host reservoir that create the opportunity and evolutionary pressures for recombination** to shape sarbecovirus genomes.

In other words, we should not be surprised to find a complex mosaic within the genomes of sarbecoviruses, because it reflects the complex mosaic of bat and virus co-evolution. A constant fight for niche survival against the merciless competition of ever-changing circumstances, one where only the most opportune genetic elements and fittest chimeric viruses will persist.

*These caves are nature's vast "gain-of-function" laboratory, all while our human encroachment on bat territory is stirring the genetic cauldron ever faster.*

So now that we know what “gain-of-function labs” we are looking for, do we know where to look? **Where exactly is the birthplace of SARS-CoV-2?**



We humans tend to underestimate the diversity and vastness of nature. Even over geographically confined regions, vast cave networks can house millions of bats, dozens of different bat species and sub-species not yet described or sequenced. Image credit: Prof. Alice Hughes, University of Hong Kong

## II) Undersampled Karst in the Wild Wild East

**Karst** is a topography formed by the **erosion of soluble carbonate rocks**, such as limestone or dolomite. With a total area of 500.000 square kilometers in China and 400.000 square kilometers spread over various nations in Southeast Asia, these countries house some of the largest and most biodiverse Karst regions in the world. But it is truly the power of water over eons that has shaped the landscape to include spiky spires, enormous sinkholes, underground rivers, and intricate cave systems below the old-growth subtropical forest and rainforests. This is bat country, and each valley and limestone formation can not only house millions of bats but is a microcosm in itself.

“Each isolated limestone hill can host more than 12 unique species found nowhere else on earth, with up to 100 microsnails, endemic begonias, orchids, and geckos, and yet an estimated 90% of cave-dependent species are undescribed”

—Prof. Alice Hughes, University of Hong Kong

**The Karst region spanning Southern China and Southeast Asia** forms a unique and biodiverse ecosystem that in considerable parts is still relatively untouched by humans, albeit that has been changing especially with deforestation, slash-and-burn, and cash crop agriculture leading to the degradation of the land. Its importance to the global climate and conservation is probably second behind the Amazon rainforest, with projections suggesting around 40% loss until the end of the century if we do not dramatically change course.

**The shifting landscape use and human encroachment** are in turn some of the major drivers of zoonotic spillover. Some studies suggest that around 65.000 spillovers of CoVs happen each year, most of them from viruses that are yet undiscovered and that do not cause an outbreak for a variety of reasons.



A guano collector in a massive limestone cave housing millions of bats. Bat guano is sold as potent fertilizer and can be harvested periodically.

The majority of viruses that spill over might just not replicate well in the host cell, or get shut down by our innate human host defense, such as the interferon pathway, without humans ever being the wiser. Even most replicating viruses might be unable to transmit between humans, or unable to spread effectively enough. Outbreaks are also always social phenomena, even a perfectly capable pandemic virus like SARS-CoV-2 would have died out with a 99,5% likelihood if it spilled over in a remote village, rather than a megacity.

*continued on next page*



Karst rock formations in Cat Ba, Northern Vietnam. Image credit: Prof. Alice Hughes, University of Hong Kong

## A granularity problem

Given most of these sarbecoviruses die out and will never be known, narrowing down where exactly SARS-CoV-2 came from within an often inaccessible, remote Karst region 2,5 times bigger than the size of Germany is a daunting challenge.

“The risk for zoonotic viral disease presence and emergence in humans also increases in geographic areas with higher mammal diversity, where previously pristine forests have been recently deforested. This makes the areas of Southeast Asia, which support large, intact natural habitats and have ongoing ecosystem fragmentation, at high risk for disease emergence”

—[Evans TS. et al., Int J. Infect Dis, 2023](#)

Yet this background of viral emergence and constant spillovers gives researchers **three lines of attack to at least narrow down** where the bat ancestor of SARS-CoV-2 likely originated.

- discover the immunological footprint of related viruses through serology in humans and farm animals
- discover the immunological footprint of related viruses through serology in wild bats
- discover related viruses through taking and sequencing bat samples

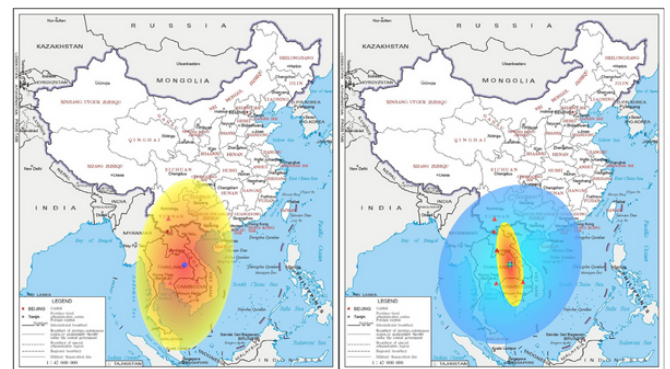


Bat researchers catch bats with nets, transfer them to cotton bags, take measurements and various samples (sliver, excrement, tissue) in *impromptu* field labs, then release the bats unharmed after

Evidence for SARS-CoV-2-related sarbecoviruses has been found in China, Laos, Vietnam, Cambodia, Thailand, Myanmar, and Malaysia.

- Most direct evidence comes from sequenced bat viruses in China (before sampling in Yunnan became forbidden), where researchers found a pool of SC2-related viruses in Yunnan province. (China has been more sampled than other nations)
- Another trove of informative viruses was found by French and Laotian researchers in Vientiane province in Laos. Here the closest match to SARS-CoV-2 broad-affinity RBD was found, proving nature unlocked that particular door to human infection; not human engineering.
- More distant SC2-related viruses have also been found in Cambodia and Thailand; which in themselves can house informative genetic elements such as insertions at the S1/S2 site reminiscent of the polybasic cleavage motif in SC2

Some have summarized the possible origin place at very low resolution (below) based on the sampling location and overall genetic similarity of the various SC2-related viruses.



Low-resolution likely geographic origin based on overall genetic and RBD similarity of a set of closely related SC2-like bat viruses and their original sampling site (red triangle) [Zhao S. et al., J Genet Genomics, 2022](#)

Yet such efforts are bound to be highly dependent on sampling biases and offer a low granularity that is almost meaningless.

*continued on next page*

## All we know is that we are flying blind

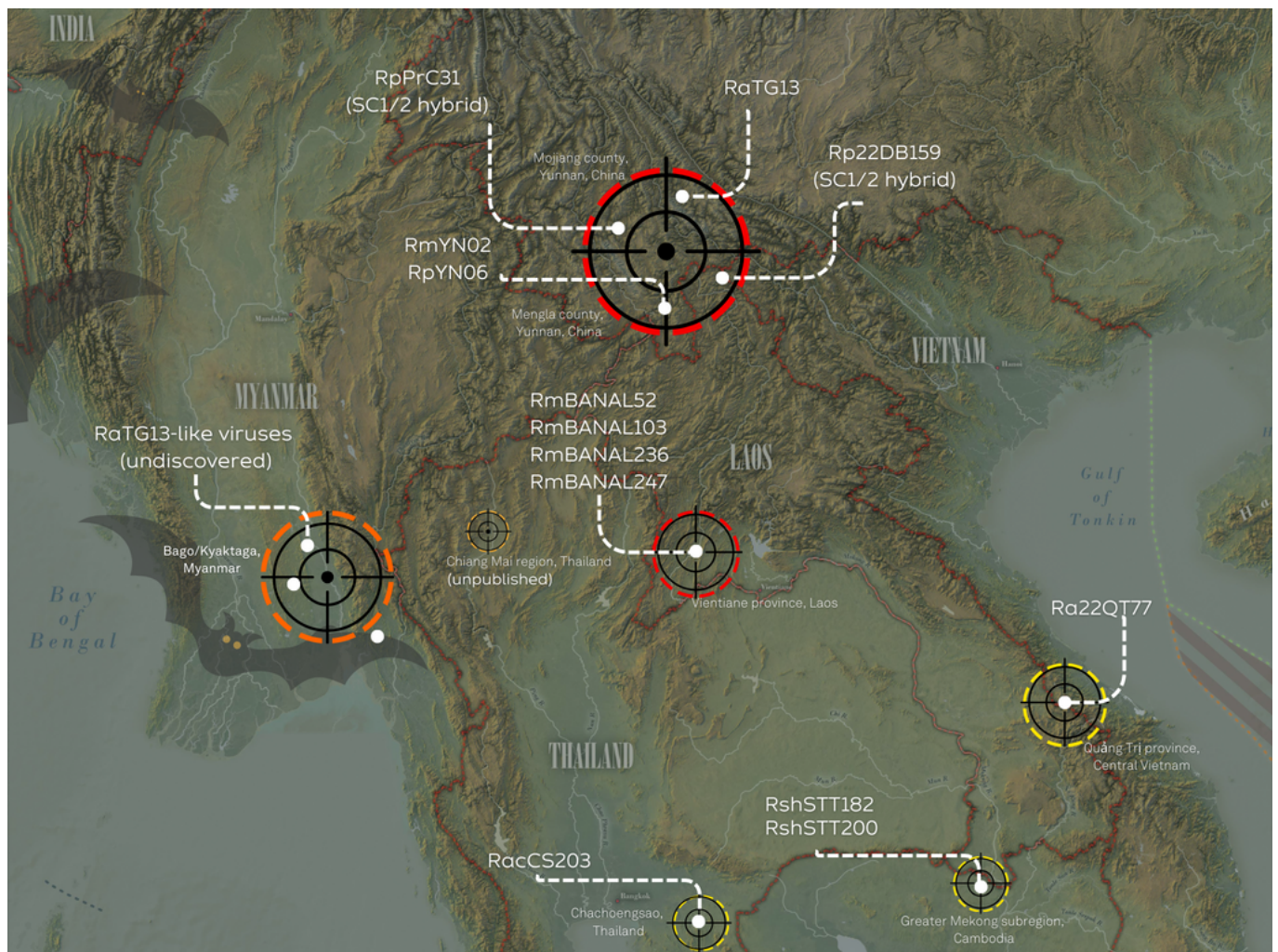
The previous map has also not considered some other relevant findings such as:

- SC2-related viruses have been found in pangolins smuggled over the Chinese and Thai borders
- Especially the karst regions bordering Laos, Northern Vietnam, and Myanmar seemed very rich in these viruses, with Myanmar lacking any bat sampling efforts there so far
- However, serological evidence from Myanmar further South found a very high prevalence of SC2-related viruses in humans that had contact with wildlife. More than 1 out of 5 had antibodies and cross-reactivity to a panel of these viruses (most prominently RaTG13), arguing for high prevalence and spillover potential in this region, but the concrete viruses remain undiscovered

- Bat sampling in Japan, Southern Vietnam, and Malaysia shows more ancestral and distantly related sarbecoviruses
- Sampling in Northern Thailand has been politically difficult to publish but might be insightful as well

*Honestly, while a lot of valuable information has been gathered in the last few years, it serves merely as a snapshot of the scope of our ongoing ignorance.*

Below, I have plotted a rough summary of scientist's efforts so far:



**Geographic location of SARS-CoV-2 relatives sampled in Southern China and Southeast Asia.** Large parts of the Karst region, especially outside of China, remain hopelessly unsampled. Technical names of virus relatives usually include the bat species as first letters. Rm ... Rhinolophus malayanus, Ra ... Rhinolophus affinis, Rsh... rhinolophus stheno, Rp ... Rhinolophus pussilis, Rac ... Rhinolophus acuminatus. The displayed color coding roughly indicates genetic similarity to SARS-CoV-2 (yellow—medium, orange—high, red—very high)

**The reality is that there is still much we do not understand about our diverse world.**

According to Prof. Alice Hughes, it is estimated that 90% of caves in Southeast Asia alone remain scientifically undescribed and uninventoried, while some estimate around 60% of bat species remain to be discovered, and a staggering many of the bat species we know about have never been sequenced.



While bats remain undersampled, cave tourism in the karst region is booming. Image: Tham Lot Cave, Northern Thailand

We still need to learn much more about viral and bat geography, how human drivers from deforestation to urbanization to tourism interface with bat immunity, viral-host interactions, bat-human contacts, and propensity for viral shedding.

*We are working with a map where many spots are and will likely remain blank because any research on bats has been so highly politicized by the distorted origins debate.*

How long can this dangerous mix of ignorance and power continue?

## Why governments are not keen on virus discovery research

There is a need to study these complex interactions between viruses, bats, and humans. However, many Southeast Asian nations, as well as China, have little interest in allowing researchers to sample more bats, discover new viruses, or even publish ones they have already found (*personal communication from multiple bat hunters*).

**Especially sarbecoviruses are a sensitive subject almost everywhere.** Governments are currently not interested in discovering where these dangerous chimeras are circulating in the wild or what drives their emergence.

*Some governments fear more scientific discoveries will lead to them being blamed by a world that has not made peace with natural pandemic risks.*

A world that asks for reparations, culprits, or impossible insurances, rather than offering a helping hand on a collective problem. A world where leaders seek power with geopolitical grandstanding, or worse, gain popularity by exerting vengeance on inconvenient scientists rather than heeding their warnings.

I believe collectively, we need to do better on that front.

Scientists are doing their part, despite these artificial limitations and political constraints. Some of them have been tinkering with new methods and ingenious ideas to learn more about the origin history of SARS-CoV-2, and there are some exciting findings to report.

### III) Ghost hunting with phylogenetic inference

Gaining knowledge from sparse data is difficult, but not impossible.

To date, only ~200 sarbecoviruses have been found. 139 of them are closely related to SARS-CoV-1, and 26 are SARS-CoV-2-like. Not much to work with to pinpoint a geographic location, and individual genomes can also not be used to reconstruct a faithful evolutionary history because all of these are chimeras, so molecular clock estimates based on mutational divergence are being misled by the “evolutionary fast-forwards” of recombination.

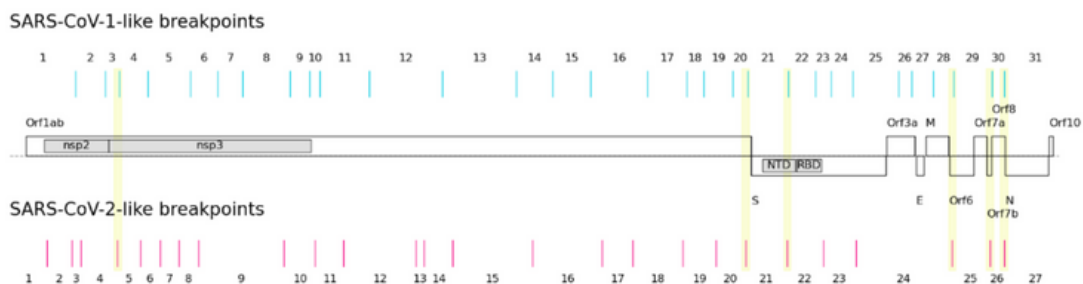
The number of sarbecovirus relatives discovered to date is however just good enough to identify **recombination breakpoints**, places where genetic similarity from one parental strain stops and similarity to a different parental strain starts. A way to think about recombination breakpoints is as a minimum number of sexual acts that must have happened in the past that can explain the segmented shape of the chimeric genome today.

The genomic segments in between recombination breakpoints are called **non-recombinant regions (NRRs)**, basically intact genomic segments from a parental strain. On these segments, molecular clock estimates work normally and can thus be informative. Virologists will just have to individually work with as many clocks and evolutionary histories as there are NRRs, which is 27 in the case of SARS-CoV-2, and 31 for SARS-CoV-1. (see figure below)

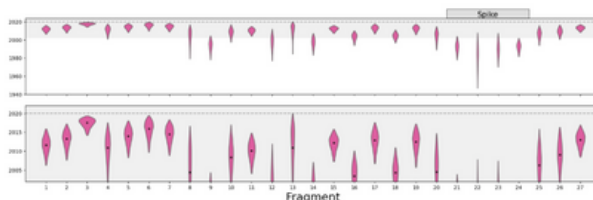
“If you align all viruses in the same place, and then chop up the genome into segments—what we call non-recombinant regions (NRRs)—then each NRR will have their evolutionary history”

— Dr. Spyros Lytras, an evolutionary virologist at the University of Tokyo

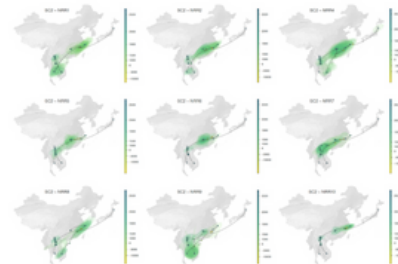
### Recombination-aware phylogenetic inference for each non-recombinant region (NRR) maps evolutionary history



1) **Recombination breakpoints** for SARS-CoV-1 and SARS-CoV-2 separate the genomes into 31 and 27 **non-recombinant regions (NRRs)** respectively.



2) Each NRR is a segment of a parental virus that became defined by a recombination event, thus has its **individual age estimate** (time period of when it was created)



3) Each NRR has its own phylogeny and thus **independent evolutionary history**

**Non-recombinant regions are informative and need to be analyzed individually** to be able to present—to the extent possible—a single evolutionary history of chimeric viruses. (Pekar et al., biorxiv, 2023)



Separating the histories of NRRs allows researchers to gain some critical information. For example, while some NRRs might be very old and just serve as reminders of sexual conduct decades ago, others might be very recent additions; the final sexual (recombination) events that gave rise to the chimera.

Indeed, researchers found that **the youngest recombinant additions** for SARS-CoV-1 and SARS-CoV-2 to their closest-inferred bat virus ancestor happened not decades ago, but were in 2001 (SARS-CoV-1 spilled over in 2002) and **in 2014 (SARS-CoV-2 spilled over in 2019)**, respectively. This is of course not the final word, as the researchers noticed the more cousins are discovered, the closer the time estimate moves up towards emergence date.

“The virus pieces that were most similar circulated in bats very recently”

—Dr. Jonathan Pekar, US San Diego.

Having identified the 27 disparate recombinant pieces constituting SARS-CoV-2 also allows researchers to do something very interesting, which is constructing the **closest common recombinant ancestor (recCA)** genome, basically looking at which viral cousin of SC2 has the highest similarity to SARS-CoV-2 for each NRR.

*The recCA genome is an aggregate, built from the ghosts of closest-inferred ancestors for each NRR that must have existed at one point.*

Tellingly, researchers have found that the more bat virus relatives got discovered in nature, the closer the genome sequence of the recCA resembled the human-infecting SARS-CoV-2 that emerged.

**In other words, we have evidence to infer that a 98.8% identical to SC2 bat virus ancestor circulated just a few years before emergence;** with temporal and genetic granularity still increasing as more viral cousins are discovered.

But what can we do with that knowledge?

## On the limits of science communication

Making science accessible is a balance act between accuracy, complexity reduction and *useful but inaccurate* abstractions (i.e “virus sex” for recombination).

Some scientific topics lend themselves to complexity reduction, others not so much. Sometimes a communicator has a very good understanding or creativity to explain a technical topic brilliantly easy, but nobody can do or know everything at the required depths.

The viral phylogenetic inference covered here is a deeply technical topic based on statistical models, probabilistic assumptions, and experienced modellers using real data for inference.

The research is currently under peer-review as this is the cutting edge of origins research.

For more in depths information about this work, a detailed interview with the two first authors can be found here:



You can also access the video here:

<https://www.youtube.com/watch?v=ROUdGeCmVYk>

I highly recommend listening to primary sources put their work into context, on top of that, it was a really fun conversation with two remarkable young scientists.

## IV) Reconstructing the phylogeographic histories

As we have observed in the karst region, sarbecoviruses, just like their bat hosts, show a degree of geographic structuring; with the border region of Laos, Myanmar, Northern Vietnam, and Southern China seemingly quite rich in sarbecoviruses closely related to SARS-CoV-2.

But does that mean SARS-CoV-2 came from that region, or is it just a mixture of undersampling and coincidence?

By using these **phylogeographic data**, a phylogeny (family tree) for each NRR can be constructed that is spatially mapped to sampling locations.

“When we build those family trees, we want to calibrate them with time”

—Jonathan Pekar, US San Diego

**Combining geographic information** from sampling locations of close ancestors **with phylogeny scaled with units of time** of their closest recombinant ancestors creates a geo-temporal record for the various genetic elements constituting the SARS-CoV-2 viral genome.

These geospatial and temporal efforts also allow us to **narrow down a dispersion zone** for these “floating genetic elements”, and with it, the viruses that arose by their (re)combination. In other words, the **most likely geographic birthplaces** of direct bat ancestors to SARS-CoV-1 and SARS-CoV-2 (see below).

This last result gained from studying sarbecovirus ancestry **unearthed one more essential clue** to how SARS-CoV-1 and SARS-CoV-2 emerged.

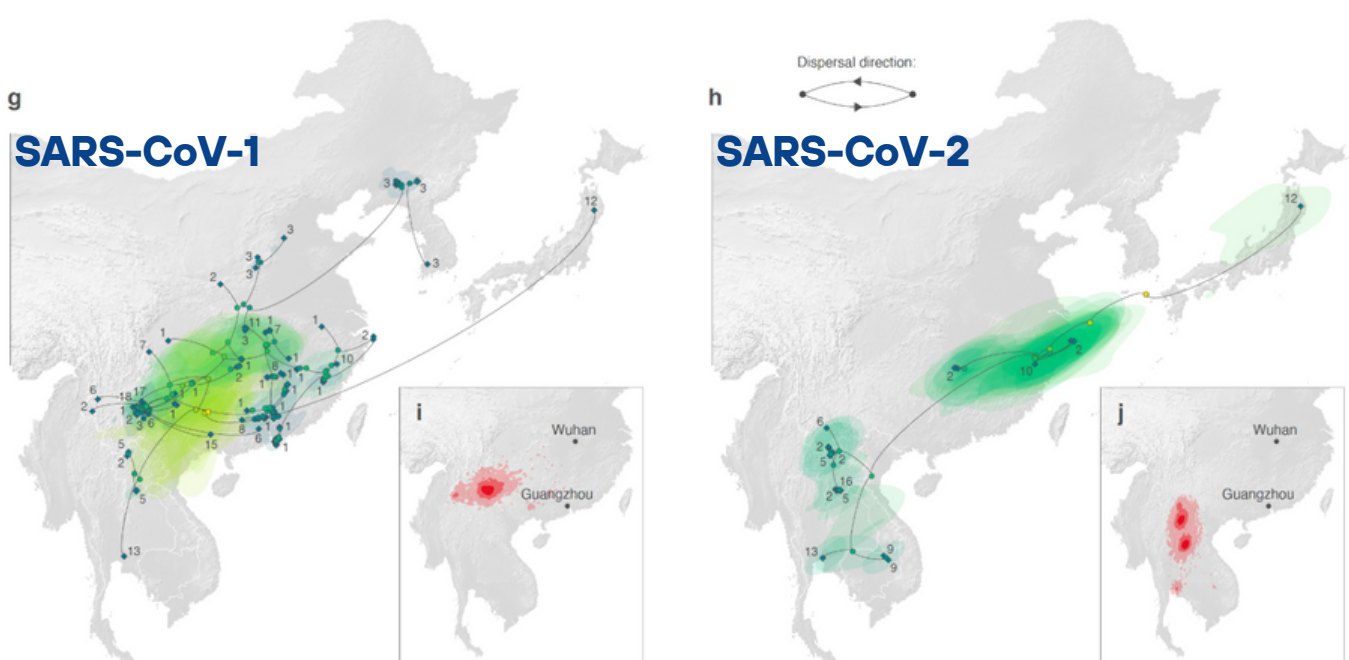
And it starts with a suspicious conclusion:

“Direct ancestors of the SARS-CoVs likely could not have reached sites of emergence via the bat reservoir alone”

— Pekar et al., biorxiv, 2023

SARS-COV-1 emerged in Guangzhou, around 1000km away from its suspected home; and SARS-COV-2 emerged in Wuhan, around 1500 km away from where the last direct bat ancestor existed.

*So how exactly did SARS-CoV-2 find its long way from the sarbecovirus “gain-of-function” heart of the Karst region to a wet market in Wuhan?*



**Phylogeographic origin dispersion map of SARS-CoV-1 (left) and SARS-CoV-2 (right).** Green density map tracks evolutionary history through time, red density (right bottom corners) are the most likely birthplaces of the direct bat ancestors to SARS-CoV-1 and SARS-CoV-2 (Pekar et al., biorxiv, 2023)

## V) Smoking guns pointing at the wildlife industry

An outbreak rarely divulges all of its mysteries. How the virus made it to Wuhan still leaves room for uncertainty and speculation. *Lab leak advocates might be quick to smell an opportunity to allege that maybe a researcher from Wuhan has gotten infected and brought the virus to the city. Weren't Zhengli's group and some other teams sampling in Yunnan after all?*

These insinuations are unfortunately quite shallow, scientifically naive, and ultimately false because the leftover uncertainties are much smaller than one might appreciate.

But one step at a time. I could spill much more digital ink on arguing how *a priori* unlikely it is for bat researchers and virus hunters to ever stumble upon a pandemic-ready chimeric virus from the small bat sampling efforts researchers ever get to conduct. One might find genetic puzzle pieces, but how nature put them together only comes to light after selection created those million-to-one lottery winners that spill over into other mammals. If Chinese researchers had collected and stored hundreds of those spillover viruses, maybe suspicions had a ground to stand on. The reality is that Zhengli Shi's team only ever found one SARS-CoV-2-related virus, even non-pandemic viruses are hard to find. One could also contend there is no evidence for any of these speculations. One might also list the eerie circumstances surrounding SARS-CoV-2 that look an awful lot like SARS-CoV-1; from the fact that both started in November (suggesting seasonality) to the finding that sampled wildlife farms around Wuhan had SARS-CoV-1 lineages potentially ancestral to the outbreak that happened in Guangzhou... *but okay, I am spilling digital ink again, so I will stop. See note (left)*

Again, it is 2024 and scientists have uncovered a lot more **evidence that completely exonerates** Chinese and other **bat researchers** who ever sampled in Yunnan or various places in South East Asia.

*In the last section, let us focus on just **two puzzle pieces** that I think are illuminating because **no type of research-related accident can explain them.***

## Note on: Uncertainty inflation versus scientific triangulation

On popular and contested topics, media manipulators utilize an arsenal of misleading communication tactics to sow doubt, imply ambiguity, or create confusion about inconvenient scientific studies. They know that **citizens can be misled into discarding nuanced findings and even whole papers** by pointing out limitations, alternative explanations or simple unintuitive findings that need further investigations.

Any scientific study or experiment has at least some area of messy or incomplete data, unknown or undefined variables, unexplored alternatives or other limitation. That is just the nature of science, if we had perfect data and perfect knowledge, nobody would need to write a paper anymore. So is science doomed by the never-ending contestations of conspiratorial communities and motivated manipulators?

Of course not. The existence of limitations and uncertainty (*artificially inflated or not*) with individual studies does not need to strand us in eternal ignorance on the origin topic, nor should it entice us to throw the baby out with the bathwater.

This is because of **scientific triangulation** or sometimes called **information convergence**; *the use of multiple methods or data sources in qualitative research to develop a comprehensive understanding of phenomena.*

The more methodologically diverse, complementary and independent the research approaches and teams behind scientific papers on the topic are, the more we can trust them if they all converge towards the same conclusions.

The body of evidence behind a non-research related origin of SARS-CoV-2 is entirely one-sided, converging and can explain all the available evidence. **Individual uncertainties and alternative speculations do not have the power to move that mountain of evidence any more.** Nor can any alternative explanation be formulated that can explain all the evidence plausibly.

## The first piece

The first one is – *what I feel* – one of the biggest media oversights in the discussion, and it has to do with the **damned furin cleavage site** again.

In Chapter 1, we spend considerable time explaining why we know that the **FCS does not have an artificial origin**. We also know from looking at sarbecoviruses that while these FCS motifs can easily be created, they do not seem to be maintained in bat viral lineages. They are even unstable in many cell culture systems. However – within certain real-life transmission contexts – the FCS is strongly **preserved as a respiratory adaption**.

That extends beyond humans; minks, deers, and other animals that got infected with SARS-CoV-2 via reverse spillover and the respiratory route from humans have been shown to preserve the FCS as well. Bats do not appear to do so, for reasons one can speculate (e.g. *bat sarbecoviruses are gut-adapted, not respiratory*) but are certainly **host-context dependent**. No genetic element acts alone.

This observation however means that one has to assume that **the closest bat ancestor of SARS-COV-2 almost certainly did not have (or maintain) an FCS**.

*If neither bats nor humans created the FCS, where does it come from?*

By just following these two evidence-based assessments to their logical conclusion, it becomes clear that an intermediate host had to be involved.

Somewhen **between as early as 2014** (likely later) when a recombination event allowed the bat ancestor to reach its final chimeric form, **but before it emerged in 2019** at the Huanan wildlife market, the **virus must have acquired and/or sustained it's FCS in a non-bat host** (*complimentary evidence like D614G stabilization also argue for a very recent FCS addition*).

*"The Furin cleavage site gave me the strongest hint that the progenitor virus of SARS-CoV-2 is not in bats",*

*–Prof. Linfa Wang, Duke-NUS Singapore*

**There is a recent precedent.** Evidence for such FCS acquisition in intermediate host species has been found in 2023 with trafficked pangolins. A bat HKU4-related merbecovirus (*bat ancestors to MERS also do not contain an FCS*) has acquired a minimal polybasic cleavage motif experimentally shown to be cleaved by furin. It seems that for CoV spillovers, **intermediate animal populations** tend to bring these polybasic cleavage sites forth and **can maintain them**.

An identical argument for another such **respiratory adaptation that gets maintained** could also be made for the T372A mutation that alters the 3D confirmation of the trimeric S glycoprotein to a more open and respiratory infectious form. Bat sarbecoviruses do not have this, humans could not have come up with it. So where did it come from? *Same story really*.

*A priori*, it is very unlikely that bats directly infect bat researchers, and even extremely unlikely to ever find a pandemic pathogen from the pithy sampling efforts bat researchers ever get to conduct. But add to this the likely requirement of a **host-context switch and respiratory adaptation** in an intermediate animal to maintain an FCS/T372A? Pretty much impossible.

I trust this should finally dispel the myth that bat-sampling researchers brought SARS-CoV-2 to Wuhan.

*"With the US Intelligence agencies, they all focus on the Wuhan Institute of Virology because they have the most bat samples...  
... I said, yes, but science says that this progenitor comes from a non-bat small mammal"*

*–Prof. Linfa Wang, Duke-NUS Singapore*

There has been a lot of hysteria and myth-making surrounding the furin-cleavage site in the media over the last four years.

To **understand that the one genetic element** many have falsely called the "smoking gun for engineering" **contributes much more to bat researchers' exoneration than to their incrimination** is certainly a fitting and long-overdue twist of fate.

Talking about non-bat intermediate hosts though...

## The second piece

Something no research-related accident can explain is the fact that the multiple independent lines of evidence point to the Huanan market as the epicenter of the pandemic.

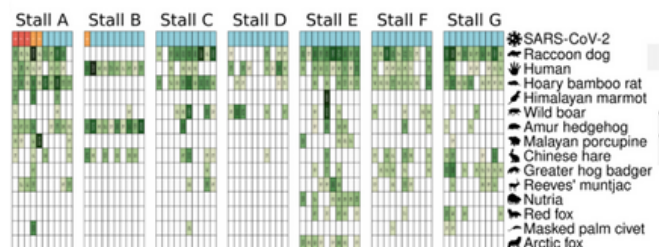
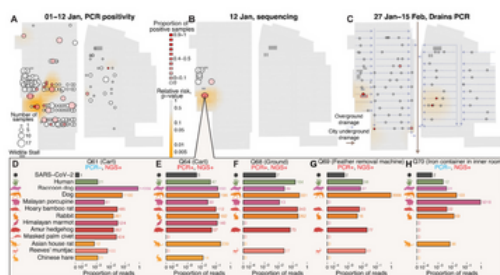
**Researchers had already established in 2022** that **SARS-CoV-2 susceptible wildlife** had been sold at the market – *despite denials and obfuscations from Chinese authorities continuing to this day* – and that the spreading pattern of two lineages centering from the market is best explained by a multi-spillover scenario from an infected pool of animals. Much more could be said here, but let's go to the cutting edge right away:

In 2023, sequencing data of environmental samples from the market were finally released by Chinese scientists which provided a trove of genetic information to sift through for new possible clues.

And new clues were found.

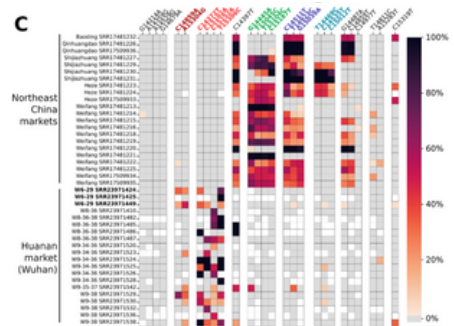
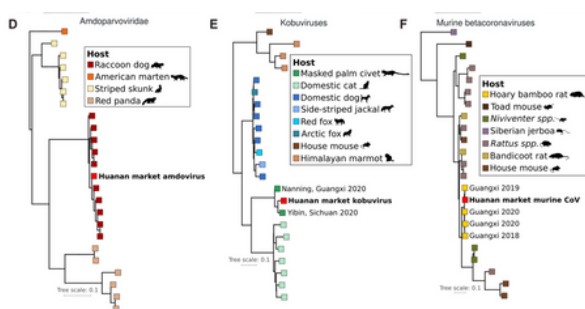
- Genetic diversity of found SARS-CoV-2 samples is **consistent with viral emergence** at the market; also contradicts the market as a mere "amplifier" or "superspreading" event
- Environmental samples that tested positive for SARS-CoV-2 contained lots of DNA/RNA from various **SARS-CoV-2 susceptible wildlife** species
- Animal **DNA/RNA correlated spatially** with stalls that housed these animals, but not other places in the market
- Environmental samples from these wildlife stalls contained **sequencing reads for wildlife-specific animal viruses** and SARS-CoV-2, but not other human viruses (*arguing for sick animals and against human contamination of these samples*)
- single-nucleotide variation (SNV) analysis of raccoon dog reads found that the **raccoon dogs at the market did not belong to a commonly bred species** used for fur production in large operations in the north, but rather a **wild-caught variety**

## Metatranscriptomics provides a direct link from animals at the market to wildlife industry



1) Environmental samples positive for SARS-CoV-2 are full of wildlife DNA/RNA

2) Wildlife mtDNA spatially co-localized with a to a small set of stalls reported to house wildlife



3) Animal viruses associated with wildlife trade, but not human viruses, are discovered in animal environmental samples

4) Raccoon dog DNA shows indications of wild-caught raccoon dogs species, rather than more commonly farmed species for fur production

**Metatranscriptomic analysis of environmental samples taken from SC2-susceptible animals** at the Huanan market provide evidence for a link to the wildlife trade, possibly in Southern China. (Figures from *Crits-Christoph A. et al., biorxiv, 2023*)

## Remaining uncertainties are small

These market data make it clear that no infected lab worker just walked into Huanan to start the pandemic, nor sneezed on the animals to make it look like they were involved.

On top of that, the metatranscriptomics data provide **multiple direct links to the wildlife trade and industry**; from the species that were identified to the types of viruses that plagued those animals; and even to the geographic region where they might have come from. These findings are consistent with the WHO report that also suggested the involvement of the wildlife industry in Yunnan and beyond.

From illegal trafficking of rare animals over the border, hunting and trapping them in the wild for later sale or sustenance, or supplementing breeding in wildlife farms, the industry has many facets. Some activities are illegal but lack enforcement, some are traditional and culturally valued, while other activities such as wildlife farms were explicitly promoted before the pandemic.

“Local officials trumpeted the wildlife trade as a way to close the rural-urban divide and to meet ambitious national targets to alleviate poverty.”  
—Emily Fang, reporting for NPR

### So that is it? This is how the virus made it to Wuhan?

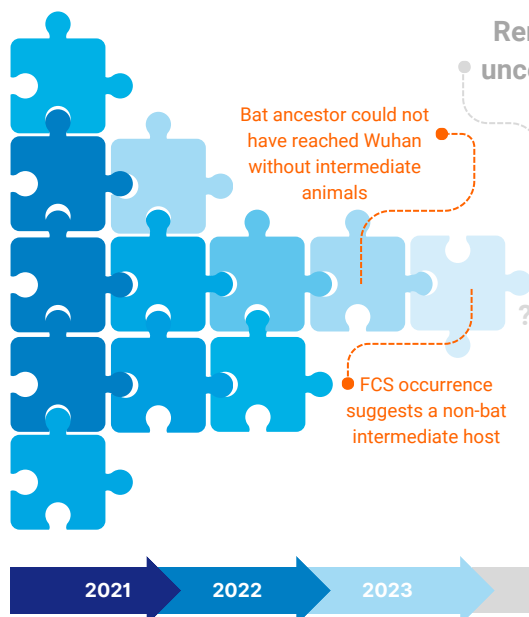
Outbreaks rarely divulge all their secrets, and insights have to be pried from nature and history through the often painstaking work of scientists, journalists, and other truthseekers.

Science is a process of approaching ever more likely explanations of reality by rooting out false hypotheses and narrowing down existing uncertainties.

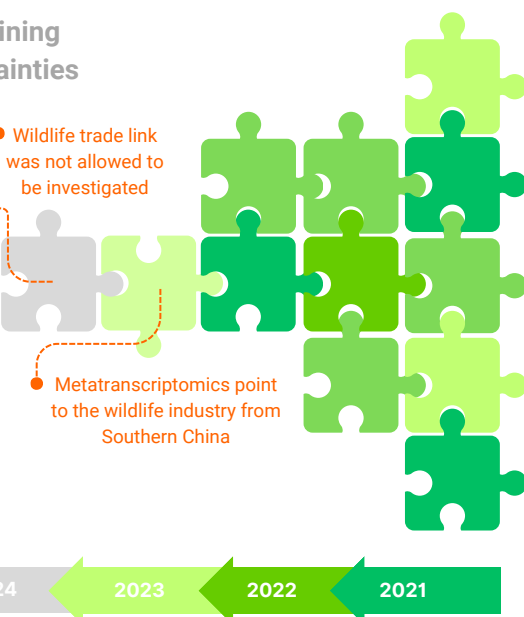
**The remaining scientific uncertainties of the origin puzzle (see below)** surround politically sensitive topics such as a **rampant regional wildlife industry**—estimated to be over USD 70 Billion in China and around USD 10 Billion in South East Asia just for trafficking—as well as more **complex patterns of how socioeconomic behavior and ecosystemic disruption influences spillover risk** factors and pandemic prevention policies at various virus/animal/human interfaces. That is why this type of “pandemic origin” research must continue, despite knowing this virus did not come from any lab. There is still **much more to learn on how to prevent SARS-CoV-3**. *But that is a topic for a different time.*

For now, I think we finally reached a (*hopefully satisfying*) end of this epic journey towards the cutting edge of origins research. So what is my personal takeaway?

### Accumulating evidence for a bat origin and direct ancestor of SARS-CoV-2 circulating in the wild



### Accumulating evidence for the Huanan market being the epicenter of the outbreak that started COVID-19



The remaining uncertainties are narrow but surround the regional wildlife industry’s role in circulating the virus and facilitating its ultimate emergence

# Conclusion: Don't bet against scientific inquiry

**Finding the origin of a new virus** is a lot more difficult than finding the needle in a haystack. Viruses evolve, adapt, branch out, and almost always die out without us ever being the wiser. It's like finding a particular grain of sand on a beach, always in danger of being washed away by the next wave. The search for any "bat patient zero" almost four years later would be a hopeless effort even before the deliberate sabotage of scientific processes by governments and malicious actors.

Yet in some twisted sense, we were lucky that the genome and evolution of sarbecoviruses are so intricate and reliant on recombination; it means that **SARS-CoV-2's evolutionary history contains many overlapping stories**. Stories that can provide a richer picture of what, when, where, why, and how it happened, akin to a set of contemporary witnesses with limited knowledge, or a set of partial fingerprints all over a crime scene.

*Puzzling these genetic and phylogeographic fingerprints together, documenting what places they touched at what time, and listening to what story they have to tell allows researchers to draw a much larger picture from the mosaic knowledge of our ignorance.*

The reality is that the observed viral ancestry of SARS-CoV-2 betrays any conceited notions we might have held about virologists—or gain-of-function research—bringing about the deadly chimeric pathogen. Scientists have the time and place of its birth and emergence narrowed down to a sufficient granularity to exclude any lab-related incidents. On top of that, we learned about some viral features that only natural selection, not human engineering, could bring forth.

Yet our understandable fears and suspicions of a "man-made" pandemic — *albeit misplaced into too simple narratives or wrongfully projected onto convenient scapegoats in suspicious white coats* — are also not entirely baseless.

As often with popular myths and sentiments, there is a grain of larger truths embedded in them.

Today, there is plenty of scientific evidence to make the case that collectively, we humans are not entirely innocent. Neither when it comes to fueling the forces that create these dangerous viruses, nor the reasons that make them spill over into us. In nature, no element acts alone. That includes us and the circumstances we create. As a now publicly smeared and unjustly character-assassinated zoologist once told me:

*"Ecosystem disruption, climate change, deforestation, wildlife trade and consumption... we see the human hand behind almost every outbreak"*

—Dr. Peter Daszak, EcoHealth Alliance

It is what we humans currently do to the planet and every other species on it that largely sets the conditions for zoonotic spillover events, and we drive them at an ever-accelerating pace.

What we can and should do about this sobering reality needs urgent societal discussion.

The unscientific speculations and fantasies of motivated manipulators do not. Citizens deserve better. They are owed clarity, not spectacle.

*continued on next page*

## False myths are counterproductive to solutions

I believe that the era where false "gain-of-function" origin myths **are wielded as weapons for persuasion, profit, or power** has to end.

Not only is this deeply unethical and scientifically irresponsible in 2024, but it has thus far proven counterproductive for scientific research, pandemic prevention efforts, and the public good.

In this article, I focused our attention on an **interconnected set of scientific knowledge related to viral recombination and its implications**.

It is important to mention that this - *while very comprehensive* - is just a mere facette of the larger origin picture. There are other comprehensive lines of investigation that this article could not go into, like epidemiology, which also points to viral emergence at the Huanan market and also contradicts any research-related origin scenario. No single article can cover it all.

**Understanding the origin story through the lens of viral recombination is incredibly powerful** because it exposes the shallow ideas and trivial falsehoods about this virus thrown into our faces every day. It also shows us that nature is much more complex, nuanced, and vast than we commonly give it credit for. It would be wise to exert more caution and compassion before trying to shape every last ecosystem to our human whims.

Maybe after four years of political myth-making and societal inaction, it is time to face scientific reality. I certainly believe we'd be better off fighting for solutions rather than for who is to blame.

**Because no matter what we want to believe**, these natural chimeras will keep haunting us if science and society don't come together to stop them.

*The end*

A comprehensive body of scientific evidence has shown that the immediate bat ancestor to SARS-CoV-2 came from one of the countless natural "gain-of-function labs" spanning the vast biodiverse Karst region from Yunnan in Southern China towards Myanmar, Laos, Thailand, Cambodia, Vietnam, and maybe even Malaysia in SE Asia.

The lingering and promiscuous endemic viral elements in that enormous geographic region constantly mix and bring forth new chimeric combinations within their socially intricate reservoir hosts; while human activities and encroachment on bat territories stir the genetic cauldron ever faster.

Once a particularly combustible set of genetic elements produced a potential pandemic pathogen with broad host tropism, the legal and illegal mammalian wildlife industry likely became the maturing vessels through which the virus we now know as SARS-CoV-2 reached its final explosive form. From there, it was dragged in front of hundreds of immune-naïve future hosts visiting the largest wet market of one particular Chinese megacity well connected with the entire world.

The rest is history, one we are about to repeat



## A word of caution about science communication versus peer-reviewed research and scientific reviews

In this article, I explained how multiple lines of scientific inquiry strongly exclude the possibility of gain-of-function research leading to COVID-19. Furthermore, the article highlights a body of literature warning about the clear and persistent danger of another zoonotic sarbecovirus spilling over into the human population.

Please consider that no single article can summarize all the valuable scientific contributions that scientists have made in support of investigating the origin question. It is a large and still-growing body of evidence spanning hundreds of papers that are mutually consistent with each other, the extant evidence, and the reality we live in.

I want deeply thank **Prof. Alice Hughes, Prof. David Robertson, Prof. Linfa Wang, Dr. Alex Crits-Christoph and Dr. Jonathan Pekar** for reviewing relevant sections of this article.

**My humble contribution** to this topic was to focus on **explaining recent key scientific arguments through the lens of viral recombination** for non-experts and also accurately present the conclusions scientists reached in their work by considering a large, detailed body of evidence.

While I did go to great lengths to avoid misrepresentations – *such as having hours of expert interviews and article reviews by domain experts before I post* – I cannot always control how my words will be interpreted. If there are uncertainties arising from my simplifications, omissions of brevity, or bad analogies, I advise you to first consult the primary literature for clarification rather than presume there is an obvious mistake in the science or reasoning of scientists.

At this point, the scientific consensus arising from a large body of evidence is pretty unequivocal on key issues surrounding the origins of SC2. However, this does not mean that there are no more questions to be asked or no more lessons to be learned on how to prevent SARS-CoV-3. That is why origin research continues, despite the “lab leak” question being mostly settled in the scientific literature.

## About scientific “controversies”

Whenever there is a (fake or real) controversy about a scientific topic, I strongly believe that only the fruits of scientific inquiry—not a toxic combination of politics, media, and activists—will produce reliable knowledge and factual insights to act upon.

All that is missing currently is for some citizen defenders of an evidence-based worldview to do the legwork of closing the rift between science and society, for example by building bridges of accessibility where there are none.

There is no infrastructure and no money in science communication, and science desks were the first to go in a crumbling newspaper economy. That is why I have to do this work without any financial aid or compensation (see *motivation statement*); and against a lot of headwinds from emotionally engaged conspiratorial communities.

To be completely honest, the lack of accessible information and resources on this topic is not only a monetary issue, **nor a coincidence in a broken information ecosystem.** Many virologists have been generous and great communicators trying to compensate for public knowledge gaps before they were publicly smeared and character assassinated into disengagement (or even defamed) by motivated actors. Anti-science pressure organizations like US-right-to-know (USRTK) target outspoken virologists explicitly for that reason.

“They extract a high cost for free speech, they coerce the informed into silence”

— Nature Magazine editorial about the anti-science pressure group USRTK

**I believe nobody in society should condone anti-science activism and abusive behavior.** It would surely help if citizens stood up for scientists unfairly targeted by politicians once in a while. That virologists are bitterly disappointed in institutions, media, and politics that facilitate such attacks is an understatement. In the end, science and democracy can not function without a society that supports them.

So I encourage you and other defenders of a “weight-of-evidence”-based worldview to talk about and speak up for the scientific method, use it to dispel popular myths, or **simply enjoy having understood something important about a dramatic world event** so you won’t be a sucker for the next manipulative “gain-of-function” charlatan that comes along.

## Motivation statement

As always, my hope and goals are to educate and equip citizens with conceptual tools and new perspectives to make sense of the world we inhabit.

This article took a lot of time and effort to conceptualize, research, and produce, actually almost irresponsibly so given that **I do not monetize my scicomm here**; and neither do the many scientists that so graciously gifted their time and expertise to help me and others understand their field of expertise.

---

*I see this work as a public good that I send out into the void of the internet in hopes others will get inspired to act*

---

## Copyright

You are also invited to deepen this work or just derive satisfaction from understanding our chaotic modern world a bit better.

So feel free to use, share, or build on top of this work, I just ask you to properly attribute (Creative Commons [CC-BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)).

---

### Cite this work:

**Markolin P., "Treacherous ancestry. A phylogeographic hunt for the ghosts of SARS-CoV-2", April 12, 2024.**

Free direct access link:  
[www.protagonist-science.com/p/treacherousancestry](https://www.protagonist-science.com/p/treacherousancestry)

*Also, since this topic is close to my heart, I'd be happy to hear your thoughts*

## References

Below I listed some of the key publications referenced in this article, with a short comment on what they are about for easy navigation and further reading.

### On recombination in CoVs:

[Sola I. et al., Annu Rev Virol., 2015](#) (recombination in CoVs is discontinuous with RdRp jumping around; deep mechanism of CoV transcription)

[Patino-Galindo JA. et al., Molecular Biology and Evolution, 2021](#) (recombination patterns and frequency between viral families)

[Boni MF. et al., Nature Microbiology, 2020](#) (recombination event between bats and pangolins not likely to have led to emergence of SARS-CoV-2, genetic elements of lineage have been circulating in bats for decades)

[Klerk A. et al, Virus Evol., 2022](#) (Conserved recombination patterns across and within coronavirus subgenera)

[Wells H. et al., Cell Host Microbe, 2023](#) (review article about CoV recombination, requirements and likely mechanisms)

[Wang J. et al, Nature communications, 2023](#) (virus-sharing networks between bats highlight ample opportunity for CoV recombination in nature)

**\*recombination as "virus sex" is not accurate mechanistically:** The exchange of genetic material between viruses is usually non-reciprocal, meaning the recipient of a genome portion does not act as donor of the replaced portion in the original source. In this respect, the term recombination does not have the same meaning in viruses that it does in diploid, sexually reproducing organisms wherein the exchange of genetic material between chromatids in the first meiosis division is reciprocal. ([Perez-Losada M., Infect Genet Evol., 2015](#))

### On the Furin-cleavage site

[Andersen KG., Nature Medicine, 2020](#) (proximal origin noticing unusual FCS and structural predictions of changed glycosylation patterns)

[Wu Y. et al, Stem Cell Research, 2021](#) (highlights FCS sites over wider CoV family)

[Lavie M. et al., J Virol., 2022](#) (cleavage at S1/S2 and S2 viral processivity studies highlight the function of the FCS)

[Jackson JB. et al., Nat Rev Mol Cell Biol., 2022](#) (FCS cleavage makes S1/S2 unstable, stabilizing mutation D614G. "This perplexing observation suggests that the acquisition of a furin-cleavage site by SARS-CoV-2 may have been a recent event.")

[Peacock T. et al., Nature Microbiology, 2022](#) (FCS cleavage of S1/S2 is essential in ferrets, pre-cleavage of the spike during viral egress enhances entry of progeny virions into TMPRSS2-expressing cells such as those abundant in respiratory tissue; and avoid endosomal IFITM proteins)

[Fraser BJ. et al., Nature, 2022](#) (Spike processing by TMPRSS2; cuts extracellular spike in 3 places including S1/S2 if not pre-cleaved by furin to enable membrane fusion)

[Garry R., PNAS, 2022](#) (FCS was not engineered and certainly not modeled after human EnAC)

[Chaudhry MZ. et al., Virology, 2022](#) (FCS gets quickly lost in cell culture because of kinetics)

[Vu MN. et al., PNAS, 2022](#) (synergistic interaction with the QTQTN motif proximal to the FCS plays a key role in infection and pathogenesis)

[Sander AL. et al., Communications Biology, 2022](#) (high sequence diversity in S1/S2 and a proto-FCS site found in European SARS-related bat CoVs)

[Alwine JC. et al., mSphere, 2023](#) (The FCS not lab adapted—initial SARS-CoV-2 isolates replicate poorly in traditional laboratory models— and not engineered—cleavage site loop length—best we can tell)

[Neil SD. Cell, 2023](#) (Furin cleavage sites in intermediate host animals, but not bat reservoirs)

[Steiner S. et al., Nature Reviews Microbiology, 2024](#) (recent review on viral entry and what is known about different mechanisms)

## References

### On the social lives of bats

Hughes AC. et al., Acta Chiropterologica, 2011 (Bat calls are distinct; rhinolophids use distinctive constant frequencies that can facilitate identification)

Vernes SC. et al., Phil. Trans. R. Soc. B., 2019 (vocalizations vary between species and include echolocation calls as well as social calls used either for parent–offspring reunions, territorial defence or maintaining group integration)

Irving A. et al., Nature, 2021 (bats are uniquely suited hosts and viral reservoirs)

Respicio JM. et al., Journal of Animal Ecology, 2024 (bat aggregation leads more negative and aggressive behavior)

### On Karst and caves

David Gillieson, Oxford Academic, 2005 (With a total area of about 400 000 km<sup>2</sup>, Southeast Asia contains some of the more extensive karst regions in the world. Many of these karst areas are of high relief with spectacular arrays of tower and cone karst. Many have now been inscribed on the World Heritage list in recognition of their unique geomorphology and biology)  
(500.000 km<sup>2</sup>)

### On pre-pandemic Sarbeco seroprevalence in South-East Asia

Wang N. et al., Virol Sin, 2018 (October 2015 serum sampling from 218 residents in four villages in Jinning County, Yunnan province, China, 6 (2.7%) positive SARSr serology)

Li H. et al., Biosafety and Health, 2019 (pre-pandemic: 9 cases (0.6%) of SARS-CoV-1 positive serology from a snowball-sampled cohort of 1600 people in 3 provinces in Southern China between 2015–2017)

Sanchez CA et al., Nature Communications, 2022 (65000 CoVs spillovers per year estimated)

Manning J. et al., Emerg. Inf Dis, 2022 (SC2 Elisa assays of pre-pandemic blood samples taken in Cambodia between 2005–2011 show 4–14% reactivity)

Evans TS. et al., Int J. Infect Dis, 2023 (neutralizing antibodies against various SC2r CoVs in Myanmar locals, high seroprevalence for rural (~20%) populations in 4 areas, 0% in city people, sampled pre-pandemic July 2017 to February 2020)

Meyer M. et al., Nature Communications, 2024 (bat diversity and disease relationship in response to human encroachment and habitat change)

### On sampling location, geographical distribution of SC2r genomes:

Li LL. et al., Emerg. Microb Infect. 2021 (RpPrC31, a viral recombinant ancestor, retrospectively discovered from bat samples in Yunnan)

Wacharaplusadee S. et al., Nature communications, 2021 (RacSC203, a cousin of SARS-2 discovered in Thailand containing S1/S2 insertion; also pangolin CoVs at wildlife checkpoint)

Delaune D. et al., Nature communications, 2021 (bat samples from 2010 cambodia find a close relative 92,6% of SARS-CoV-2 (RSHTT200), one bat was simultaneously co-infected with 4 viruses, 2 of them sarbecoviruses)

Lytras S. et al., Gen Evo., 2022 (recombination analysis and geographic distribution of Sarbecoviruses)

Temmam S. et al., Nature, 2022 (discovery of human-infectious SC2 relatives in Laos, “Banal” viruses and genome sequences)

Zhao S. et al., J Genet Genomics, 2022 (geogenomic distribution patterns)

Han Y. et al., Nature communications, 2023 (CoV sampling all over China)

Muylaert RL. et al., Nature communication, 2023 (landscape predictions and drivers of SC2 spillover)

Gilbert M. et al., Tropical Biomedicine, 2023 (SC2 related coronaviruses in Malaysia, no full genome sequencing yet but 99% similarity for a fragment)

Forero-Muñoz NR. et al., Virus evolution, 2024 (phylogeographic structure in bat hosts creates a landscape of selective pressure)

## References

### On moral panics

\*moral panic: As a distinct species of collective behavior, moral panics represent contentious and intensely affective campaigns to police the parameters of public knowledge and morality. As such, they are necessarily dependent upon and constituted by claims-making, with interested parties historically seeking to actuate alarm by influencing the imagery and representations of the mainstream press. (Walsh JP, *International Journal of Cultural Studies*, 2020)

### Recent preprints:

[Hassanin A. et al., biorxiv, 2023](#) (phylogenetic divergence of close SC2 relatives between sub-tropical Northern Vietnam and tropical Southern Vietnam)

[Crits-Christoph A. et al., biorxiv, 2023](#) (metatranscriptomics data analysis from Huanan market identifies wildlife species)

[Pekar et al., biorxiv, 2023](#) (geospatial history and origins of SARS-CoV-1 and SARS-CoV-2, video interview [here](#))

[Si J. et al., biorxiv, 2024](#) (multi-species ACE2 adaptiveness, Sarbecos can infect humans from the get-go)

### Other video interviews with origin researchers

Video: [Long-form discussion with bat ecologist Dr. Alice C. Hughes and the risk of SARS-CoV-3](#)

Video: [Long-form discussion with Dr. Michael Worobey and Dr. Kristian Anderson about their research establishing the Huanan market as the origin of the SARS-CoV-2 pandemic](#)

Video: [Long-form discussion with Dr. Angela Rasmussen and Dr Stephen Goldstein debunking common bad lab leak argument](#)

Video: [Long-form discussion with SAGO member Dr. Carlos Morel \(SAGO is the WHO's scientific advisory group on the origins of pandemics\)](#)

### Bonus

Wow, I am surprised you scrolled all the end to the references. This curiosity (or due diligence) is of course rewarded with a little bonus information:

I have been working on an origins book with a twist that I believe will make some waves. It raises questions nobody else has yet dared to tackle: Where exactly does the "lab leak" myth come from, how did it move through society, and why were so many citizens susceptible to it? From the remote villages of the Lahu mountain tribes via chaotic conspiracy theorists and needy journalists to the halls of power and even the US presidency, THE LAB-LEAK MYTH will invite readers into my most challenging project yet; making sense of where the current rift between science and society really comes from.

You can subscribe for free to keep following my science writing and receive further updates and announcements per email.

<https://www.protagonist-science.com/> (Substack newsletter)

You can also follow/engage on socials, preferably Bluesky @philippmarkolin.bsky.social, (still exist on Twitter as well but maybe not long)  
<https://medium.com/@protagonist-science>