Covid Origins Debate Day 2
Genetic clues

The Wuhan institute of Virology

Perception

Inaccessible fortress full of deadly viruses

vs.

Reality

Shi Zhengli, a few staff scientists, and some grad students

The lab leak theory says they secretly made SARS-CoV-2 in collaboration with Ecohealth Alliance and Tony Fauci.

I don't think that's true.

But if that is, here's what they got paid to do it:

Personnel	Hourly Rate	# Months	# Hours
Dr. Zhengli Shi (Co-Investigator)	$25.56	3.00	528
Dr. Peng Zhou (Senior Scientist)	$18.26	6.00	1056
Dr. Ben Hu (Research Fellow)	$10.95	3.00	528
Associate Professor	$13.69	6.00	1056
Senior Technician	$10.95	6.00	1056
Technician 1	$7.30	6.00	1056
Technician 2	$7.30	6.00	1056

Budget summary from the DEFUSE grant.

The full WIV is pretty large. Their website lists 46 research fellows, and the total lab staff is a few hundred people.

The Institute contains several research centers:

- Center for Emerging Infectious Disease
- Chinese Virus Resources and Bioinformatics Center
- Center of Applied and Environmental Microbiology
- Department of Analytical Biochemistry and Biotechnology
- Department of Molecular Virology

Shi Zhengli's group researching coronaviruses is just one smaller part of the lab.

Shi runs the "Research Group of Emerging Diseases".

Her lab did not do a lot of gain of function research. Going over the list of experiments, I count three since 2007.

Her group did do a lot of sampling trips, looking for bat viruses.

They took about 20,000 samples. That's not 20,000 coronaviruses.
It's about 2,000 samples with coronaviruses and 200 with sarbecoviruses.

This kind of sampling activity is quite safe – the samples tend to have enough genetic material to sequence but it's not usually enough to grow a virus from. Shi's group only ever managed to culture and grow 3 of those viruses.

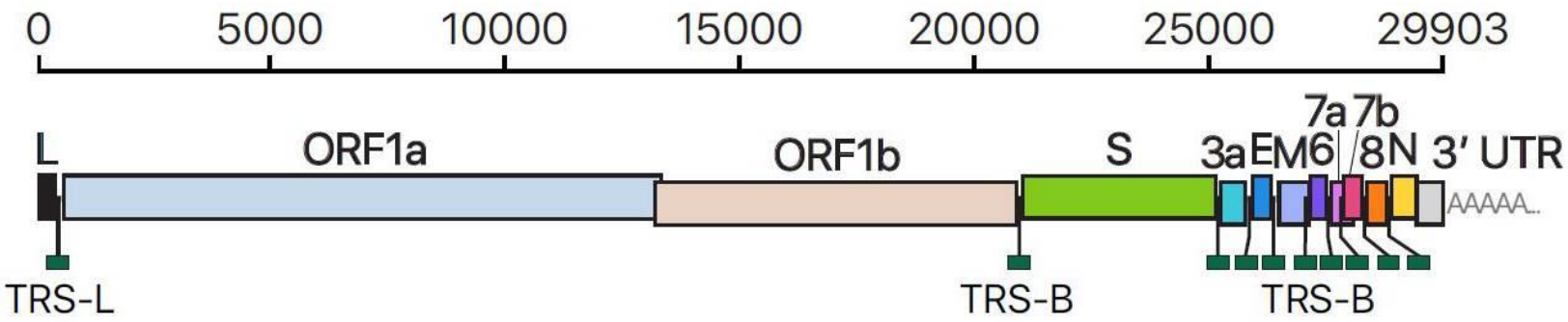
They have also recreated parts of other viruses using genetic techniques – for instance, in a 2017 experiment, they put the spike of 8 sequenced viruses into the backbone of another known virus.

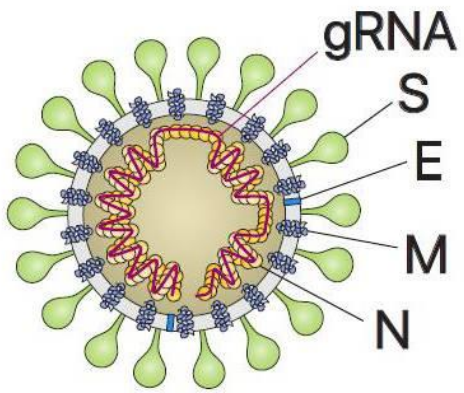
Let’s define some terms:

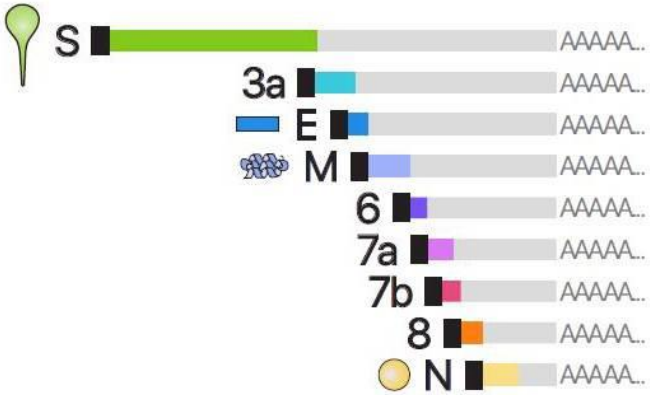
SARS-CoV-2 is an RNA virus, 30,000 nucleotides long.

Every 3 letters encodes an amino acid. A string of amino acids folds up into a protein.

Different sections of the genome encode for different proteins:







ORF1 makes non-structural proteins.

There are a few structural proteins, like spike, envelope, membrane, and nucleocapsid.

Let’s call everything besides spike the “backbone of the virus”.

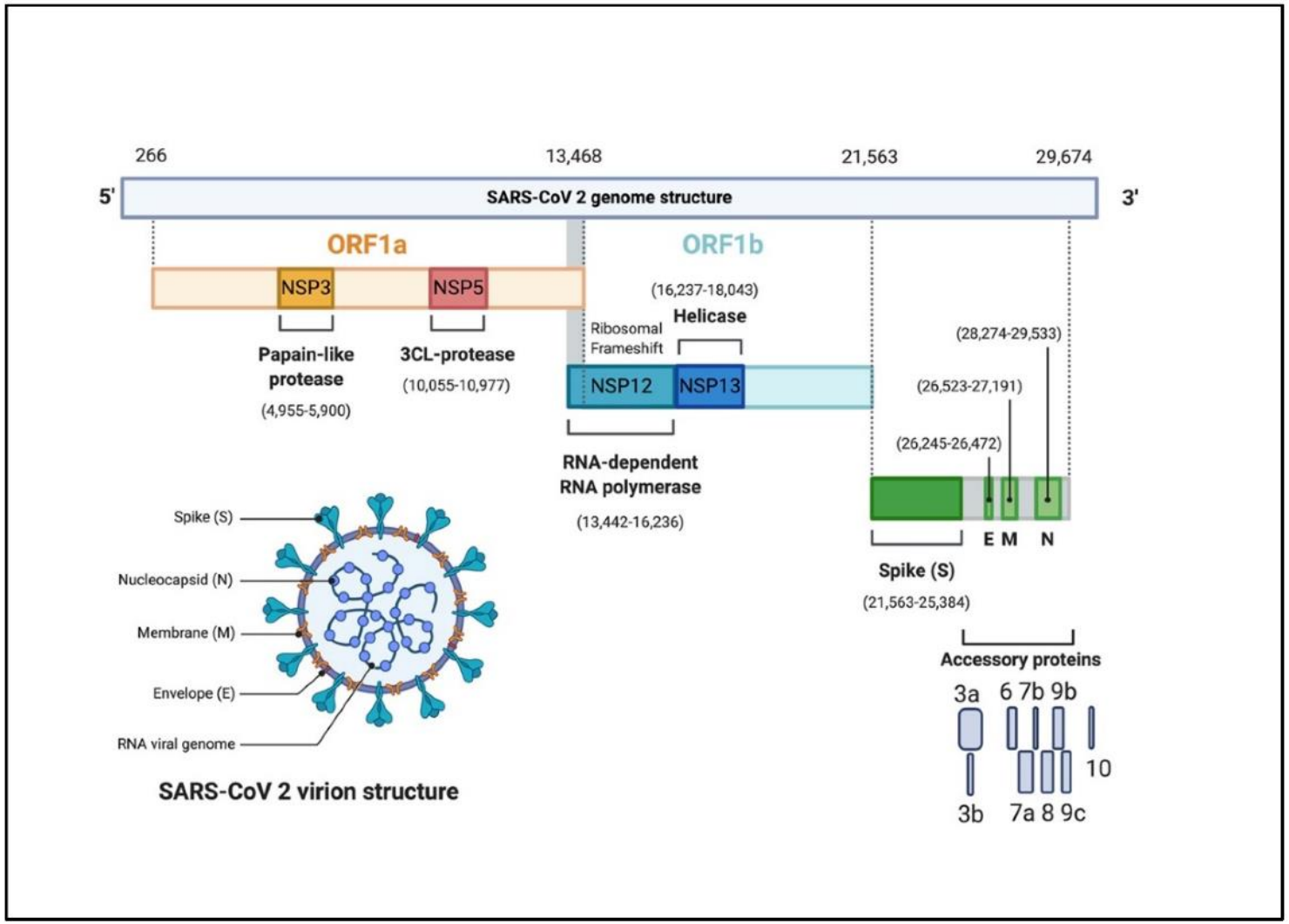


Figure from Alanagreh et al., 2020

Some virologists have put the spike of one virus into the backbone of another.

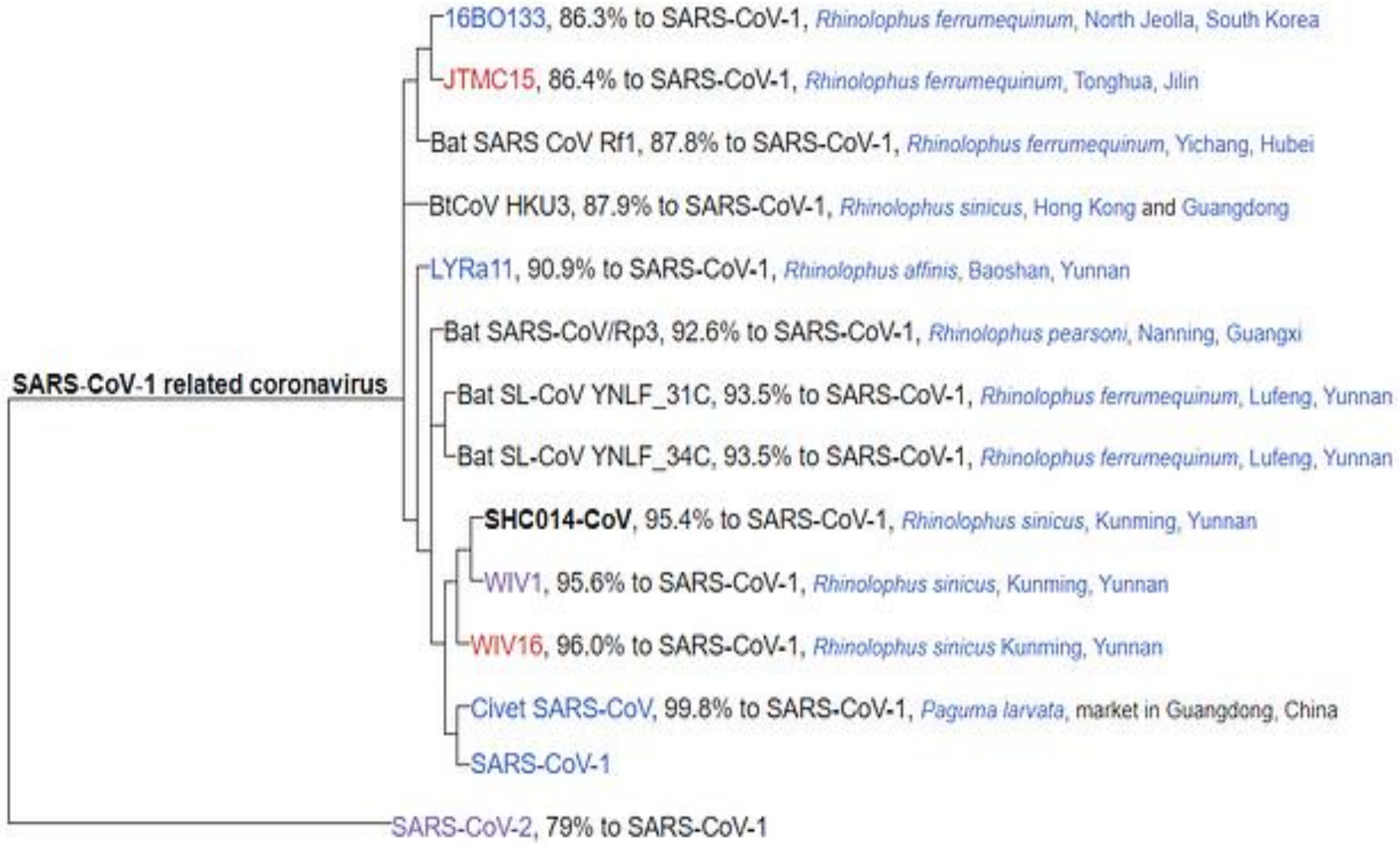
A 2015 experiment at the University of North Carolina made a chimera with a SARS backbone:

> The emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome (MERS)-CoV underscores the threat of cross-species transmission events leading to outbreaks in humans. Here we examine the disease potential of a SARS-like virus, SHC014-CoV, which is currently circulating in Chinese horseshoe bat populations[1]. Using the SARS-CoV reverse genetics system[2], we generated and characterized a chimeric virus expressing the spike of bat coronavirus SHC014 in a mouse-adapted SARS-CoV backbone. The results indicate that group 2b viruses encoding the SHC014 spike in a wild-type backbone can efficiently use multiple orthologs of the SARS receptor human angiotensin converting enzyme II (ACE2), replicate efficiently in primary human airway cells and achieve *in vitro* titers equivalent to epidemic strains of SARS-CoV.

A 2017 experiment at the Wuhan Institute of Virology made 8 chimeras with a WIV1 backbone:

> In the current study, we successfully cultured an additional novel SARSr-CoV Rs4874 from a single fecal sample using an optimized protocol and Vero E6 cells [17]. Its S protein shared 99.9% aa sequence identity with that of previously isolated WIV16 and it was identical to WIV16 in RBD. Using the reverse genetics technique we previously developed for WIV1 [23], we constructed a group of infectious bacterial artificial chromosome (BAC) clones with the backbone of WIV1 and variants of S genes from 8 different bat SARSr-CoVs. Only the infectious clones for Rs4231 and Rs7327 led to cytopathic effects in Vero E6 cells after transfection (S7 Fig). The other six strains with deletions in the RBD region, Rf4075, Rs4081, Rs4085, Rs4235, As6526 and Rp3 (S1 Fig) failed to be rescued, as no cytopathic effects was observed and viral replication cannot be detected by immunofluorescence assay in Vero E6 cells (S7 Fig).

These backbones were chosen because they're the viruses closest to SARS:



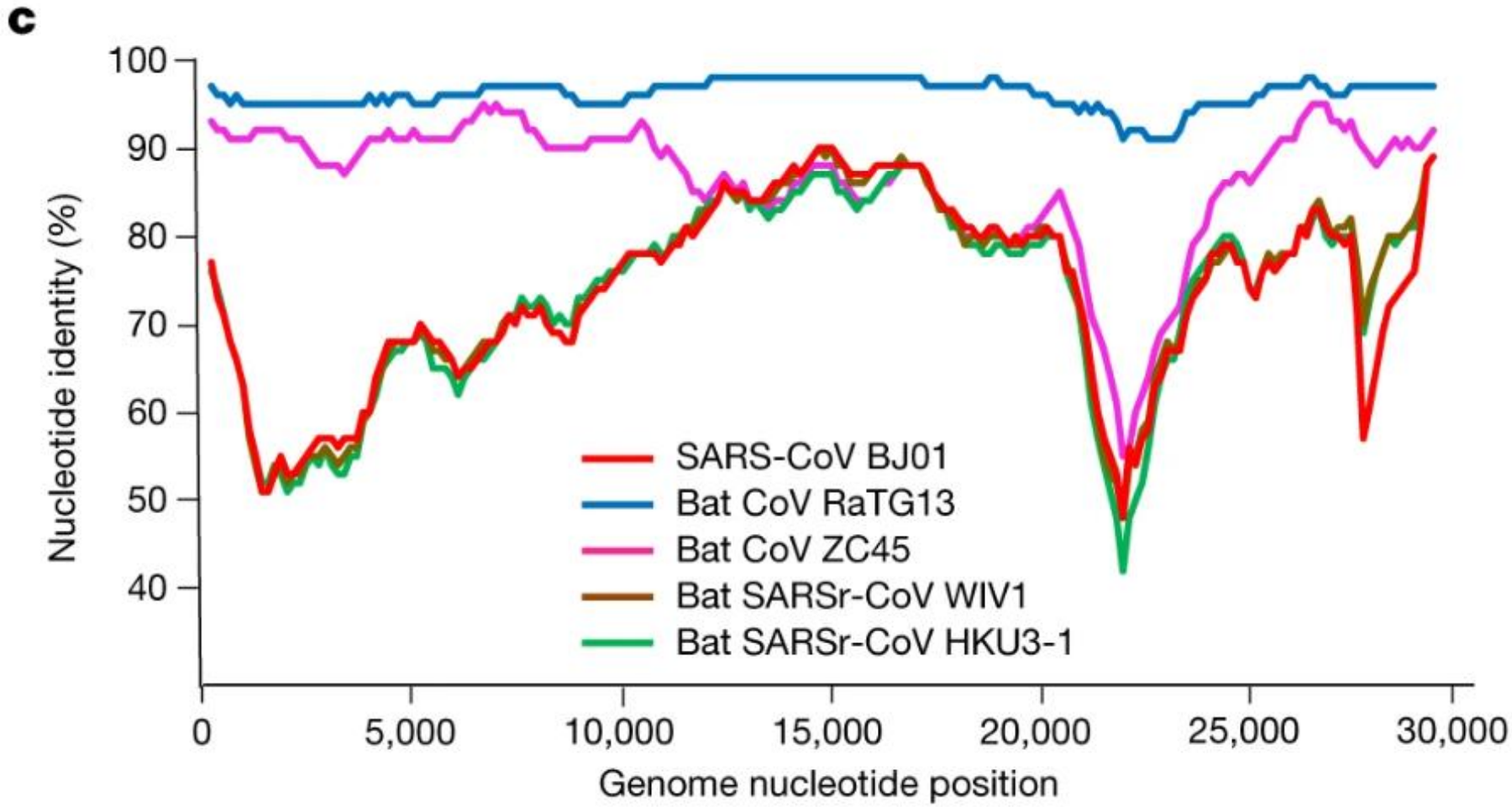
SARS virus family tree, [from Wikipedia]

A 2018 research proposal (the DEFUSE grant) proposed putting novel spikes into two of these viruses – WIV1 and SHC014. Those were also chosen because they're similar to SARS.

Technical Approach: Our goal is to defuse the potential for spillover of novel bat-origin high-zoonotic risk SARS-related coronaviruses in Asia. **In TA1** we will intensively sample bats at our field sites where we have identified high spillover risk SARSr-CoVs. We will sequence their spike proteins, reverse engineer them to conduct binding assays, and insert them into bat SARSr-CoV (WIV1, SHC014) backbones (these use bat-SARSr-CoV backbones, not SARS-CoV, and are exempt from dual-use and gain of function concerns) to infect humanized mice and assess capacity to cause SARS-like disease. Our modeling team will use these data to build **machine-learning genotype-phenotype models** of viral evolution and spillover risk. We will uniquely

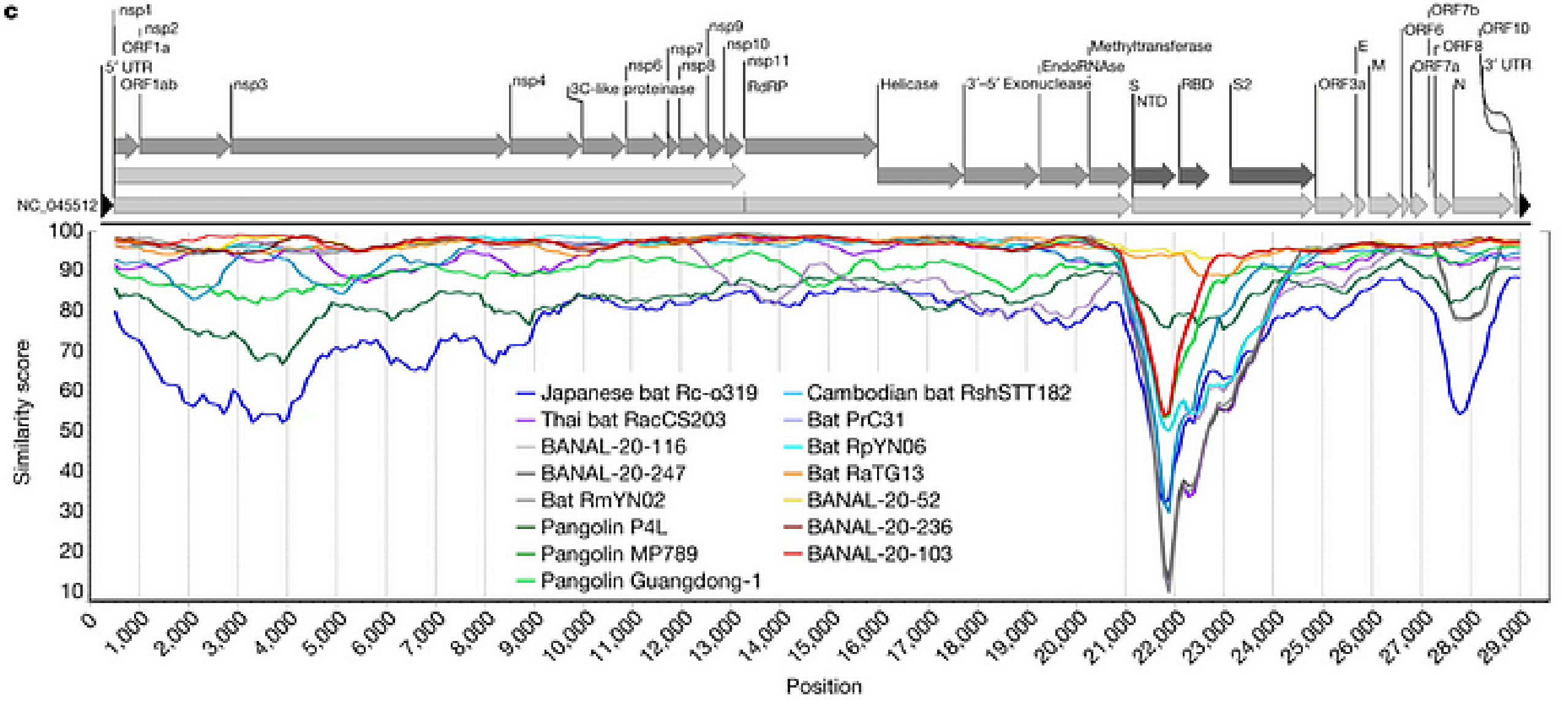
Which backbone was used to create SARS-CoV-2, if it was lab created?

Here's covid's genome, as compared to a few other viruses.



It's not close to WIV1, the virus used for previous experiments and mentioned in the DEFUSE grant.
Li Meng Yan said Covid was made from ZC45, it's still quite far from that.
Some other people say it was made from RATG13, which is closer.
Others say that it was made from some secret virus.

Two viruses are [closest to SARS-CoV-2](#).
RATG13 is 96.1% similar. It was found in 2013, mentioned in a few papers, and disclosed fully in January 2020.
BANAL-20-52 is 96.8% similar. It was found in Laos, after the pandemic started.
Neither one is close enough to create Covid.



RATG13

Soon after the pandemic started, Shi Zhengli's group disclosed a virus they had previously found, called RATG13, which was 96% similar to covid. There are ~1,200 mutations between the two viruses, or 40 years of evolution. This is not close enough to turn into covid, but it's still featured in many lab leak theories.

One theory says that RATG13 was used to create covid.
Another says the Wuhan lab used drugs to mutate it into covid.
Li Meng Yan called it fake. Other people also said it was fake.
One theory says it was passaged through animals to create covid.
There's also a theory that RATG13 was suspiciously renamed, from another virus in the database called BtCov/4991, to somehow hide where the virus was found.

RATG13 stands for **R**hinopholus **A**ffinis (the bat it was found in) **T**ong**G**uan (the place it was found) 20**13** (the year it was sampled). That should dispel the theory that it was renamed to hide something.

Most of these theories are strange — if the Wuhan lab used RATG13 to create covid, and was trying to hide that fact, why would they disclose RATG13 at all?

The lab would either have to use some very complicated process to turn RATG13 into Covid, or they would need to have other bat viruses that they never disclosed, and they altered one of those.

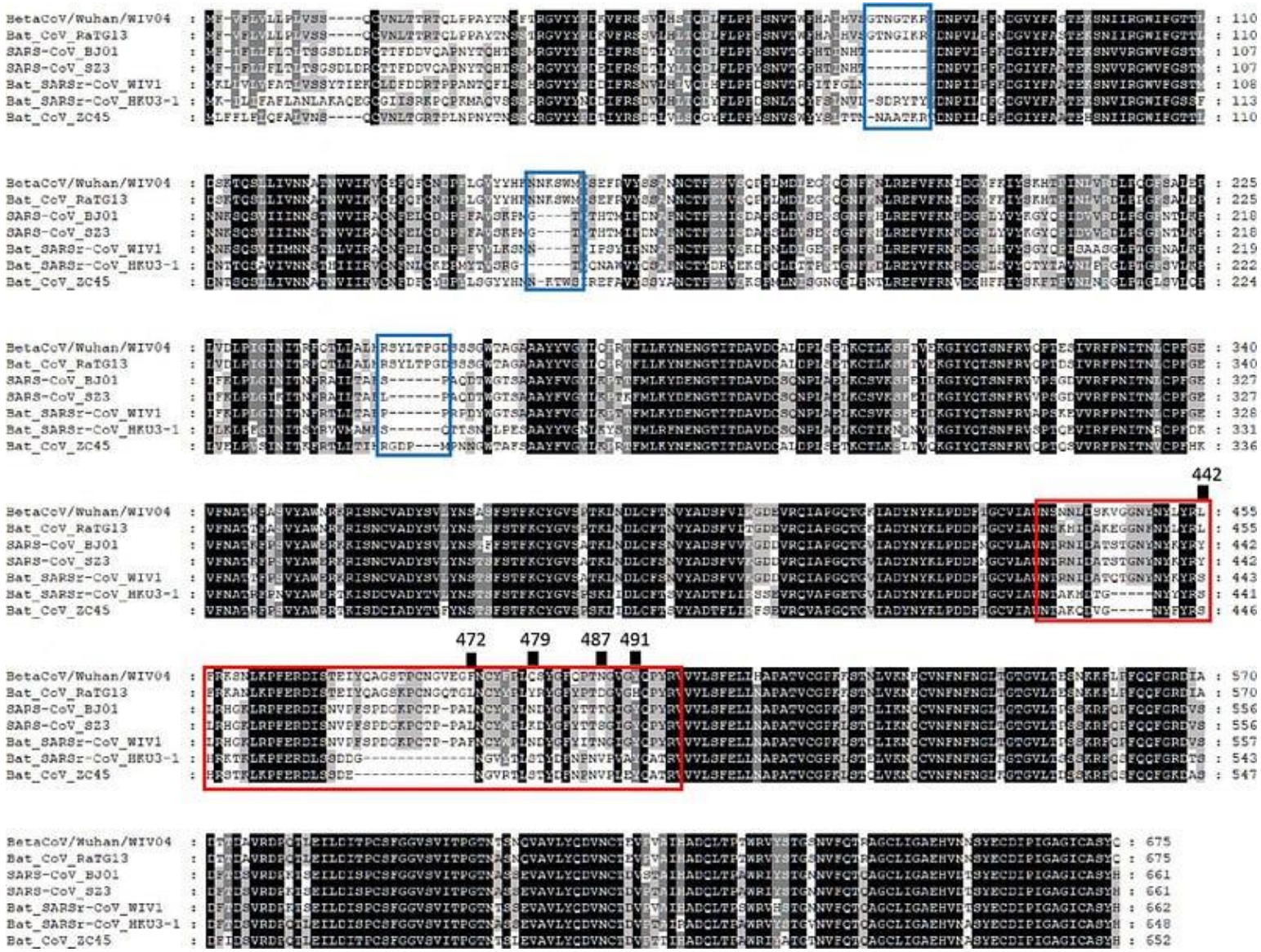
RATG13 was likely seen as an uninteresting virus, prior to the pandemic.

WIV lab director Wang Yanyi explains that RATG13 was not seen as relevant, as it was only 80% similar to SARS.

Linfa Wang says that no one would have been able to predict which virus to use, before the pandemic,
they would have started work with SARS or something close to SARS.

One lab leak theory claims RATG13 is real, but WIV scientists tried to hide the furin cleavage site in Covid when they disclosed it in the first paper, [submitted January 20th 2020](#).

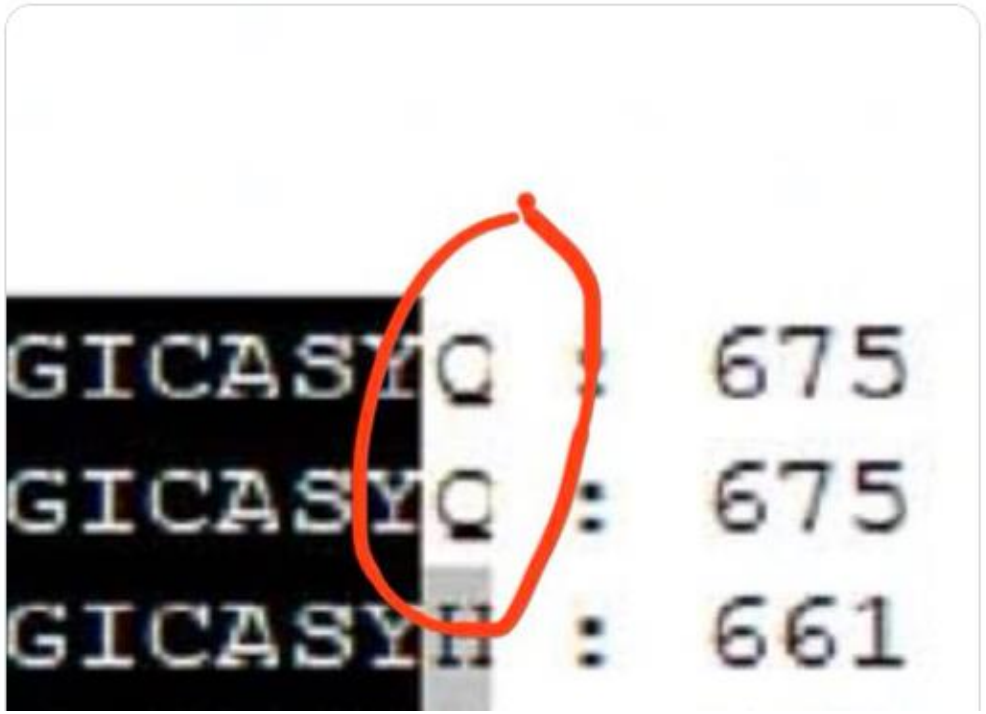
The paper shows a comparison of spike genes for RATG13 and SARS-CoV-2, but stops a bit shy of the furin cleavage site.



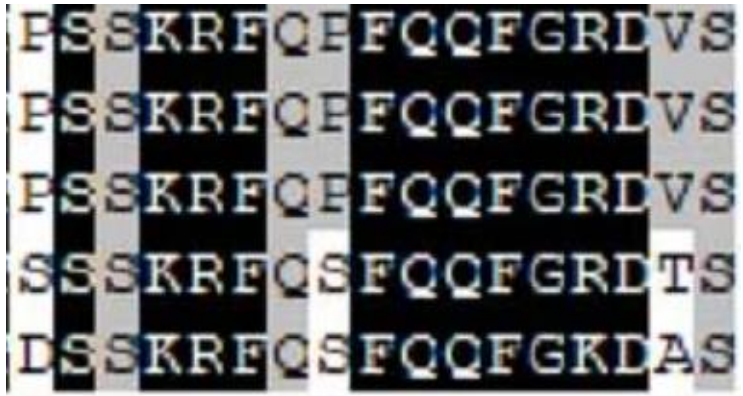
Yuri Deigin claims you can see where it was cut off:

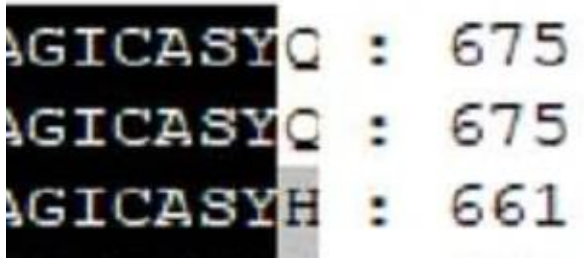
Yuri Deigin
@ydeigin

Also, you can see it was cut off quite sloppily by hand and whoever did that COULD NOT have missed the conspicuous PRRA insertion relative to RaTG13:



But the Q's, and some other letters are always cropped that way in the full image. That's just the font.



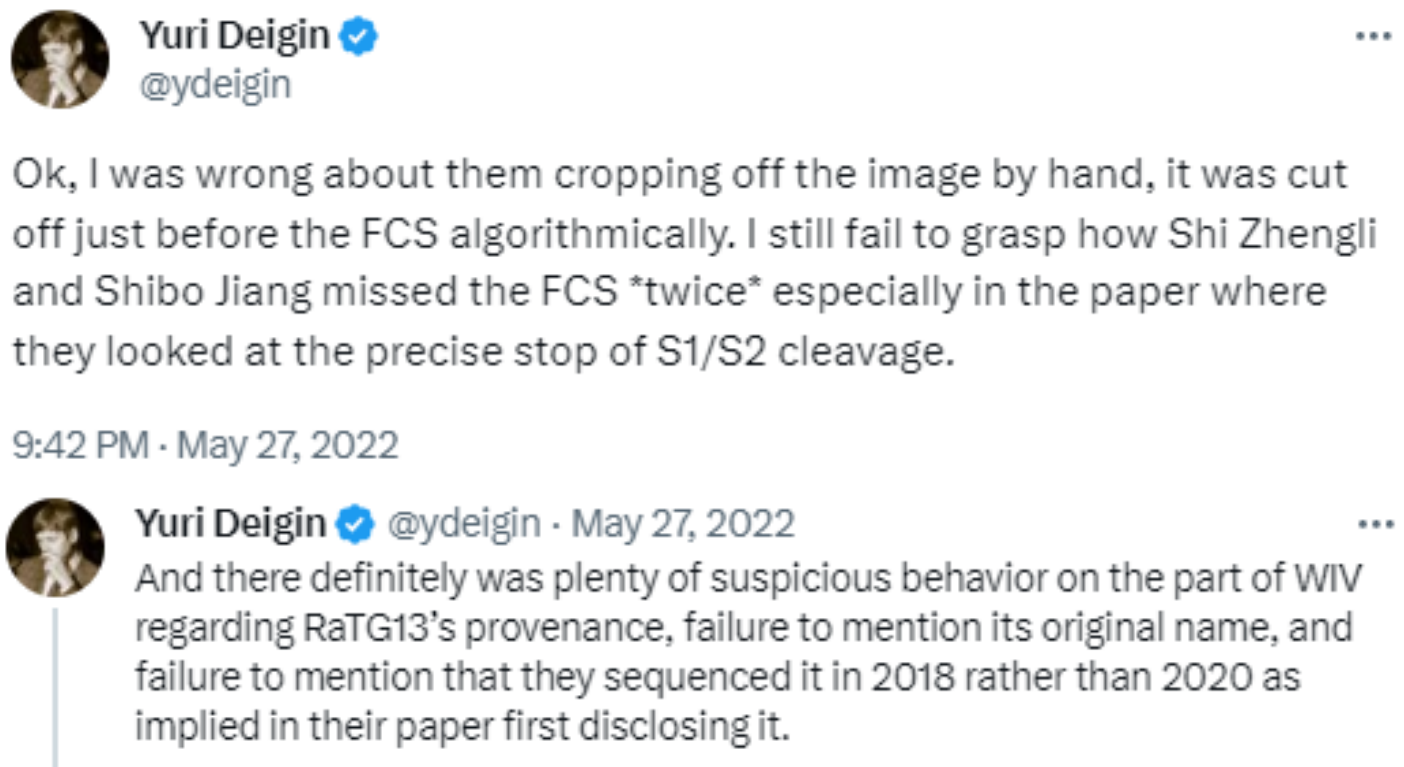


Also, it's not cut off right at the furin cleavage site, it's cut off 6 amino acids earlier.
Another scientist figured out which software was used to create this exact graphic and reproduced it.
Previous papers by this group (in 2017) compared viruses and also cut them off at this exact spot.
Also, a 2nd group who described the novel virus at the same time missed the furin cleavage site. So did a 3rd group.
Also, why would they even disclose RATG13 at all, if they'd used it to create Covid?

Yuri repeated this theory through 2020 and 2021.

I did find a 2022 tweet where Yuri admitted he was wrong about this.

But he just pivoted to saying that RATG13 was suspicious in other ways:



In 2020, Yuri talked about it being [suspicious that Ra4991 was renamed RATG13]:

In 2021, Yuri Deigin claimed that Ra4991 was not the same as RATG13, but that RATG13 is a version of 4991 which had gone through passaging in humanized mice.

Later in 2021, Yuri moved on to saying RATG13 is not the backbone, so none of that suspicion mattered at all.

Peter Jacobs @past_is_future · Sep 28, 2021

Oh Yuri already solved the backbone mystery a while ago. He declared unequivocally that it was RaTG13, likely with WIV inserted pangolin RBD in addition to a WIV inserted FCS. Open and shut. Who are we to doubt Yuri?

Yuri Deigin
@ydeigin

Dude, you are quoting my Medium article from April 2020. Since then I have said many times that RaTG13 itself is not the backbone. Please do some actual research on the topic before embarrassing yourself 😂

12:58 PM · Sep 28, 2021

← He's admitting here that RATG13 is irrelevant, but he's still talking about that suspiciously cut-off diagram 2 months later.

By 2022, he's decided to just use both theories:

Yuri Deigin
@ydeigin

I think the

"let's insert an FCS into these CoVs we just got from Laos/Yunnan in 2018/19 like we just proposed in DEFUSE"

hypothesis of SARS2 origin fits better the known facts than

"let's passage the Mojiang miners virus we got in 2013"

but I don't feel strongly about it.

11:42 AM · Apr 15, 2022

It was later proven by Flo Débarre and @Etaitlife that 4991 is the same virus as RATG13.

This was also confirmed by a later revealed 2018 copy of the Ra4991 genome.

Ra4991 was also renamed multiple times, there's nothing suspicious about that:

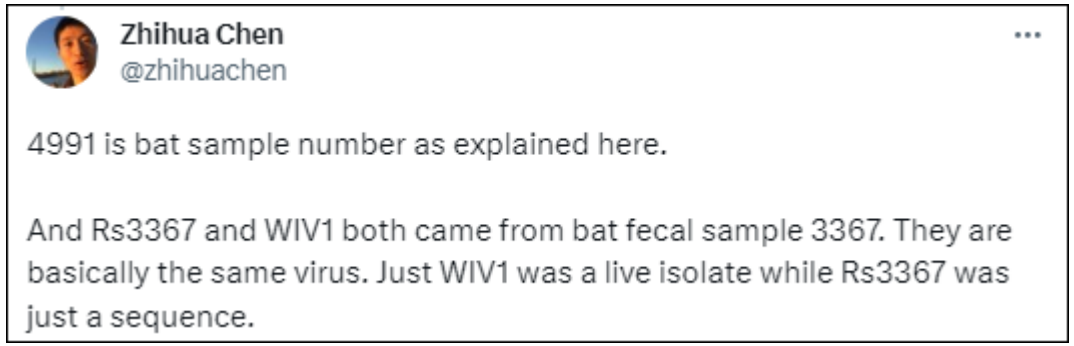
Zhihua Chen
@zhihuachen

RaTG13 names

4991-NP, Wang Ning (2014 master's thesis in Chinese, 4991 only showing up in one figure)
RaBt-CoV/4991 Ge XY, Wang Ning (2016 Virological Sinica in English)
Ra4991_yunnan Yu Ping (2019 master's thesis in Chinese)
RaTG13, Peng Zhou et al, (2020 Nature in English)

11:55 AM · Jun 22, 2021

The lab also renamed other viruses before, that's what They did every time they found an important virus:



Zhihua Chen
@zhihuachen

4991 is bat sample number as explained here.

And Rs3367 and WIV1 both came from bat fecal sample 3367. They are basically the same virus. Just WIV1 was a live isolate while Rs3367 was just a sequence.

The Mojiang mine also wasn't mentioned in other papers about 4991:

Zhihua Chen
@zhihuachen

Well hold on. For someone who's seemingly suspicious of the fact that Zhou et al never mentioned the mine, the miners, or cited the 2016 paper, you seem completely uninterested in the fact that Yu Ping never mentioned the mine, the miners, or cited the 2016 paper! Shouldn't we ...

4:21 AM · Sep 20, 2022

Zhihua Chen @zhihuachen · Sep 20, 2022
... we try to figure out the conspiracy that's required for Yu Ping to do that, before diving into other rabbit holes?

2

The Trump state department fact sheet also mentioned RATG13, after telling the 3 sick researchers story.

> •The U.S. government has reason to believe that several researchers inside the WIV became sick in autumn 2019, before the first identified case of the outbreak, with symptoms consistent with both COVID-19 and common seasonal illnesses. This raises questions about the credibility of WIV senior researcher Shi Zhengli's public claim that there was "zero infection" among the WIV's staff and students of SARS-CoV-2 or SARS-related viruses.

The fact sheet claims the lab "conducted experiments involving RATG13", which is probably not true, the only thing we know they did is sequence it:

> •Starting in at least 2016 – and with no indication of a stop prior to the COVID-19 outbreak – WIV researchers conducted experiments involving RaTG13, the bat coronavirus identified by the WIV in January 2020 as its closest sample to SARS-CoV-2 (96.2% similar). The WIV became a focal point for international coronavirus research after the 2003 SARS outbreak and has since studied animals including mice, bats, and pangolins.
> •The WIV has a published record of conducting "gain-of-function" research to engineer chimeric viruses. But the WIV has not been transparent or consistent about its record of studying viruses most similar to the COVID-19 virus, including "RaTG13," which it sampled from a cave in Yunnan Province in 2013 after several miners died of SARS-like illness.
> •WHO investigators must have access to the records of the WIV's work on bat and other coronaviruses before the COVID-19 outbreak. As part of a thorough inquiry, they must have a full accounting of why the WIV altered and then removed online records of its work with RaTG13 and other viruses.

You could not passage RATG13 to become SARS-CoV-2

Just to get that number of mutations, it would take 15 years of passaging.

And even if you did that, it wouldn't give the right kind of directed evolution,
it would be adapted to cell culture, and it wouldn't match the recombinant history.

> *"The Wuhan Institute of Virology (WIV) group would have needed to passage it in cells or animals for years to accumulate 3.8% sequence divergence.*
>
> *For example, the mutation rate of SARS-CoV during passages in cell cultures was found to be 9×10^{-7} substitutions per nucleotide per replication cycle (approximately 12 h). Serial passage of SARS-CoV in animals resulted in comparable numbers. Following cultivation in mouse lungs for more than 30 days, the coronavirus accumulated only six nucleotide mutations (the divergence of 0.02%) after 15 passages. Based on these mutation rate estimates, the accumulation of 3.8% genetic difference via cell or animal passage would require more than 15 years. It is fair to assume that SARS-CoV-2 has similar mutation rates. Therefore, given that the RaTG13 virus was discovered in 2013, the accumulation of 3.8% differences in this coronavirus by 2019 seems improbable."*

Quote from Tyshkovskiy and Panchin, 2021

If you used some kind of drugs to accelerate the rate of mutation, that would leave some signature.

SARS-CoV-2 has the same profile of mutations relative to RATG13 as SARS does compared to a close bat virus

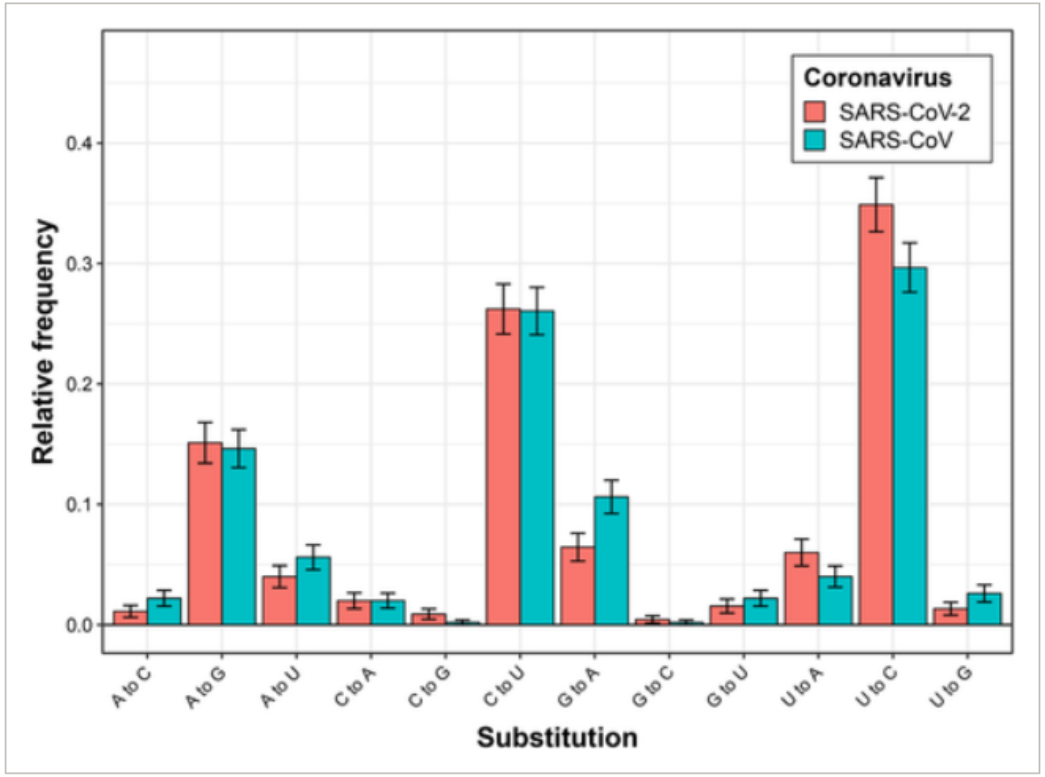


FIGURE 1 Open in figure viewer | PowerPoint

Relative frequencies of different single nucleotide substitutions, which distinguish SARS-CoV-2 (red) and SARS-CoV (blue) from their bat relatives (RaTG13 and Rs4231, respectively).[7] Differences across substitution frequencies are not significant, as assessed with Pearson's chi-squared test ($p = 0.12$)

Figure from Tyshkovskiy and Panchin, 2021

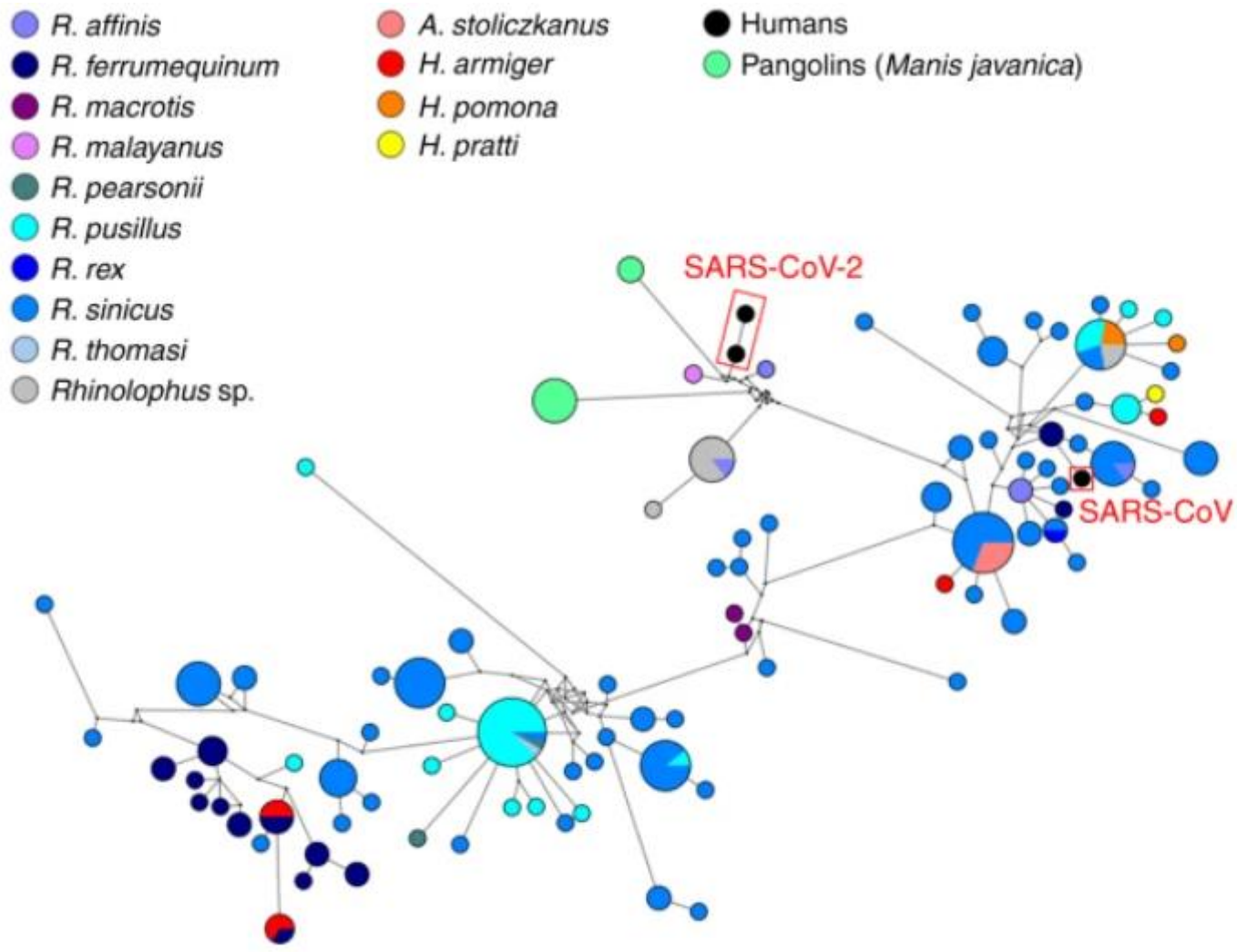
Did the WIV have undisclosed viruses closer than RATG13?

Ecohealth [disclosed 200 sarbecoviruses in 2020]():

Article | Open Access | Published: 25 August 2020

Origin and cross-species transmission of bat coronaviruses in China

Alice Latinne, Ben Hu, Kevin J. Olival, Guangjian Zhu, Libiao Zhang, Hongying Li, Aleksei A. Chmura, Hume E. Field, Carlos Zambrana-Torrelio, Jonathan H. Epstein, Bei Li, Wei Zhang, Lin-Fa Wang, Zheng-Li Shi & Peter Daszak



Maximum clade credibility tree (**a**) including 202 *RdRp* sequences from the *Sarbecovirus* subgenus

This paper was published in August 2020 but submitted in August 2019, before the start of the pandemic.

The authors should have originally had no reason to hide any viruses, back in August 2019.
Lab leak theorists thought maybe one key virus was removed during peer review, that could be the source of Covid.
A FOIA request got an [original copy of the paper](). Nothing had been changed, no viruses had been removed.

Alina Chan @Ayjchan · May 12, 2021

Honestly, I'm very curious to see the first version of this Latinne et al. manuscript that was sent to @NatureComms and the peer review that it went through.

A preprint was posted in May 2020, the paper was submitted Oct 2019 and published in August 2020.

8 15 81

Richard H. Ebright @R_H_Ebright · Feb 6

Replying to @Ayjchan and @NatureComms

I too am eager to see the original version. Especilly to see one particular sequence that was in the original version submitted before emergence of the pandemic , but was deleted from the version published after emergence of the pandemic,

1 8 47 2,153

https://twitter.com/R_H_Ebright/status/1622447800918343680?s=20&t=Rv8LGq7kpGJOB6Won6p4gQ

Francisco de Asis @franciscodeasis

[Thread] FOIA from @USRightToKnow regarding Latinne et al. (2020) and clade 7896
TLDR: No sequence was deleted/modified since Aug-2019, but it seems they wanted to buy time for not publishing the viruses very early in the pandemic.

usrtk.org
FOI documents on origins of Covid-19, gain-of-function research and biolabs...
Public records obtained by U.S. Right to Know from our investigation into the origins of Covid-19 and biolab safety

2:55 PM · Oct 12, 2021

Also, Francisco's tweet is misleading – Ben Hu actually [asked for the sequences to be made public]() earlier than planned.

There's only a very short window, between August 2019 and November 2019, where the lab could have discovered a new virus and started experiments with it.

There was also a theory that the WIV was hiding 8 viruses from the mineshaft where RATG13 was discovered.

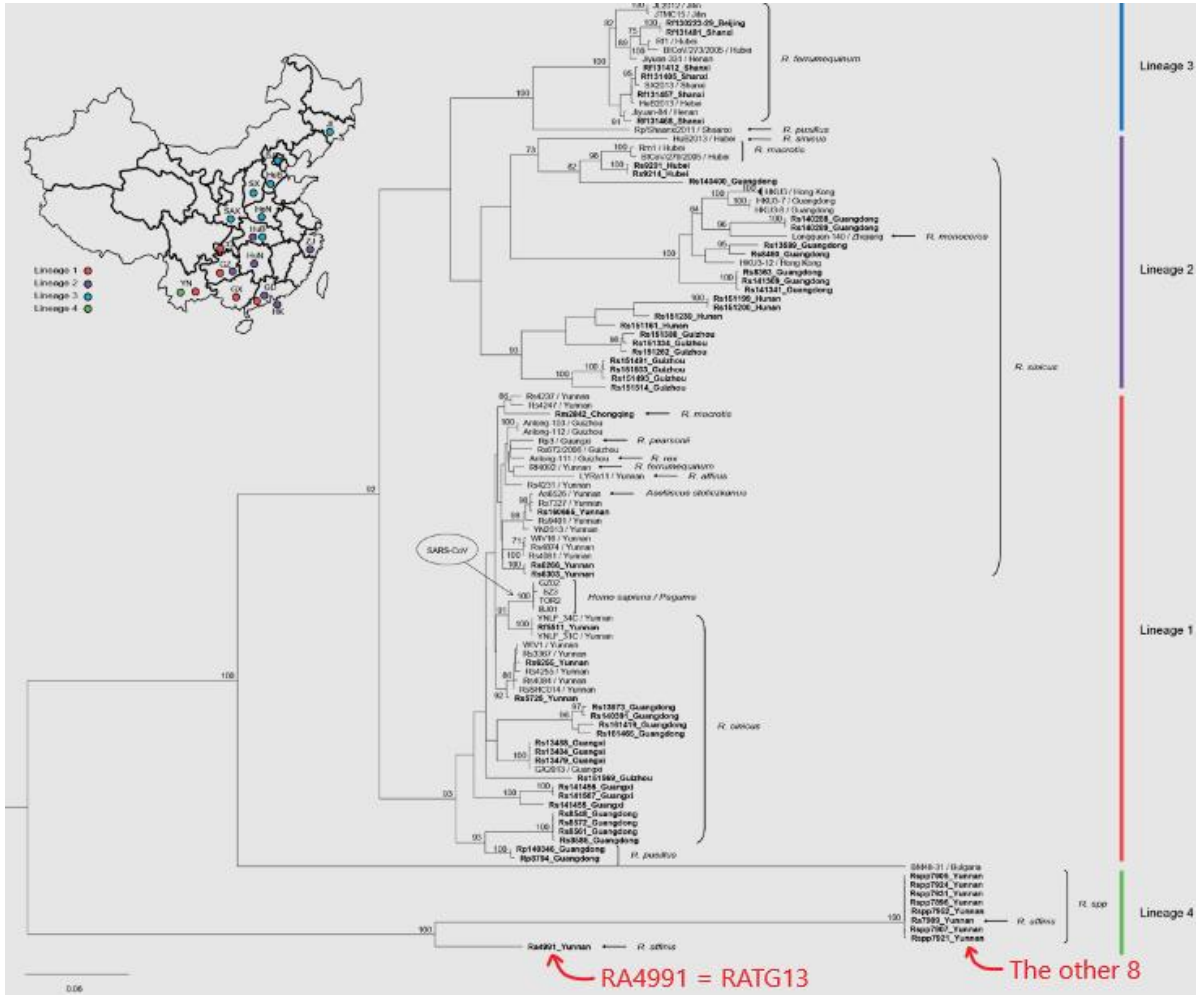
Here's one mention, from a paper by Yuri Deigin and Rossana Segreto:

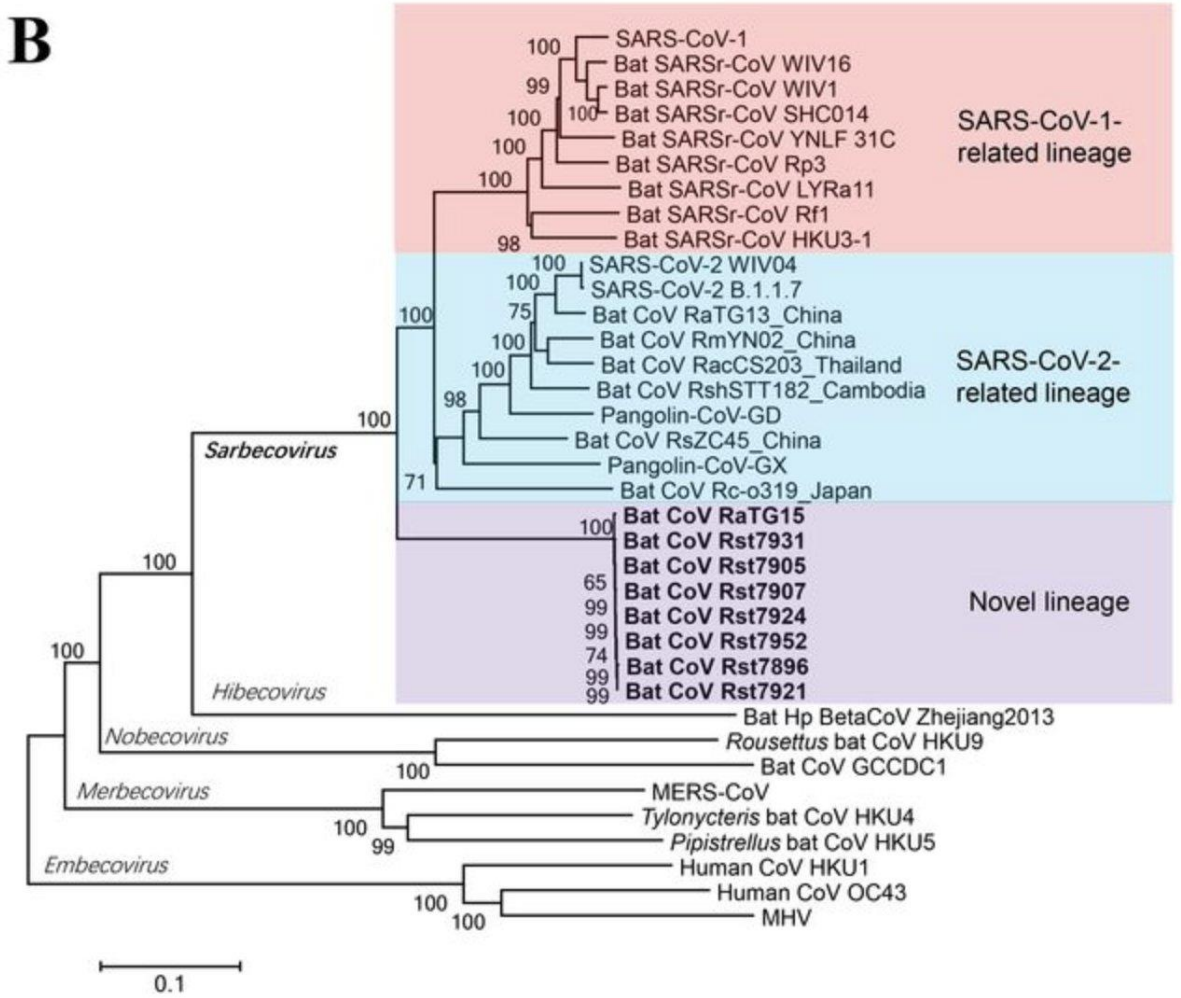
> In addition, new information revealed by the Addendum is that eight other beta-SARSr-CoVs distantly related to SARS-CoV were also isolated from the same Mojiang mine, and sequenced together with RaTG13, but neither their genomes, nor information about their sample names and eventual accession numbers is provided. It is not known how these sequences relate to RaTG13.

Those viruses were disclosed in 2021. They weren't very closely related to SARS-CoV-2.

The delay is because Shi Zhengli's group was writing a paper about them.

There's also an unpublished paper from 2018 with access to all these viruses.
Genomes were submitted to Genbank with a 4 year data embargo.
3 journals didn't take the paper, so everyone forgot about it. In 2022, the genomes were released automatically.
It confirms that RATG13 is real, the other 8 are uninteresting, and there were no other viruses similar to Covid in 2018.
There could still be secret viruses, but there would have been no need to hide anything, prior to the pandemic.





2020 phylogeny showing RATG13 and the other 8 (in purple)

This has now become a zombie talking point:

Alex Washburne
@WashburneAlex

The >50 novel strains Daszak claims to have discovered in 2019...

If SARS-CoV-2 were truly zoonotic, Daszak
& the WIV would know it and could share these sequences to exonerate themselves.

They haven't shared these sequences. The progenitor to SARS-CoV-2 is likely among them.

> **Rebecca** @Rebecca21951651 · Apr 2
> Replying to @emilyakopp and @BiophysicsFL
> Unfortunately not all >50 novel SARS CoVs strains discovered (as of Nov 21 2019) were published in Nature:
>
> **Peter Daszak**
> @PeterDaszak
>
> Not true - we've made great progress with bat SARS-related CoVs, ID'ing >50 novel strains, sequencing spike protein genes, ID'ing ones that bind to human cells, using recombinant viruses/humanized mice to see SARS-like signs, and showing some don't respond to MAbs, vaccines...
>
> > **Andrew Rambaut** @arambaut · Nov 21, 2019
> > Replying to @PeterDaszak @GlobalVirome and 2 others
> > The more we look the more new viruses we find. The problem is that we have no way of knowing which may be important or which may emerge. There is basically nothing we can do with that information to prevent or mitigate epidemics. nature.com/articles/d4158...
>
> 10:42 PM · Nov 21, 2019 from Manhattan, NY · Twitter for iPhone
>
> **2** Retweets **13** Likes

Peter Daszak @PeterDaszak · Apr 2
Replying to @CadhlaFirth and @WashburneAlex
Dr. @WashburneAlex, Cadhla is correct: every single one of the **SARSr-CoV** sequences @EcoHealthNYC discovered in China is already published, incl. "50 strains" from a 2019 tweet you mentioned above. Details in Latinne et al. & the following thread.

> **Peter Daszak** @PeterDaszak · Mar 31
> Replying to @lab_leak @franciscodeasis and 3 others
> Dear Dr. "lab leak", the answer is "yes": literally 100s of novel sequences published in 2020 paper linked here. We then published a handful of remaining SARSr-CoV RdRp in Genbank as shown in response to NIH letter, which was also made public by FoIA. nature.com/articles/s4146...
>
> 
> ve have provided below within the 5-day time limit given to us by Dr. Tabak
> ttempt to supply available data as quickly as we are able. These include:
> etic tree of two new SARSr-CoV RdRp sequences reported in ***Figure 5*** *from*
> se are from *Rhinolophus sinicus* bats sampled in Hubei and Yunnan provinc
> respectively, in 20
> below clearly dem
> they these viruses
> red) are not relate
> SARS-CoV-2, and t
> recently-describec
> coronaviruses are
> relatives of the pa
> The RdRp sequenc
> two viruses are no
> through approval
> Chinese authoritie
> can be uploaded b
> the earliest possib
> opportunity.
> quested that the 13 other novel RdRp sequences (in bold) included in Figur
> year 5 report to be uploaded to Genb
> now going through the approval proc
> All highlighted CoV RdRp sequences shown in Figure 2 have been uploaded to GenBank. Their accession numbers are given below:
> • DA180047: GenBank Accession No. OK382029
> • DA180043: GenBank Accession No. OK382030
> • DA180030: GenBank Accession No. OK382031
> • CA180419: GenBank Accession No. OK663604
> • DA180058: GenBank Accession No. OK663605
> • LA180233: GenBank Accession No. OK663606
> • LA180255: GenBank Accession No. OK663607
> • CA180367: GenBank Accession No. OK663608
> • CA180305: GenBank Accession No. OK663609
> • CA180390: GenBank Accession No. OK663610
> • LA180234: GenBank Accession No. OK663611
> • LA180239: GenBank Accession No. OK663612
> • LA180240: GenBank Accession No. OK663613
> • AA180355: GenBank Accession No. OK663614
> • DA180057: GenBank Accession No. OK663615

3 2 6 5,066

This talking point even gets repeated by mainstream journalists:

Either:

Daszak is lying, and he knows which secret virus was used to create Covid-19.

Daszak is telling the truth, but the WIV secretly has more viruses that he doesn't know about.

Daszak is telling the truth and the lab doesn't have any secret viruses. But in that case, there's no way to exonerate himself. He's already shared everything but lab leak theorists don't believe him.

Peter Daszak interview, December 9th, 2019

He talked about manipulating coronaviruses and inserting the spike of one virus into another.

If there was a lab leak before this, Daszak does not know about it. He wouldn't still be talking about this work.

January 2nd and January 11th he's still talking about infecting humanized mice with SARS viruses.
The Covid genome was released to the world on January 10th.
If Daszak knows about a lab leak, or knows how the virus was made, he's doing a very bad job at hiding it.

Peter Daszak @PeterDaszak · Dec 31, 2019
...the China CDC & Provincial CDCs are working effectively already, and there is an openness and transparency right now that wasn't there during the first SARS cases. Also no sign of wider outbreak yet....Let's hope this is not a novel viral agent, and that it resolves rapidly!
3 8 16

Peter Daszak @PeterDaszak · Jan 2, 2020
Some background on the risk of re-emergence of #SARS-coronavirus or a novel SARS-related CoV. @EcoHealthNYC has been working since 2004 in China to trace back origins of SARS-CoV... @ProMED_mail @nycbat @hongying_li @TheMenacherryLab @Laurie_Garrett
1 7 10

Peter Daszak @PeterDaszak · Jan 2, 2020
We showed it likely originated in bats & that there's a large SARSr-CoV diversity in bats in China, see papers: ecohealthalliance.org/wp-content/upl... ecohealthalliance.org/wp-content/upl...
2 4 7

Peter Daszak @PeterDaszak · Jan 2, 2020
...We isolated SARSr-CoVs that bind to human cells in the lab & @Baric_Lab @TheMenacheryLab + others showed some of these have pandemic potential, able to infect humanized mice & not protected by candidate SARS vaccine or Monoclonal therapeutics. ecohealthalliance.org/wp-content/upl...
2 8 8

Peter Daszak @PeterDaszak · Jan 2, 2020
We worked with Wuhan Institute of Virology (Zhengli Shi, Ben Hu), @dukenus (Linfa Wang) to design serol tests, and showed that people in rural China are exposed to #bat coronaviruses...ecohealthalliance.org/wp-content/upl...

Peter Daszak @PeterDaszak · Jan 11, 2020
This phylogenetic tree verifies the information we were getting out of China at on New Year's Eve, that the outbreak is caused by a novel #coronavirus approx. 80% similar to SARS....
2 8

Peter Daszak @PeterDaszak · Jan 11, 2020
..It's supports the hypothesis we put forward in our new @NIAIDNews grant with @Baric_Lab, Linfa Wang & Dani Anderson of @dukeNUS Zhengli Shi at Wuhan Inst. Virol & others, that SARSr-CoVs that are more than 10% divergent in the RBD domain are a high risk for emergence because.
2 2 3

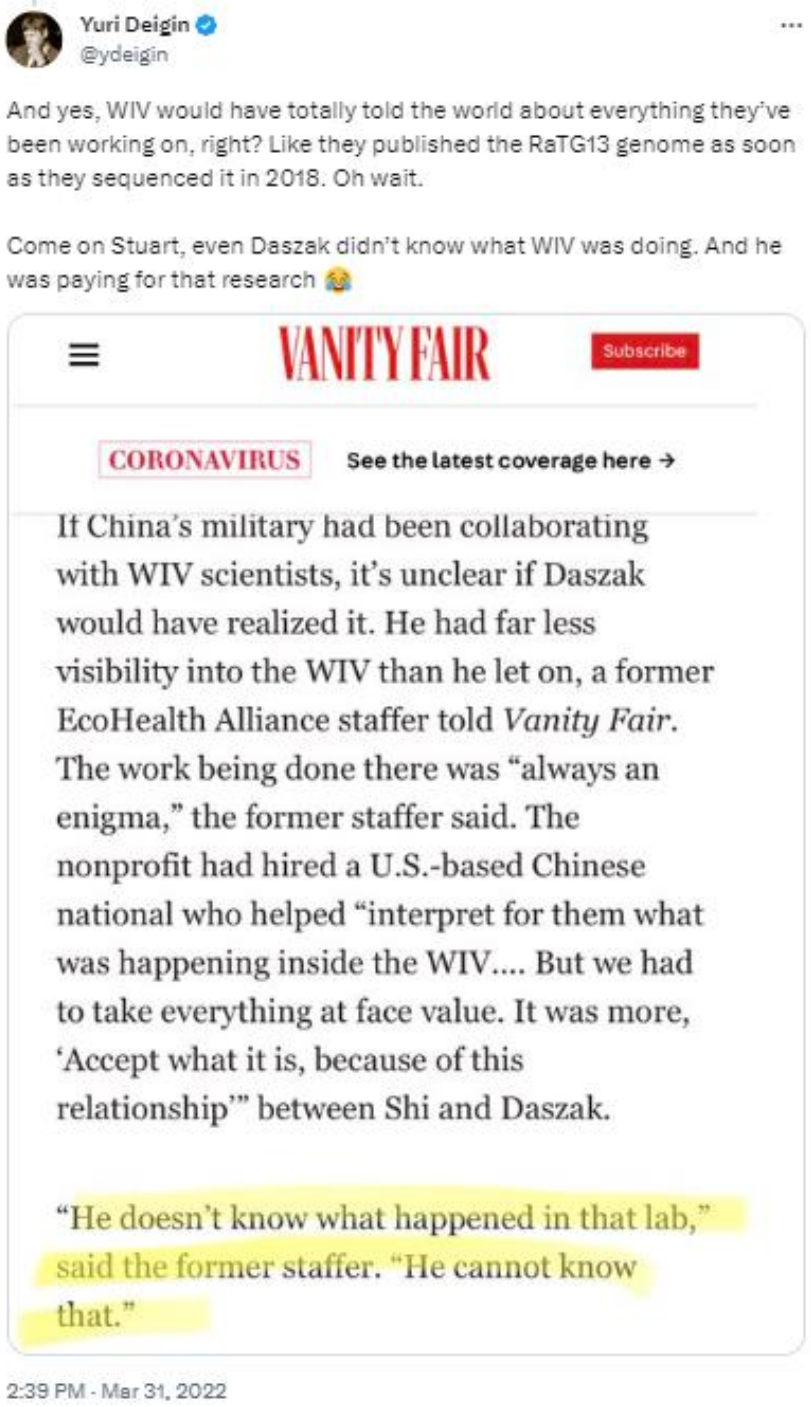
Peter Daszak @PeterDaszak · Jan 11, 2020
...These are sufficiently close to SARSr-CoVs that we've shown can infect human cells, cause SARS-like illness in humanized mice, and are not treatable with most of the monoclonal therapeutic candidates, nor preventable with the SARS-CoV based vaccine candidate, and...
2 2 2

Peter Daszak @PeterDaszak · Jan 11, 2020
...because this is more divergent, it means that any potential therapies currently being developed may not be able to work against this virus. This clade of SARSr-CoVs is a clear and present danger for pandemic emergence!
1 1 5

By January 25th, lab leak rumors have already started, someone asks Peter Daszak about them and he's still talking about the 50 similar coronaviruses and chimeric lab-made viruses:

If these "50 viruses" were proof that Daszak was hiding the source of the pandemic, it probably wouldn't be the first thing he mentioned, when asked about a lab leak…

Even DRASTIC [can't decide](#) whether Daszak is in on the conspiracy:



Yuri Deigin
@ydeigin

And yes, WIV would have totally told the world about everything they've been working on, right? Like they published the RaTG13 genome as soon as they sequenced it in 2018. Oh wait.

Come on Stuart, even Daszak didn't know what WIV was doing. And he was paying for that research 😂

VANITY FAIR Subscribe

CORONAVIRUS See the latest coverage here →

If China's military had been collaborating with WIV scientists, it's unclear if Daszak would have realized it. He had far less visibility into the WIV than he let on, a former EcoHealth Alliance staffer told *Vanity Fair*. The work being done there was "always an enigma," the former staffer said. The nonprofit had hired a U.S.-based Chinese national who helped "interpret for them what was happening inside the WIV.... But we had to take everything at face value. It was more, 'Accept what it is, because of this relationship'" between Shi and Daszak.

"He doesn't know what happened in that lab," said the former staffer. "He cannot know that."

2:39 PM · Mar 31, 2022

Rootclaim said the WIV had 180 secret viruses.
I tried to track that down, I think maybe it comes from the DEFUSE grant?

But that is not 180 secret viruses, it's 180 total viruses. The next year, they published ~200 viruses in Latinne at al.

Section II **D. TECHNICAL PLAN**

Technical Area I:

Choice of site and model host-virus system. For the past 14 years, our team has conducted CoV surveillance in bat populations across S. China, resulting in >180 unique SARSr-CoVs in ~10,000 samples (>5% prevalence, including multiple individuals harboring the same viral strains)[2,21,33] and a per-bat species prevalence up to 10.9%. Bat SARSr-CoVs are genetically diverse, especially in the S gene, and most are highly divergent from SARS-CoV. However, our test cave site in Yunnan Province, harbors a quasispecies (QS) population assemblage that contains all the genetic components of epidemic SARS-CoV[34]. We have isolated three strains there (WIV1,

If SARS-CoV-2 was engineered, they would need some secret precursor virus.

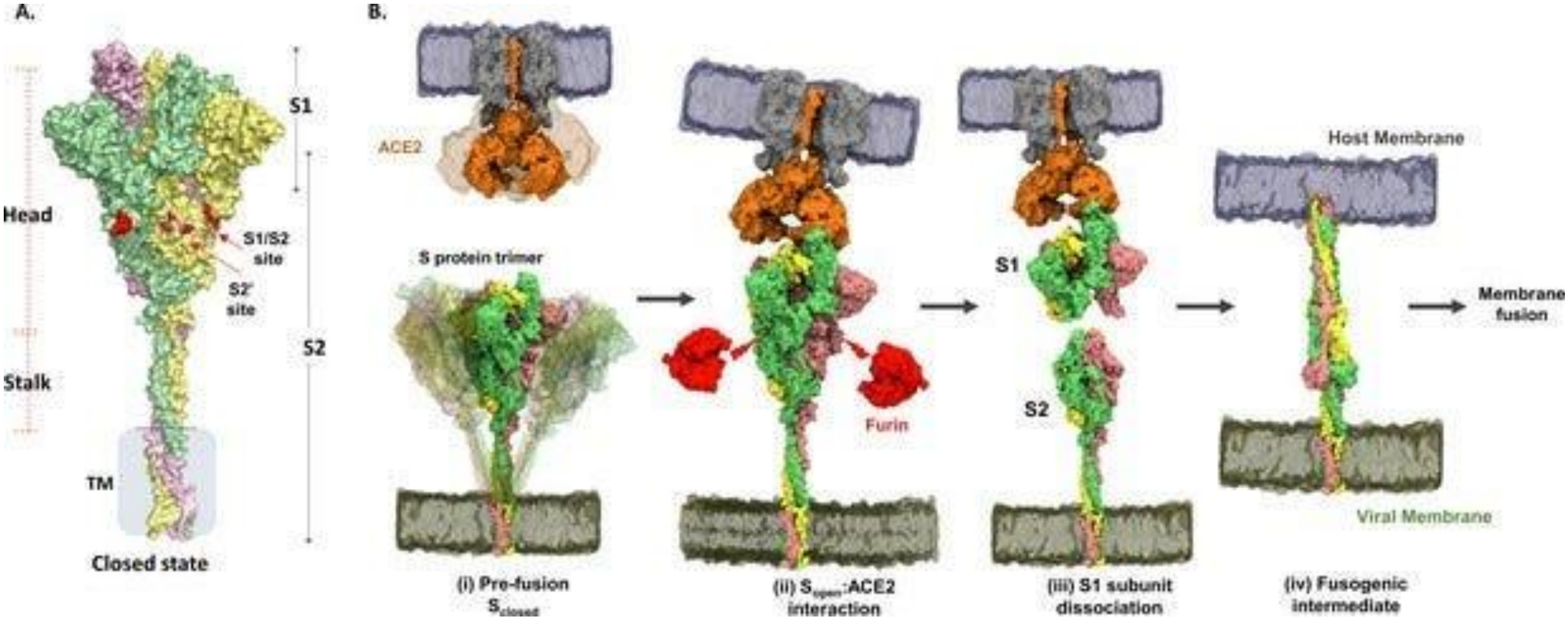
The lab would have no reason to keep this a secret, before the lab leak happened.

Because no other viruses are similar, we can’t just look for what’s been been changed.

The best we can do is look for features that seem unnatural.

The main feature that lab leak theorists point to is the “furin cleavage site”.

The spike protein fills two functions: S1 binds to ACE2, S2 fuses to the cell membrane

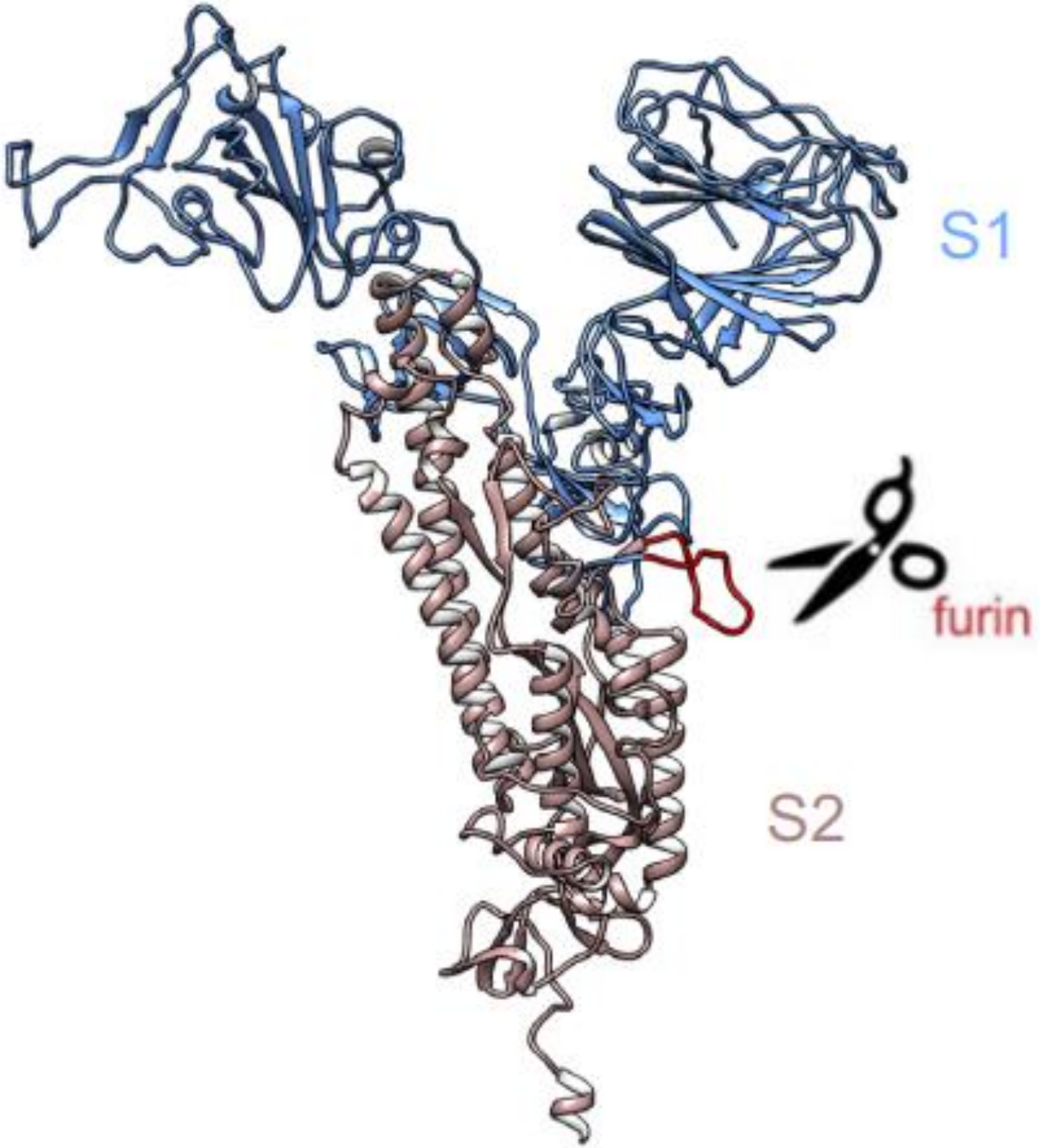


Furin cuts proteins everywhere it sees the amino acids RxxR (R is arginine, x is any amino acid).

Furin works better if it sees RRxR or RxRR.
RRxRR also works well.

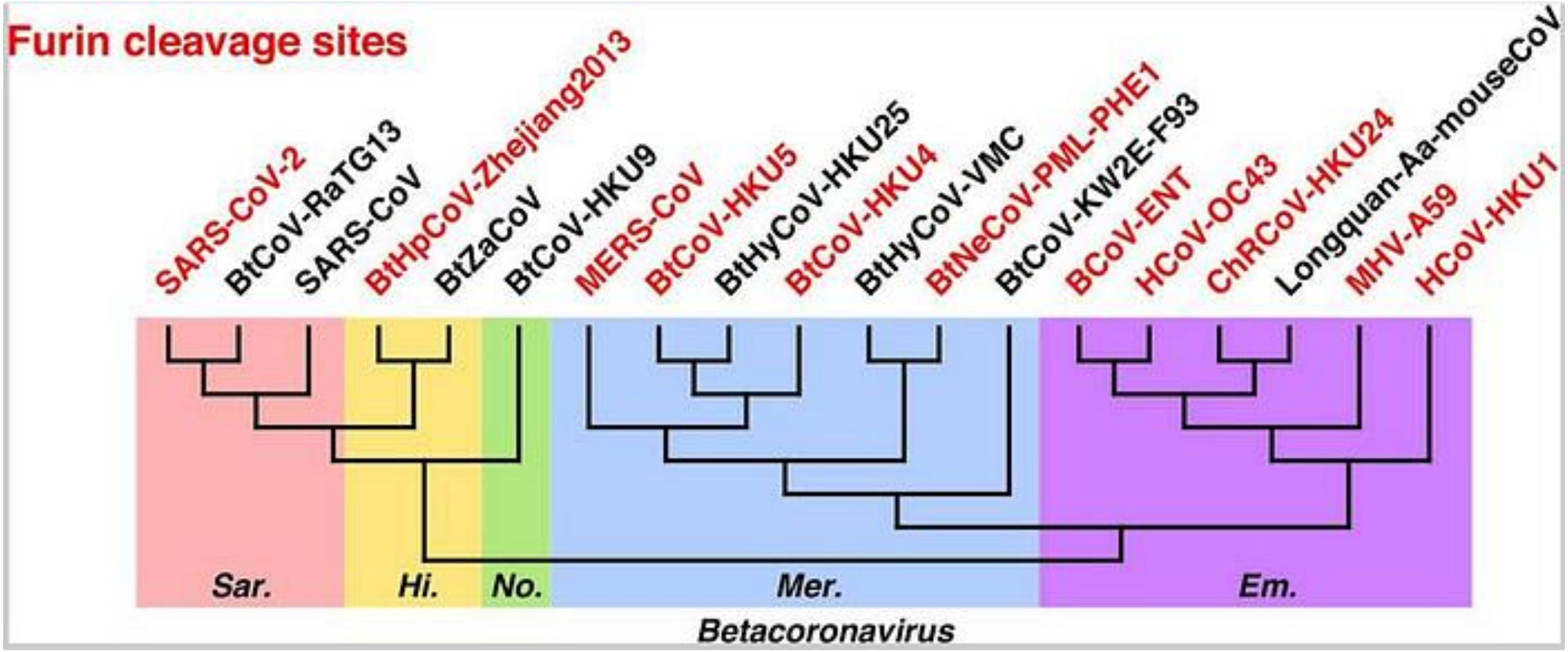
Previous experiments used RRKR or RRSRR.

In covid's case, the amino acids are PRRAR.



Are furin cleavage sites rare?

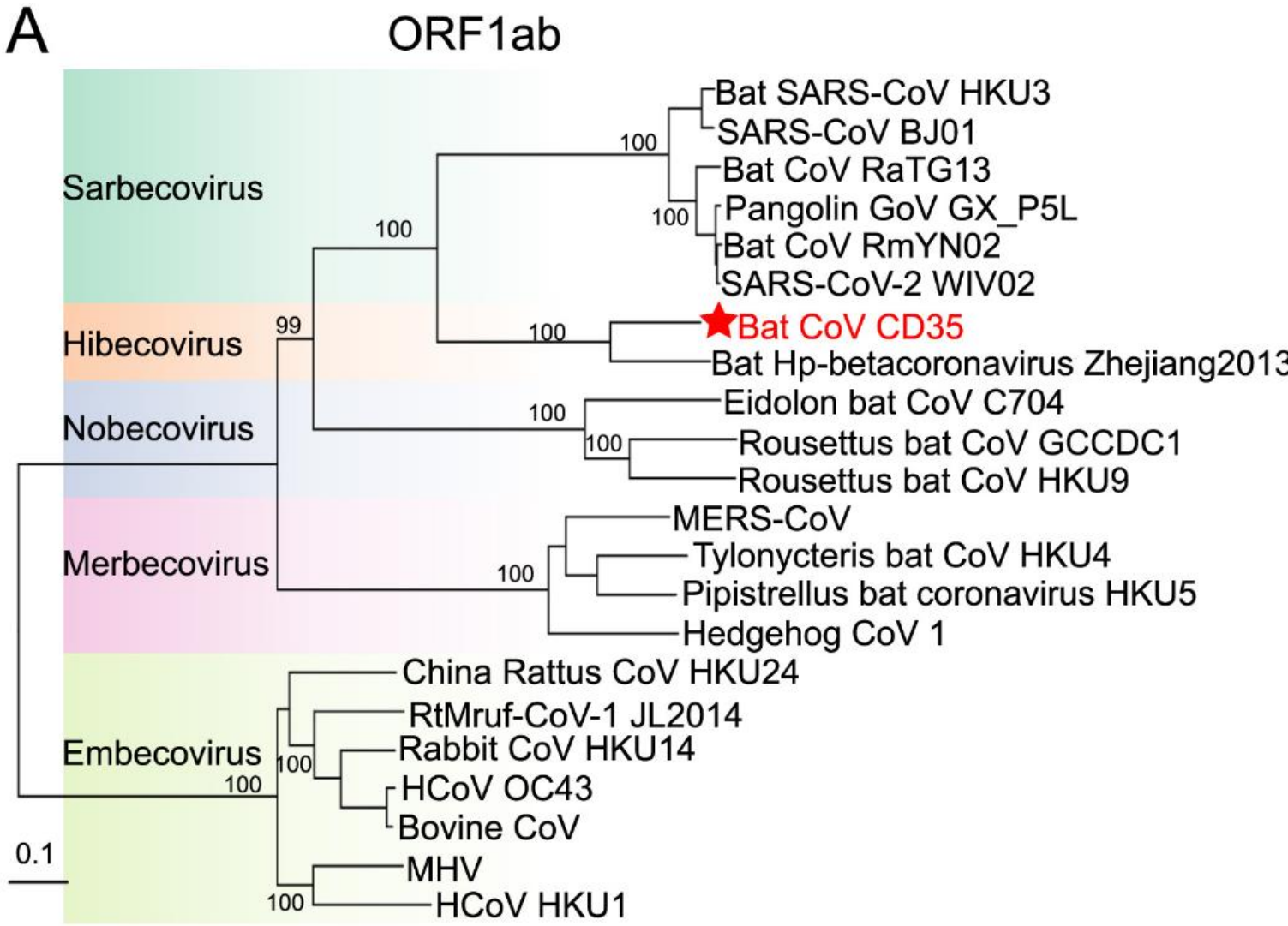
Furin cleavage sites are [found in lots of natural coronaviruses], but none were previously known in sarbecoviruses.



Since the pandemic started, scientists have been looking for more cleavage sites

A sarbecovirus found in UK bats is one mutation away from a FCS. It has RAKQ, one nucleotide change from RAKR.

Scientists recently found a bat hibecovirus with a furin cleavage site (RAKR):



4 out of 7 human coronaviruses have a furin cleavage site at S1/S2:

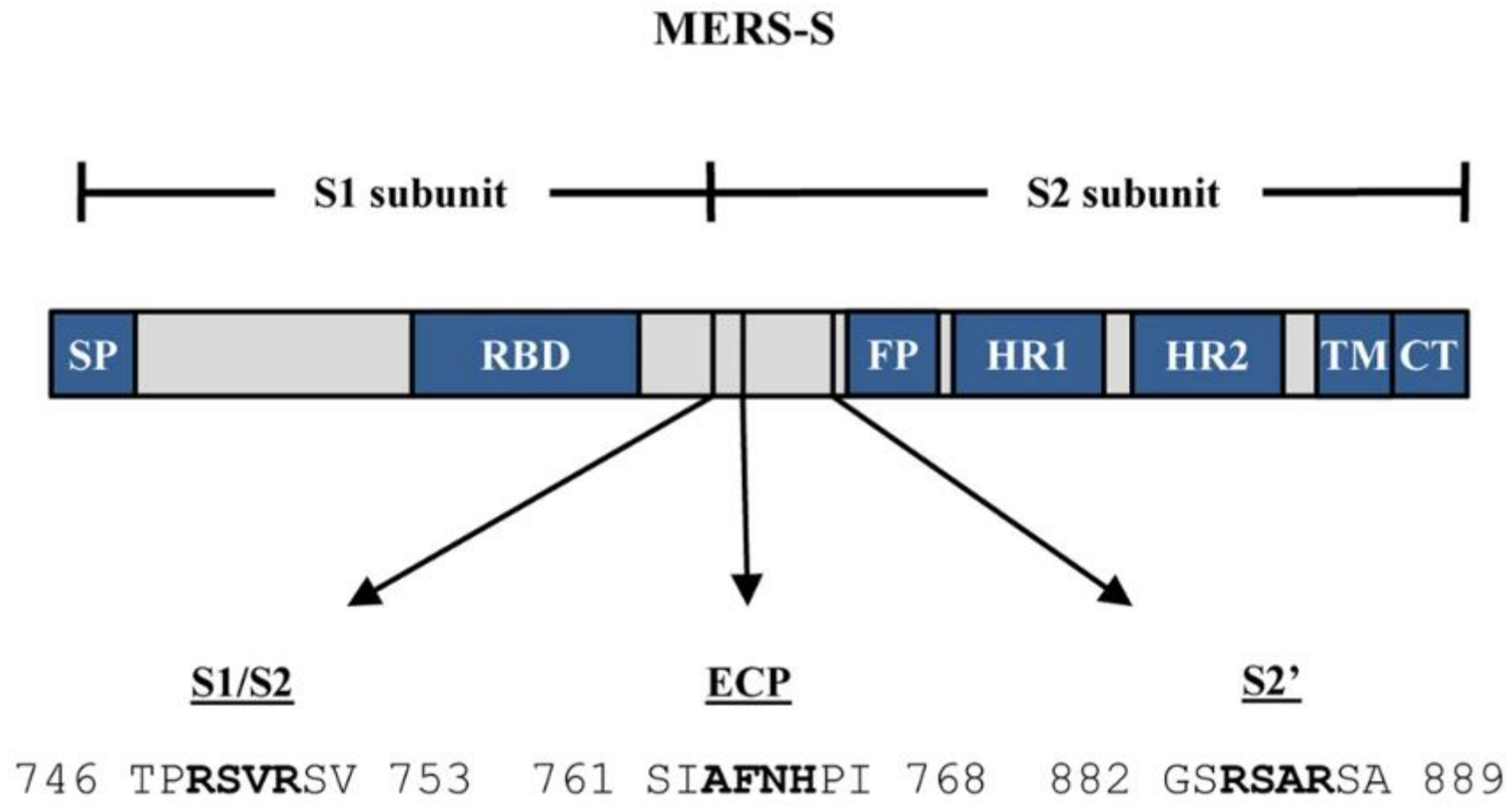
Alpha	HCoV-NL63	735	-	GICADGSLI----PVRPRNSS	-	751
Alpha	HCoV-229E	554	-	GVCADGSII----AVQPRNVS	-	570
Beta 2a	HCoV-OC43	753	-	GYCVDYSK-----NRRSRGAI	-	768
Beta 2a	HCoV-HKU1	742	-	GFCVDYNSPSSSSSRRKRRSI	-	762
Beta 2b	SARS-CoV	655	-	GICASYHTVS-L----LRSTS	-	670
Beta 2b	SARS-CoV-2	669	-	GICASYQTQT-NSPRRARSVA	-	688
Beta 2c	MERS-CoV	734	-	SLCALPDTPSTLTPRSVRSVP	-	754

(OC43 may actually be RRSRR)

There are 2 places a furin cleavage site can be placed:

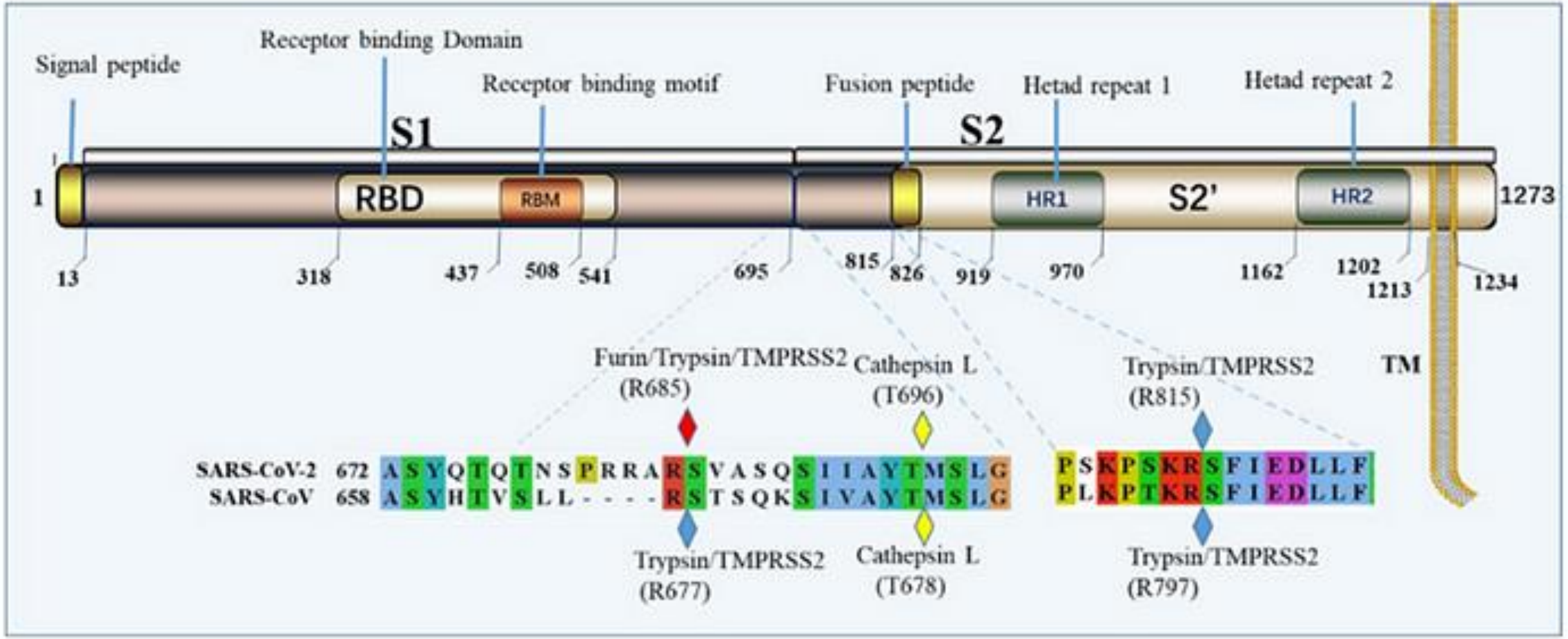
SARS-CoV-2 has a furin site at S1/S2, but not at S2’

MERS has 2 furin cleavage sites



MERS also has a cathepsin cleavage site (ECP = endosomal cysteine protease)

Cleavage happens via multiple enzymes, not just furin

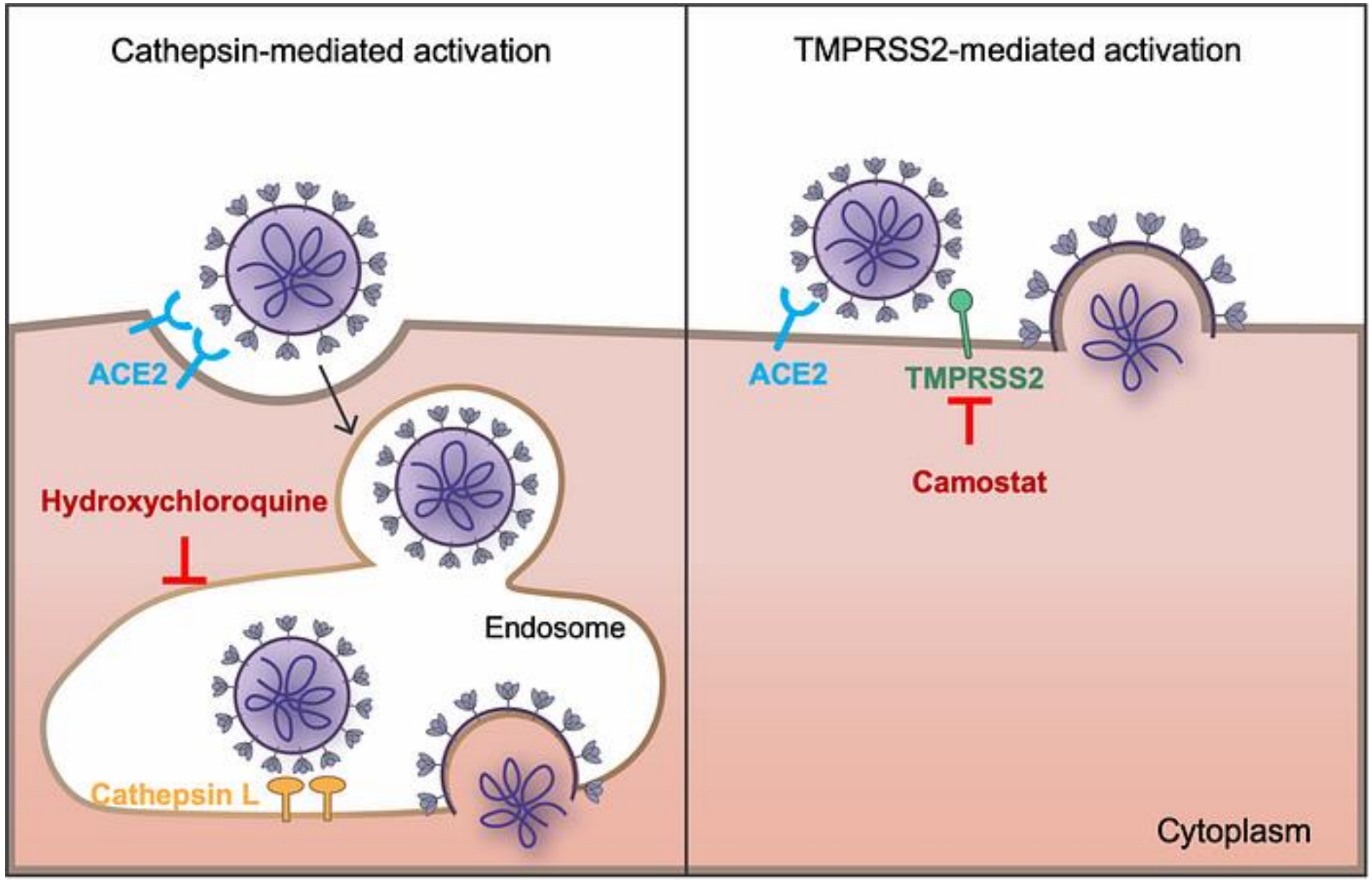


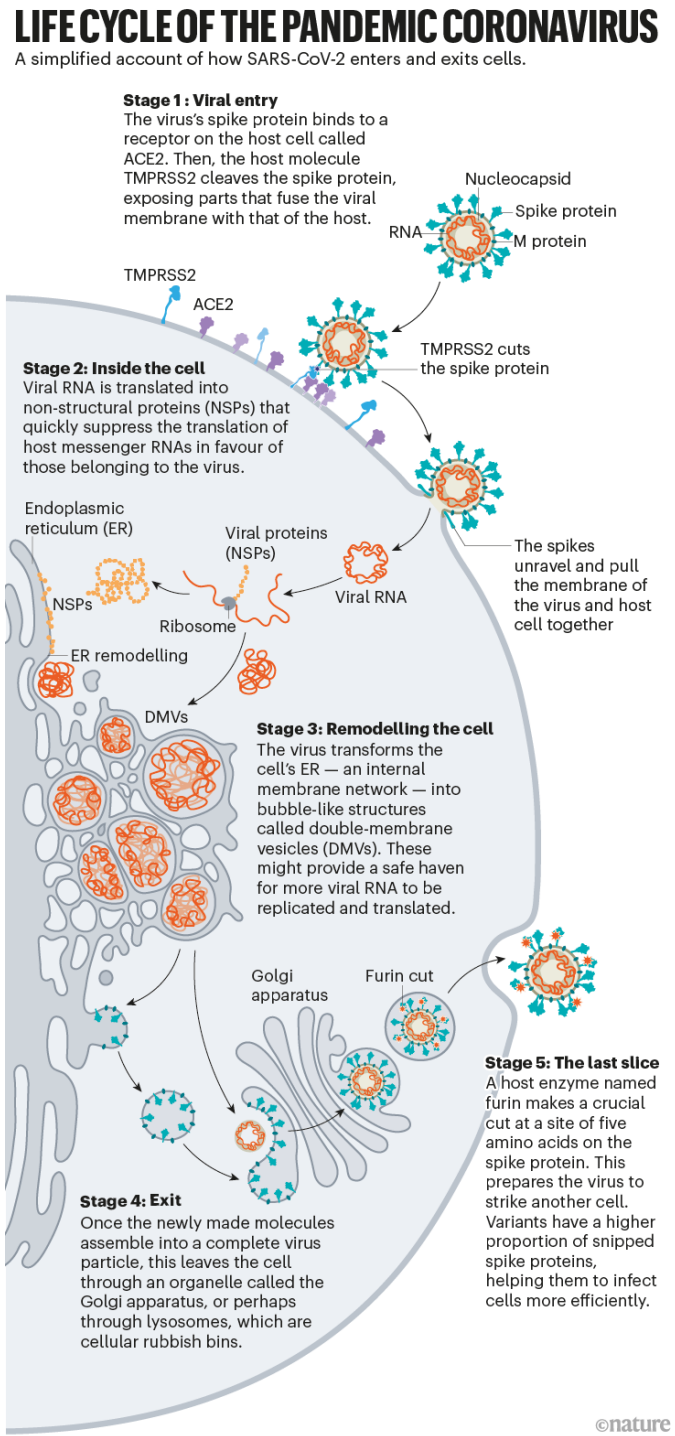
Cleavage happens via multiple enzymes, not just furin

"the S-protein is cleaved at a conserved sequence AYT↓M (located 10 amino acids downstream of S1/S2), by target cells' proteases such as elastase, cathepsin L or TMPRSS2"

Coronavirus	S1/S2, site 1	S1/S2, site 2	S2'
2019-nCoV	SP**R**RA**R**↓S**V**AS	IAYT↓MS	SKPS**KR**↓S**F**
SARS-CoV	TVSLL**R**↓STGQ	IAYT↓MS	LKPT**KR**↓S**F**
MERS-CoV	TP**R**SC**R**↓SVPG		GS**R**SA**R**↓SA
HKU1	SR**R**KR**R**↓SISA		CGSSS**R**↓S**F**
HCoV-OC43	KN**R**RS**R**↓GAITT		SKASS**R**↓SA
HCoV-229E	IAVQP**R**↓NVSYD		SRVAG**R**↓SA
HCoV-NL63	IPVRP**R**↓NSSDN		SRIAG**R**↓SA

Each of these enzymes enables a different method of cell entry, and can possibly be blocked by drugs





Furin can cut the spike protein as it is leaving the old cell

The virus is primed to attach to another nearby cell.

This enhances cell to cell fusion, causing damage in the lungs, for instance.

Other types of viruses also use furin cleavage sites. This is a common feature:

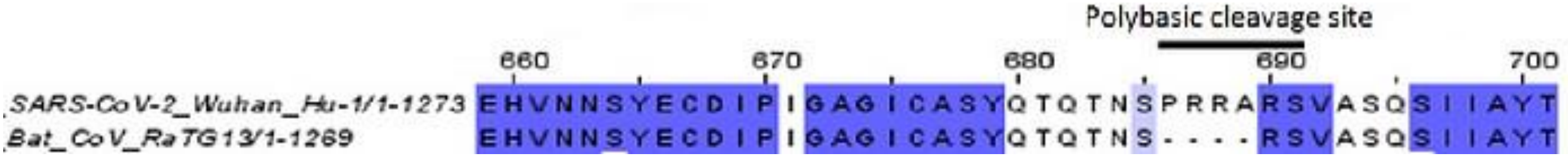
Virus	Cleavage site
HIV	VQ**R**E**KR**↓A**V**
Influenza Virus H5	**R**K**R**K**KR**↓G**L**
Avian H5N1 A/HK/98	**R**E**R**K**R**K**KR**↓G**L**
Avian H5N1 TKY/ENG	**NTPQR**K**KR**↓G**L**
Human CMV	**H**K**R**T**KR**↓S**T**
Human RSV	**KK**R**K**R**R**↓F**L**
Yellow Fever Virus	S**R**R**S**R**R**↓A**I**
Zika Virus	A**R**R**S**R**R**↓A**V**
Ebola virus	G**R**R**T**R**R**↓E**A**

Furin cleavage sites aren't rare, they've been found in 4 out of 5 families of betacoronaviruses.

But they haven't been seen in a bat sarbecovirus before.

This one sticks out, when you compare SARS-CoV-2 to similar viruses:

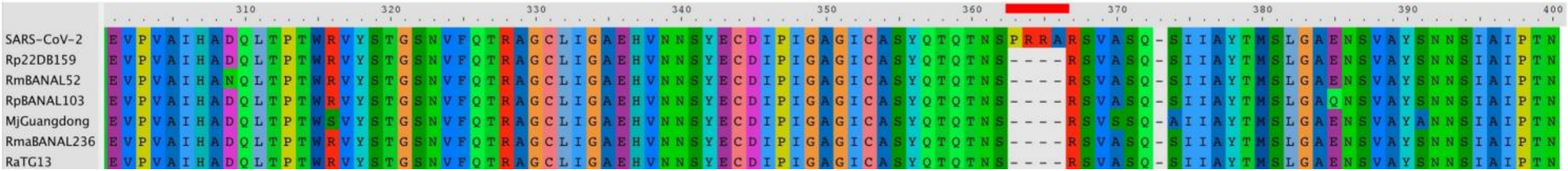
The closest known bat virus, called RATG13, has the exact same amino acids in much of its spike protein, except for those 4 at the cleavage site:



Since these two are so similar, except for the PRRA, people began to wonder if a lab had inserted that.

The over-all similarity of the spike protein is not unique.

Here’s SARS-CoV-2 compared to 6 similar coronaviruses:



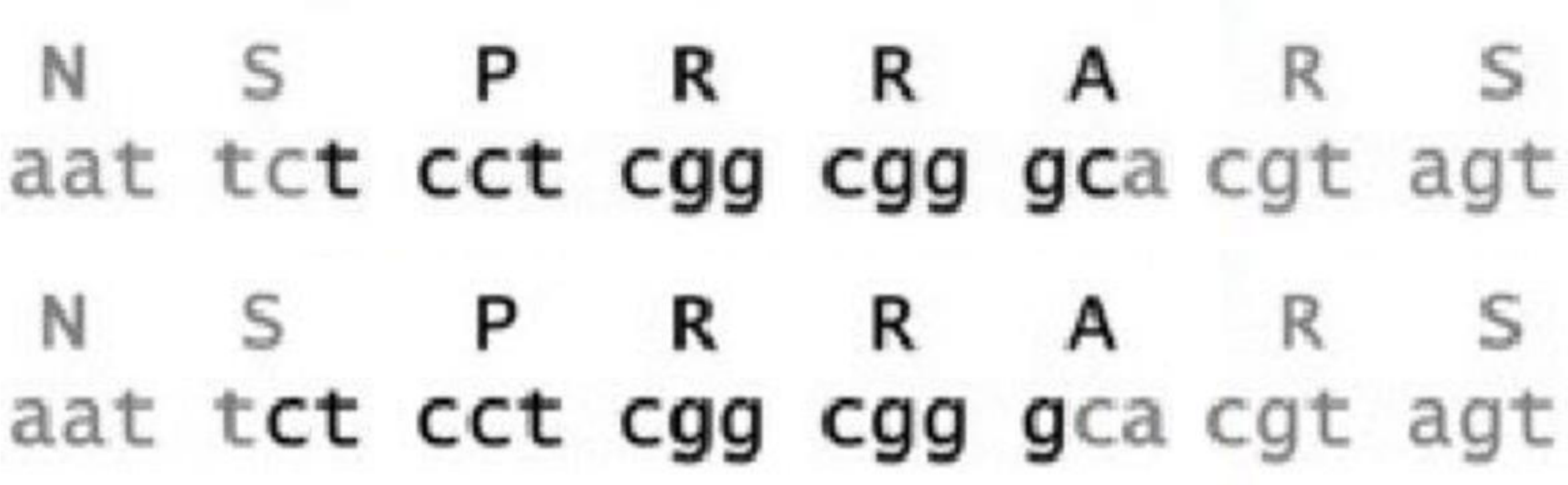
PRRA could theoretically have been inserted by a lab into some undisclosed coronavirus.

But there’s no reason to assume RATG13 was the starting point.

The 12 RNA letters that were “added” don’t line up with the PRRA

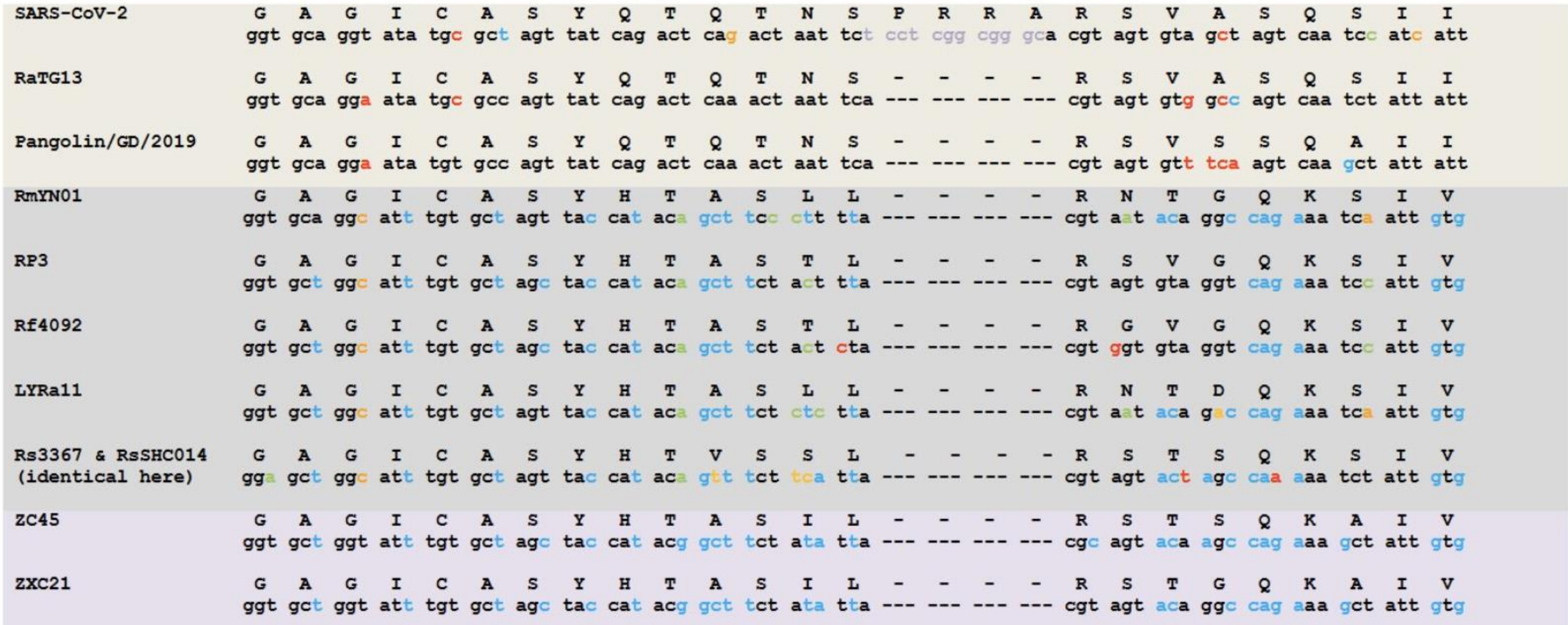
They are “out of frame”.

(there are actually 2 possible ways to align the insert, but both are out of frame)

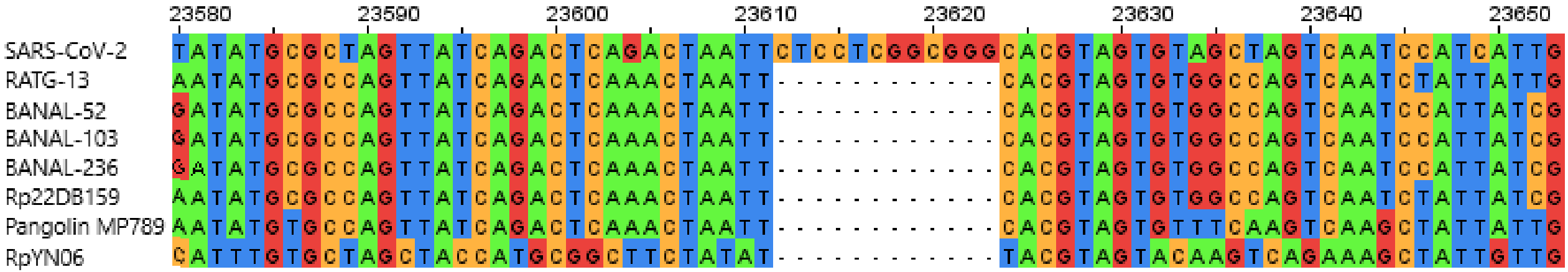


That’s the kind of random thing you’d find in nature, but it’s not what a human designer would likely do.

The PRRA insert is also out of frame relative to the similar pangolin coronavirus, and also relative to many other bat viruses:



It’s also out of frame relative to BANAL-52 and the other viruses from Laos:



Covid is also not the same as RATG13, besides this insert.

The amino acids in this section are the same but the RNA is not:

```
RaTG13          G   A   G   I   C   A   S   Y   Q   T   Q   T   N   S   -   -   -   -   R   S   V   A   S   Q   S   I   I
                ggt gca gga ata tgc gcc agt tat cag act caa act aat tca --- --- --- --- cgt agt gtg gcc agt caa tct att att

SARS-CoV-2      G   A   G   I   C   A   S   Y   Q   T   Q   T   N   S   P   R   R   A   R   S   V   A   S   Q   S   I   I
                ggt gca ggt ata tgc gct agt tat cag act cag act aat tct cct cgg cgg gca cgt agt gta gct agt caa tcc atc att
```

Of the 288 letters on either side of the cleavage site, 19 are different between those 2 viruses.

Over the full genome, about 1,200 letters of RNA are different.

There are large differences in the spike protein. The receptor binding domain is different.

What does the pattern look like, for other viruses with cleavage sites?

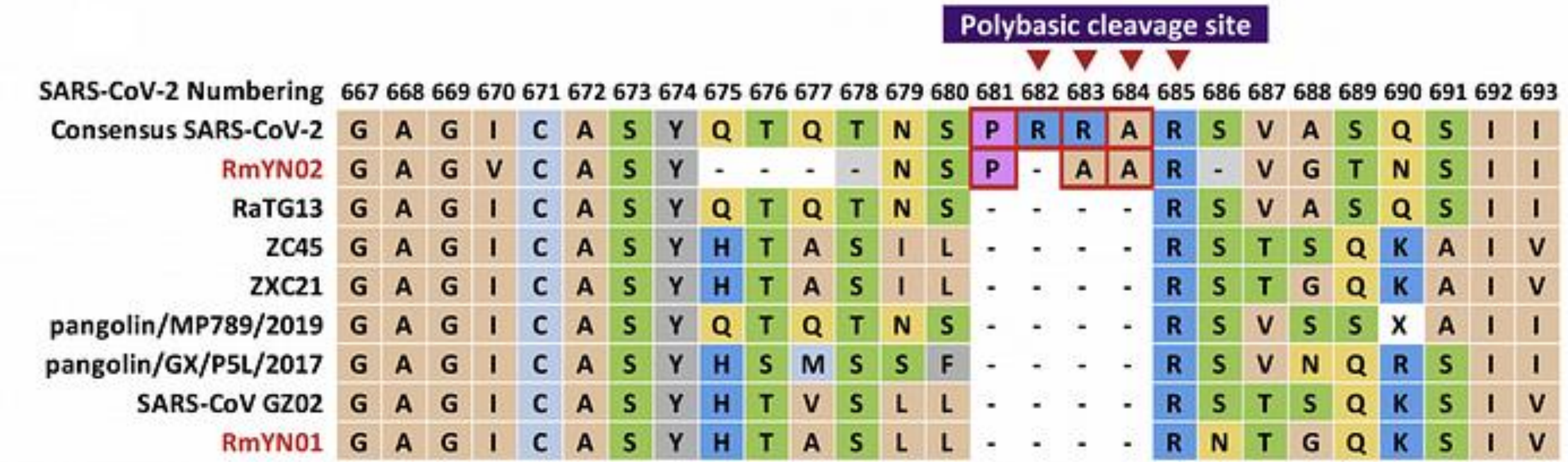
Here are a bunch of flu strains. If you compared two different flu strains, you'd also find examples that look like a 12 nucleotide insertion. So, maybe this isn't as abnormal as we think it is:

Examples of Natural Isolate Cleavage Site Sequences

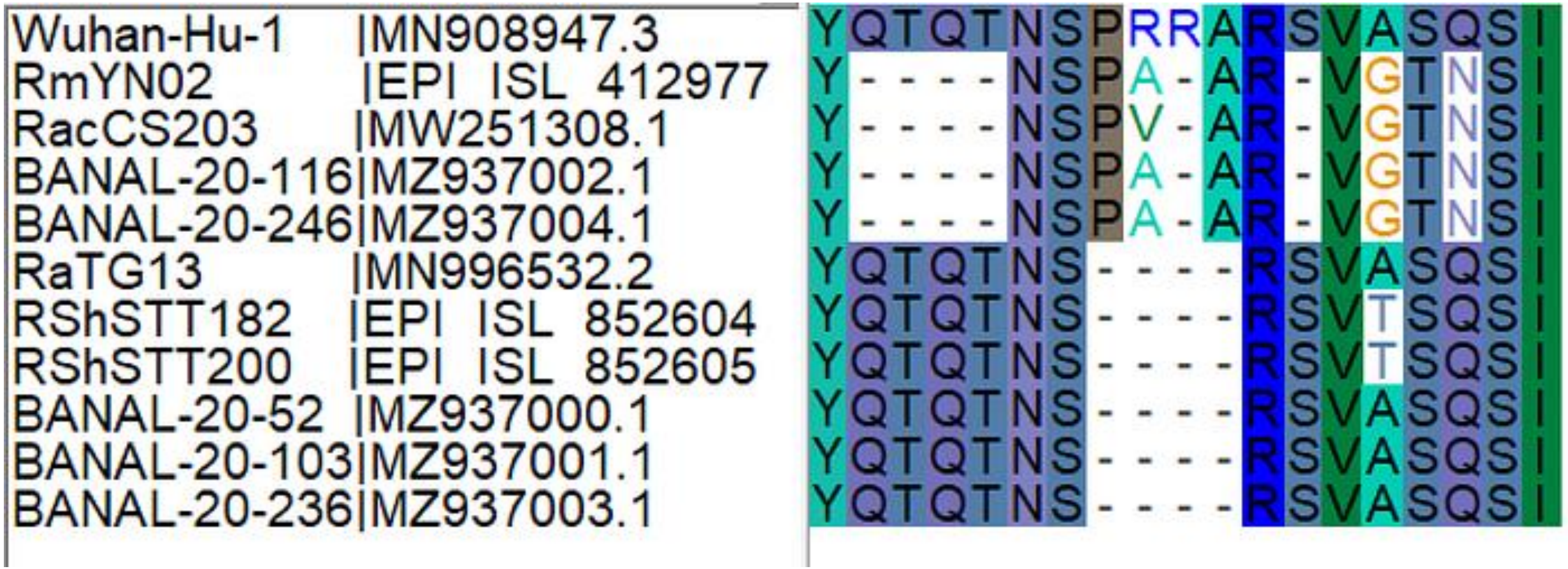
Isolate	Cleavage site	Pathogenicity
H5 subtypes		
A/chicken/Mexico/31381/94	P Q R E - - - - T R ↓ G	−
A/chicken/Pueblo/94	P Q R K R K - - T R ↓ G	+
A/chicken/Queretaro/20/95	P Q R K R K R K T R ↓ G	+
A/duck/Ireland/113/83	P Q R K R K - - K R ↓ G	+
A/turkey/Ireland/1378/83	P Q R K R K - - K R ↓ G	+
A/chicken/Pennsylvania/1/83 (CHO+)	P Q K K - - - - K R ↓ G	−
A/chicken/Pennsylvania/1370/83 (CHO−)	P Q K K - - - - K R ↓ G	+
A/duck/Singapore/645/97	P Q R E - - - - T R ↓ G	−
A/chicken/Hong Kong/990/97	P Q R E R R R K K R ↓ G	+
A/Hong Kong/156/97−(human)	P Q R E T R R K K R ↓ G	+
A/Hong Kong/486/97−(human)	P Q R E R R R K K R ↓ G	+
H7 subtypes		
A/tern/Potsdam/79	P E I P K - - - - G R ↓ G	−
A/chicken/Leipzig/79	P E I P K K K - - G R ↓ G	+
A/goose/Leipzig/137/79	P E I P K R K - - G R ↓ G	+
A/goose/Leipzig/187/79	P E I P K K K K - G R ↓ G	+
A/goose/Leipzig/192/79	P E I P K K K K K G R ↓ G	+
A/duck/Victoria/76	P E I P K - - - - K R ↓ G	−
A/chicken/Victoria/76	P E I P K K K E - K R ↓ G	+
A/chicken/Victoria/1/85	P E I P K K R E - K R ↓ G	+
A/starling/Victoria/5156/85	P E I P K K R E - K R ↓ G	+

Scientists went looking for more bat viruses to see if any had a cleavage site here, or something close to a cleavage site.

In 2020, someone published one that had 3 amino acids (or 9 nucleotides) [at this site]:



By 2021, 3 more SARS family viruses had been found with similar sequences at the S1/S2 junction



The function of the P in PRRA isn't immediately clear.

The suspicious "insert" is PRRA, but the furin cleavage site is just RRAR.

There's no clear reason why a designer would add the Proline.

The proline isn't obviously necessary, and it might actually be detrimental.

Cleavage sites can be evaluated by software, designers would likely run this before doing any experiments.

In software models, adding PRRAR is worse than just RRAR.

Using a textbook furin site like RRKR is better than either.

Cleavage site scores from Prop 1.0 model:
PRRAR: 0.626
RRAR: 0.782
RRKR: 0.884

Prop 1.0 model has a scale from 0 to 1.0, 0.5 is a barely functional cleavage site.

PRRAR is mediocre, it should be better without the P, no scientist would predict this as the best choice.

What furin cleavage sites do experimenters usually add to viruses?

2006 US study inserted RRSRR into the S1-S2 site in a SARS-CoV-1 pseudovirus

2009 US study inserted RRSRR into the S1-S2 and S2’ site in a SARS-CoV-1 pseudovirus

2008 Japanese study inserted KRRKR into S2’ site in a SARS-CoV-1 pseudovirus

2014 Dutch study inserted RRRRR, into S2’ in a mouse hepatitis coronavirus (pseudovirus).

2019 Chinese study inserted RRKR into S2’ in a chicken virus (gamma-CoV infectious bronchitis virus).

These are all efficient cleavage sites, as predicted by theory and software.

4 experiments were done with safe pseudoviruses and the last one doesn’t infect humans.

Why use the proline?

Minimal furin cleavage site is RxxR

One theory, from Yuri Deigin, is that it's because MERS has a proline before the cleavage site: PRSVR

MERS does have a leading proline, but if you put PRSVR into Covid, it wouldn't work.

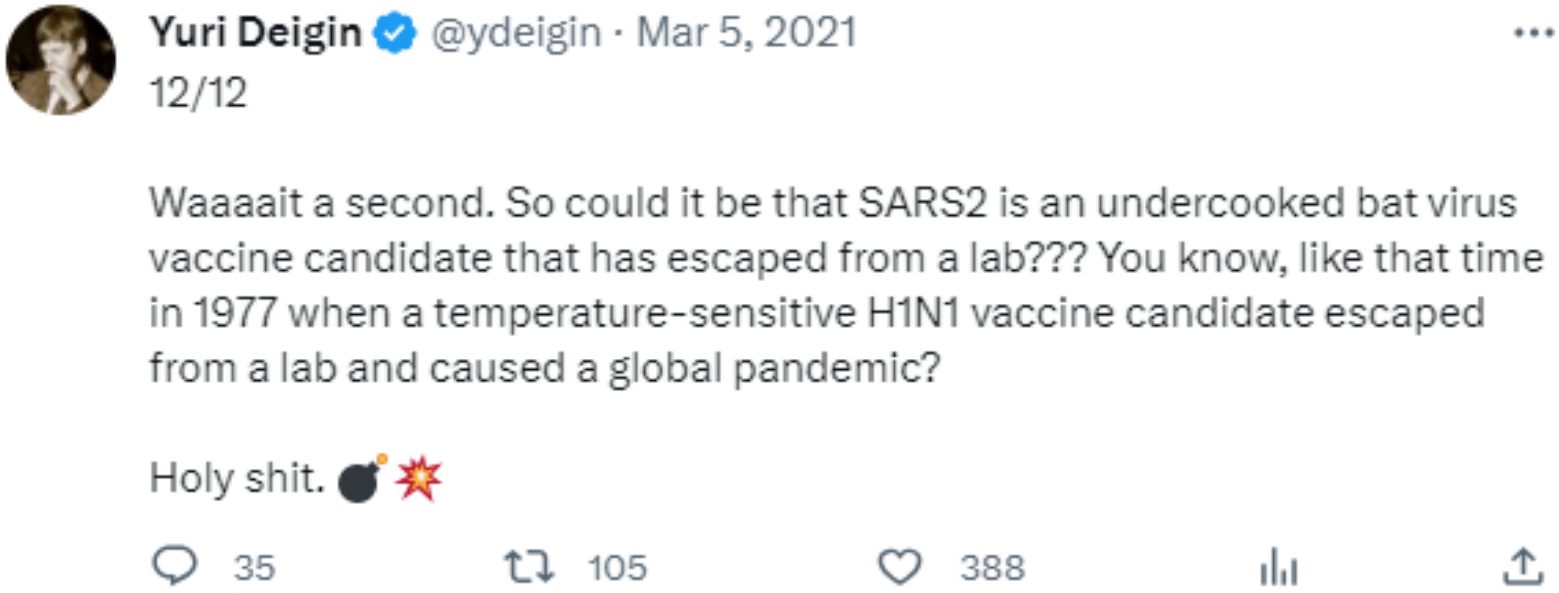
Experiments have shown that RxxR doesn't work in SARS-CoV-2, it needs RRxR.

Prop 1.0 scores:
PRRAR in Covid: 0.626
PRAAR in Covid: 0.593
PRSVR in MERS: 0.556

Yuri also speculates the lab started with the cleavage site from MERS, then added another R. But that would still be PRRVR.

In the podcast linked above, Yuri speculates this was done to test a pan coronavirus vaccine.

In another thread, Yuri speculates that Covid is itself a “self-spreading vaccine”:



Yuri Deigin @ydeigin · Mar 5, 2021

12/12

Waaaait a second. So could it be that SARS2 is an undercooked bat virus vaccine candidate that has escaped from a lab??? You know, like that time in 1977 when a temperature-sensitive H1N1 vaccine candidate escaped from a lab and caused a global pandemic?

Holy shit. 💣💥

35 105 388

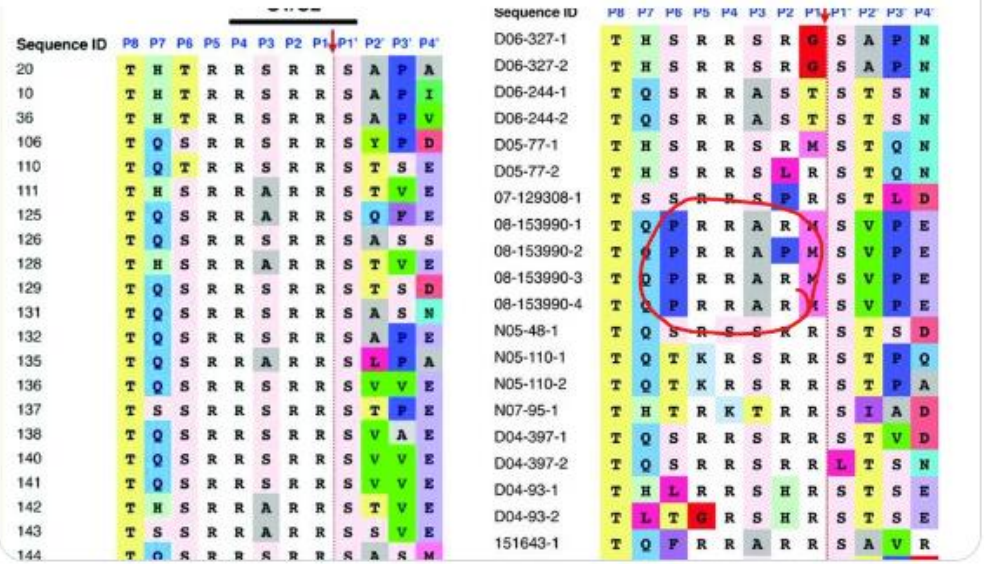
Yuri [also speculates](#) that PRRAR was somehow inspired by a feline coronavirus.

This theory is also [promoted](#) by Alex Washburne.

Yuri Deigin ✓
@ydeigin

This is an interesting idea as to why one might have wanted to engineer a non-canonical FCS into proto-SARS2: they could have been inspired by how a mutated FCS in feline CoVs turns a harmless FECV into a deadly FIPV. Some of those deadly strains had their FCS mutate into PRRAR:

Raccoon Dog's Breakfast ✓ @breakfast_dogs · Sep 24, 2022
Replying to @breakfast_dogs
They showed that in FIPV the FCS had much more variability than the well conserved FECV sequence R-R-[S/A]-R-R. Note also that a couple of these sequences are very close to SARS-COV-2's PRRARS (the additional M breaks the sequence though).



8:21 AM · Sep 25, 2022

←Raccoon Dog's breakfast is some random guy that also [thinks that SARS 2003 was a lab leak](#).

There's also a theory, popularized by Jeffrey Sachs, that tries to explain the RRAR by comparison to human ENaC.

Yuri noticed this back in 2020.

But this theory doesn't explain the proline, you have to mix and match with some other lab leak theory to explain why both the P and A were chosen.

Also, the codons are different from human ENaC.

Also, this theory likely requires the University of North Carolina consulting with the WIV on this issue: you'd need some collaboration between someone at WIV who has seen a bat virus with RSVAS and someone at UNC who's written papers about human ENaC.

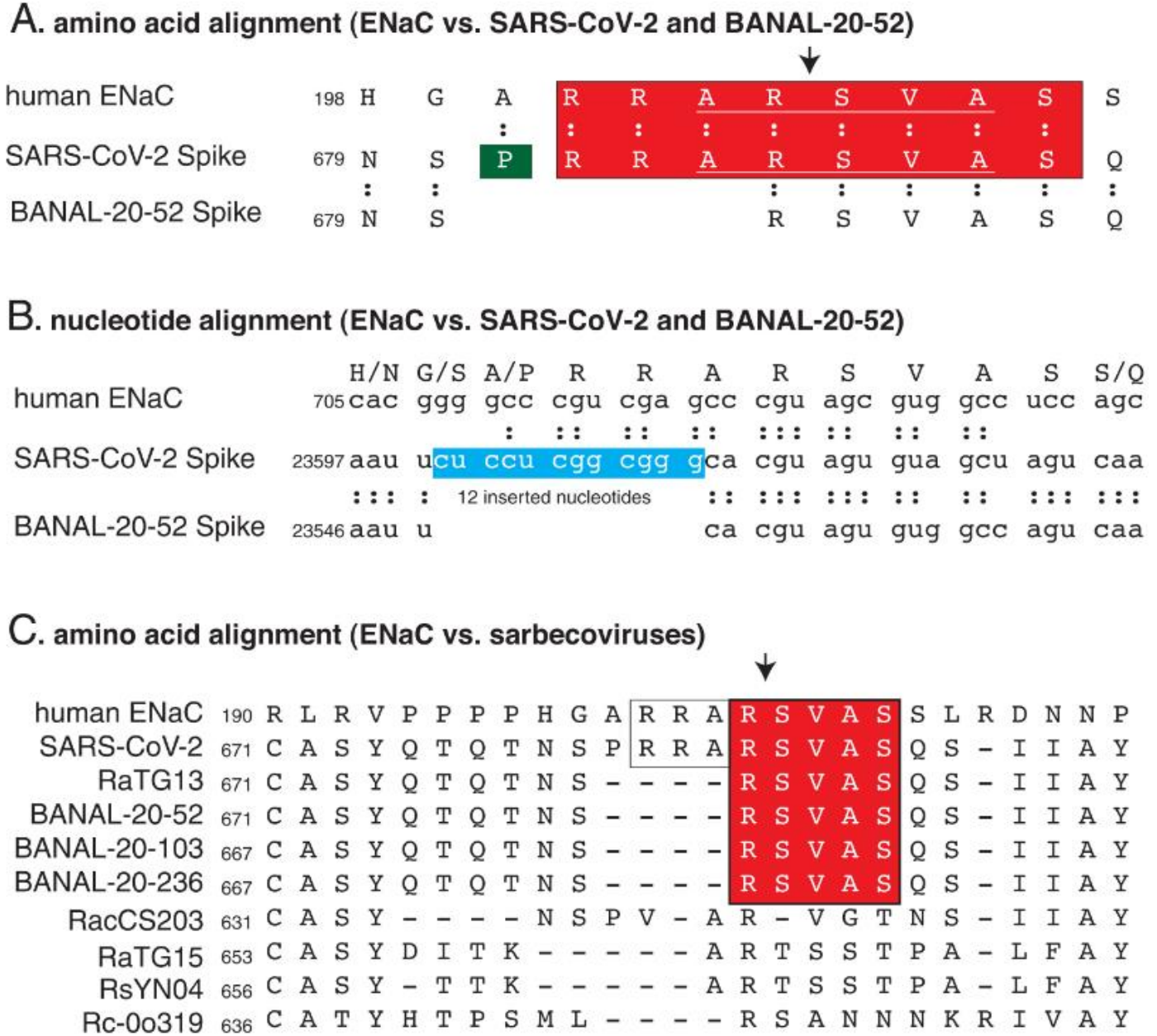


Figure from Garry 2022.

The Human ENAC theory is not consistently supported, even by the lab leak community:

Holtz @Biorealism · Feb 17, 2022

The SARS-CoV-2 cleavage site (RRAR|SVAS) is only found in hENaC, which was characterised precisely at UNC. UNC of course were named in EcoHealth's 2018 DEFUSE proposal with WIV and there are close ties between the respective labs. Seems fairly damning?

Alina Chan
@Ayjchan

It's one of those observations, like the CGGCGG or Moderna patent, that at first seems damning, but once you think about it more, the hypothesis falls apart.

5:25 PM · Feb 17, 2022

Alina Chan @Ayjchan · Feb 17, 2022

If you look at many of SARS-CoV-2's closest relatives, they have the same last 5 letters "RSVAS" in the stretch of 8 letters "RRARSVAS" that are similar between SARS-CoV-2 and ENaC.

Alina Chan @Ayjchan · Feb 17, 2022

It's not clear to me that inserting an "RRA" upstream of the conserved "RSVAS" sequence would've required the input or suggestion by a lung/ENaC specialist. Especially a specialist who didn't have access to SARS-CoV-2-like sequences prior to the pandemic.

2 2 8

Alina Chan @Ayjchan · Feb 17, 2022

Prior to the pandemic, the public sequences of SARS-like viruses did not include those with the "RSVAS" sequence. RaTG13's spike sequence was only published in 2020 and the rest of the close relatives were only characterized and published post-pandemic.

1 1 6

Alina Chan @Ayjchan · Feb 17, 2022

Unless the Wuhan scientists had sent these SARS2-like spike sequences to their collaborators at UNC, there would have been no match to "RSVAS".

If UNC had received SARS2-like sequences prior to the pandemic and are keeping this info from the public, that would be horrifying.

8 6 27

Alina Chan @Ayjchan · Feb 17, 2022

What seems much more plausible to me is that, among the hundreds of diverse SARS-like virus sequences collected by the scientists over the years, they found one or more with a furin cleavage site in the spike S1/S2. And this might've inspired the insertion of PRRA into SARS2.

1 4 11

Alina Chan @Ayjchan · Feb 17, 2022

This requires way fewer coincidences or improbable events than the ENaC hypothesis.

2 2 8

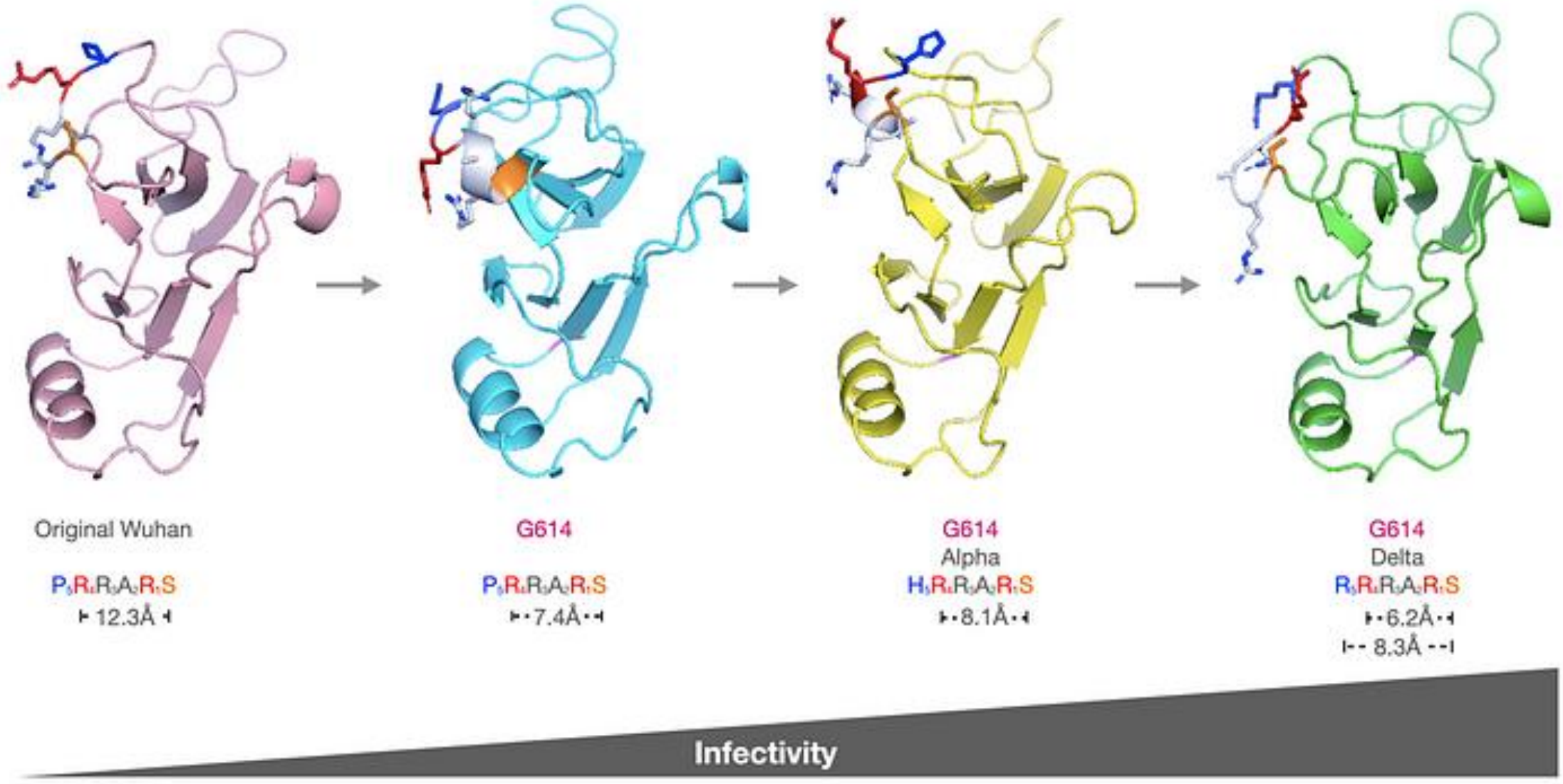
Shunsuke Otani @Shunsukepere · Feb 25, 2022

horrifying indeed..it's about time you could figure it out..they (CIA and Ralph Baric) made sars..WIV is just a red hering to put a blame on and to start a war against China like the current one against Russia

Because PRRAR is not optimal, Covid keeps mutating.

Alpha used HRRAR and Delta used RRRAR.

An early mutation (called D614G) also helped stabilize the FCS, indicating the virus hadn't spent much time in humans cells.



These changes improve the cleavage site, both in simulations and in reality.

The Prop 1.0 model does a pretty good job of explaining these mutations.

```
wuhan-hu1   0.626
alpha       0.706
delta       0.704
omicron     0.725
```

Models aren't perfect. Like, this model doesn't seem to recognize the D614G mutation as important, and maybe delta is actually better than alpha.

But it's still obvious that designers would not choose (out of frame) PRRAR as a good cleavage site.

Proline creates an inflexible protein, whereas the furin cleavage site usually needs the opposite:

"Position 681 of the envelope glycoprotein S of SARS-CoV-2 is being kept under monitoring since variants carrying mutations at this position have been linked to VOC forms of the virus. Originally, the envelope glycoprotein accommodated a proline residue at this position. Among Furin substrates, prolines are not very popular at the cleavage site because of their intrinsic rigid structure that confers restricted grades of freedom to the characteristic five-atom ring of proline. Accordingly, it is generally accepted that the substrates of Furin are characterized by a vast degree of flexibility, to allow optimal fitting into the catalytic site. In this respect, it is not surprising that the proline at the 681 position is subject to replacements. Moreover, the mutations at this particular position further make the stretch more accessible since the close-by P681-dependent (O-)glycosylation, which may hamper the substrate–enzyme docking, is lost. Overall, there is a clear pressure on the 681 position for the acquisition of a more suitable residue. P681R (Beta) and P681H (Delta/Omicron) are the most predominant over all possible amino acidic alternatives available."

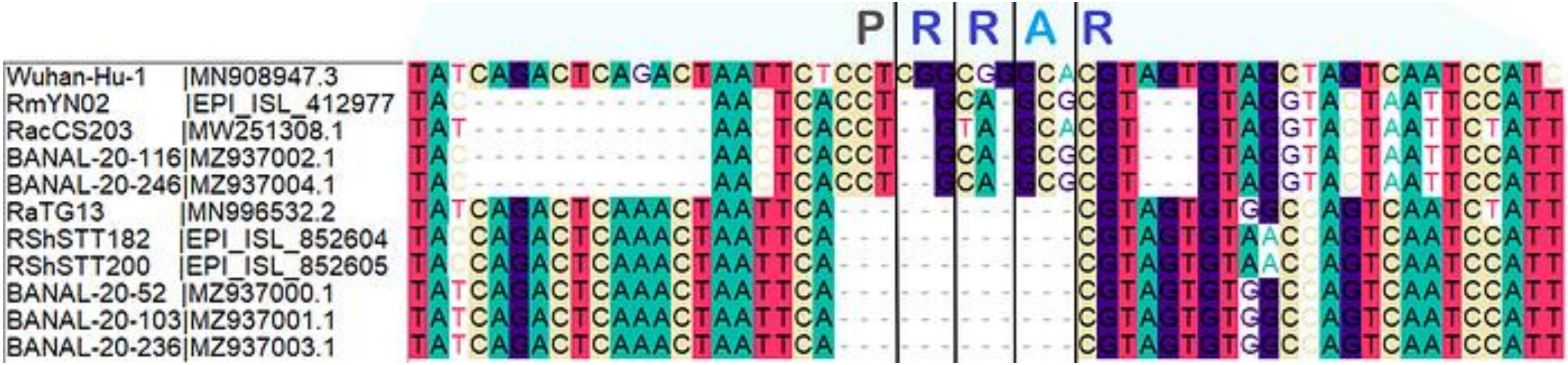
4 bat coronaviruses with a partial insert at S1/S2 do have the proline and the alanine

This might explain where the proline came from.

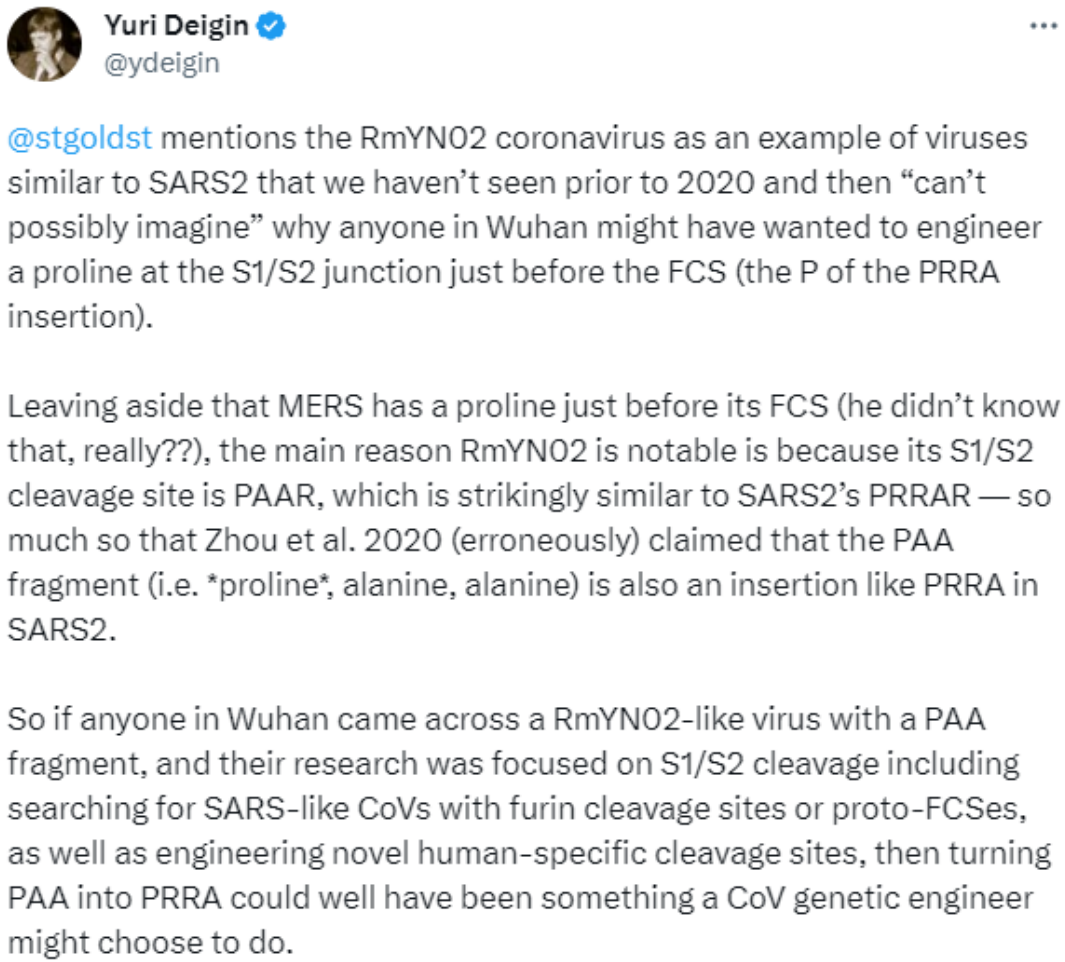
P can be spelled 4 different ways. It could be CCT, CCC, CCA, or CCG.

The P found in SARS-CoV-2 happens to use the same spelling as those other 4 bat viruses with a partial FCS.
That could be meaningful, could just be a coincidence.

Evolution would have to change 4 letters, to turn one of these natural viruses into Covid's FCS.



These viruses were all discovered or published after the pandemic started, but Yuri flips the theory on its head and imagines that maybe the lab had secretly found one of them and used it for inspiration to create Covid.



Yuri Deigin
@ydeigin

@stgoldst mentions the RmYN02 coronavirus as an example of viruses similar to SARS2 that we haven't seen prior to 2020 and then "can't possibly imagine" why anyone in Wuhan might have wanted to engineer a proline at the S1/S2 junction just before the FCS (the P of the PRRA insertion).

Leaving aside that MERS has a proline just before its FCS (he didn't know that, really??), the main reason RmYN02 is notable is because its S1/S2 cleavage site is PAAR, which is strikingly similar to SARS2's PRRAR — so much so that Zhou et al. 2020 (erroneously) claimed that the PAA fragment (i.e. *proline*, alanine, alanine) is also an insertion like PRRA in SARS2.

So if anyone in Wuhan came across a RmYN02-like virus with a PAA fragment, and their research was focused on S1/S2 cleavage including searching for SARS-like CoVs with furin cleavage sites or proto-FCSes, as well as engineering novel human-specific cleavage sites, then turning PAA into PRRA could well have been something a CoV genetic engineer might choose to do.

This makes it difficult to disprove lab leak theories: as soon as you find evidence of similar viruses in nature, the lab leak theories change to say that the Wuhan lab secretly had that evidence, as well.

He's created a theory that now requires the lab have 2 secret viruses, not just one.

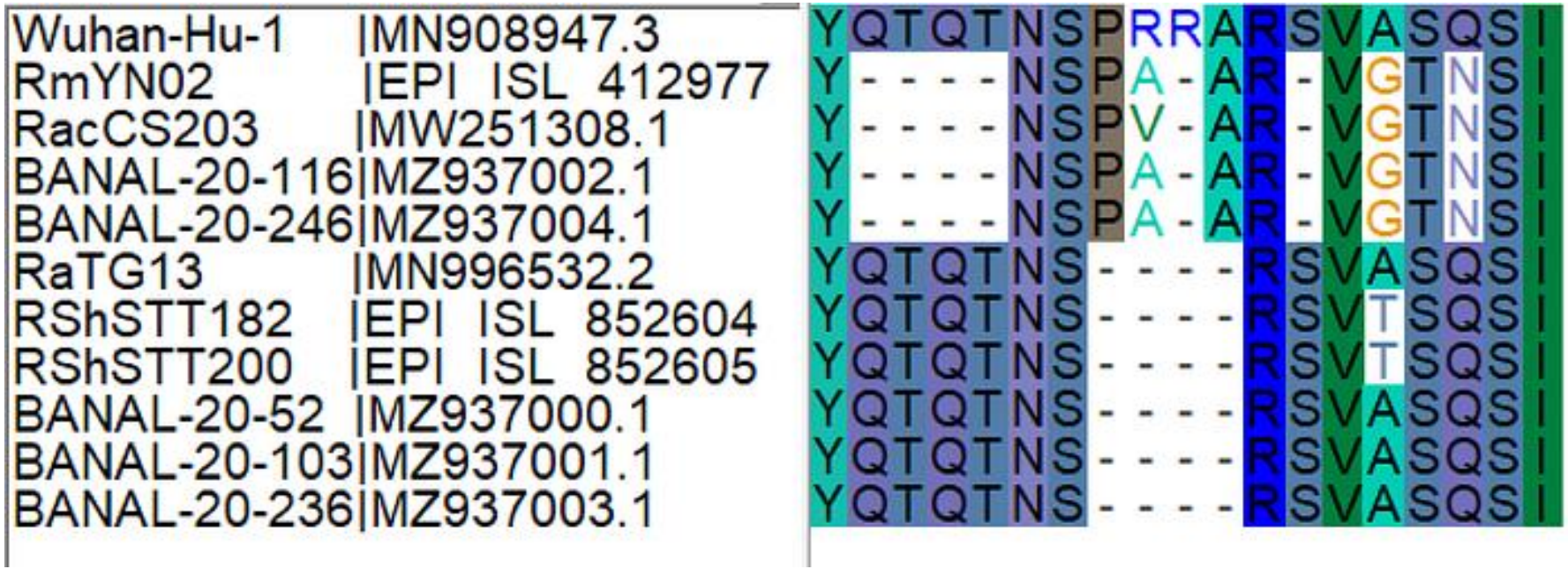
But, assume for a second that's true. I think this theory may only work in retrospect.
It's not obvious if you've never seen SARS-CoV-2.

If you saw PAAR, you could turn that into RRAR, not PRRAR
If you saw YNSPAAR, it's not obvious why you'd also turn that into YQTQTNSPRRAR.

You could also take RATG13 and make YQTQTNRRSR.

You can add an optimal furin cleavage site to any of these bat viruses by just adding RRK.

There's no reason why PAA would stand out, before Covid started and we were looking for P and A.



It is hard to predict, from first principles, some of the necessary features of a good cleavage site.

Before the pandemic, we knew that RRKR is best and PRRAR would be worse.

But during covid research we discovered the leading QTQTN is important. It extends the furin cleavage site loop:

Fig. 1.

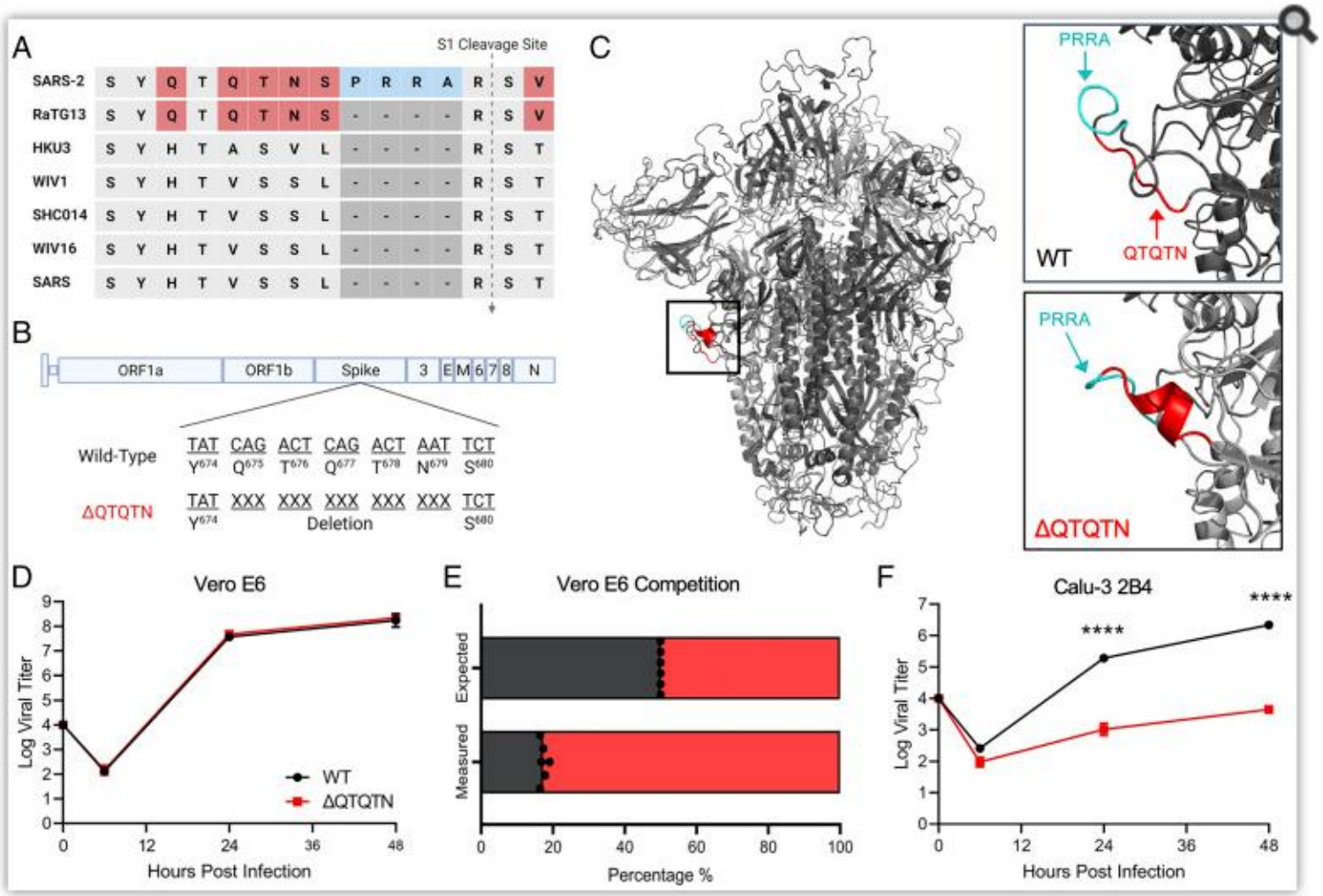
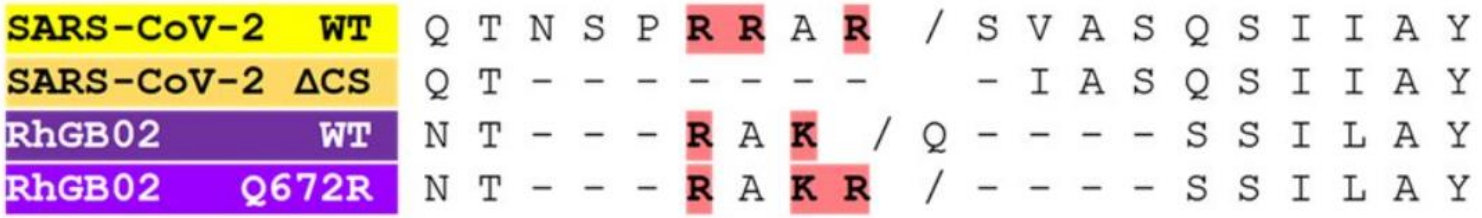


Figure from [Vu et al, 2022]

We found a sarbecovirus in UK bats that's [one mutation away from a FCS](#).

It has RAKQ, one nucleotide change from RAKR

But mutating it doesn't quite work, the FCS accessibility is too low for furin to reach it, it might need something else to extend the loop:

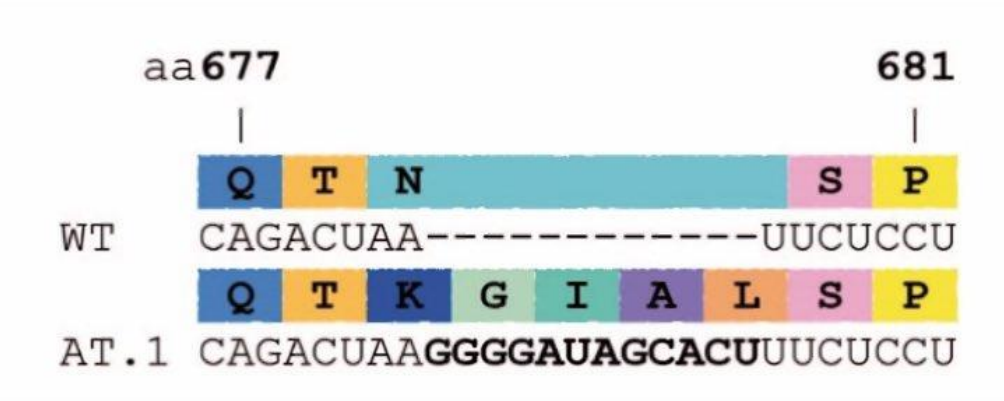


Remarkably, the RhGB01-like sarbecoviruses already possess a R-A-K-Q sequence (spike residues 669-672; Supplementary Fig. 10a), which is one nucleotide away (Gln/C**A**A to Arg/C**G**A) from the canonical R-X-K/R-R motif, a furin cleavage site (FCS) that allows cleavage by host furin-like proteases, enhancing the ability of many coronaviruses, including SARS-CoV-2, to infect human cells[47,48]. This R-A-K-Q motif is also found in Khosta-2[49], a sarbecovirus recovered from *R. hipposideros* in Khosta, Russia, which is at the south-eastern extremes of Europe, but not in BtKY72 from *Rhinolophus* sp. in Kenya[50] or other sarbecoviruses isolated from Asia. However, western blot analyses indicated that even when we mutated R-A-K-Q to R-A-K-R (i.e., a Q672R mutation), the RhGB07 spike is not cleaved by any human host protease (Supplementary Fig. 10b). Previous studies have shown that the FCS on SARS-CoV-2 (681-RRAR-684) lies on an extended flexible loop that protrudes out of the spike structure, which allows access by host furin[51,52]. Also, it has been shown that deletions that shortened this extended loop prevented efficient cleavage of SARS-CoV-2 spike, which was likely due to reduced accessibility of the FCS[48]. This loop is seven residues shorter in RhGB01-like viruses (Supplementary Fig. 10b), which may explain why no cleavage was observed for the RhGB07 R-A-K-R pseudovirus mutant.

These insertions can also happen naturally, in a single step, rather than as mutations

Here are 2 that we've observed in Covid, as the pandemic has progressed:

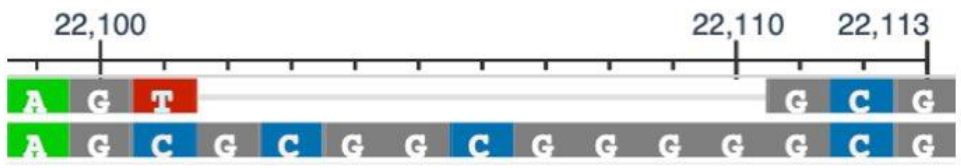
12nt out-of-frame insertion near the FCS



https://twitter.com/PeacockFlu/status/1558109368364797962?s=20

https://cov-spectrum.org/explore/World/AllSamples/AllTimes/variants?nucInsertions=ins_23598%3AGGGGATAGCACT&

9nt out-of-frame insertion with CGGCGG (RR)



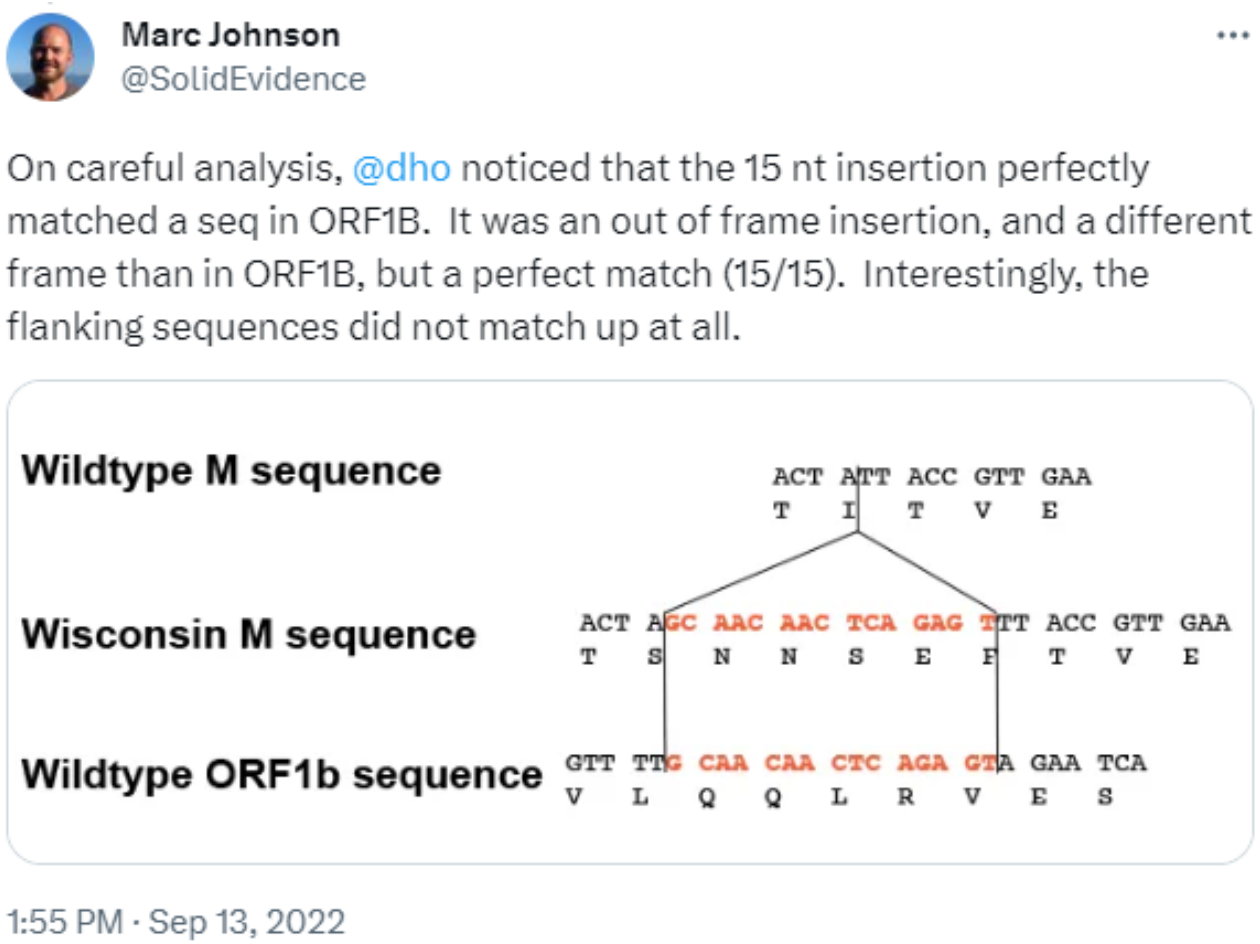
EPI_ISL_14446884, UW Virology Lab

https://twitter.com/alchemytoday/status/1615658595055443969?s=20

https://cov-spectrum.org/explore/World/AllSamples/AllTimes/variants?nucInsertions=ins_22198%3AGCGCGGCGG&

Examples from Flo Debarre

**Sometimes the insertions are [copied from elsewhere in the same virus](): **



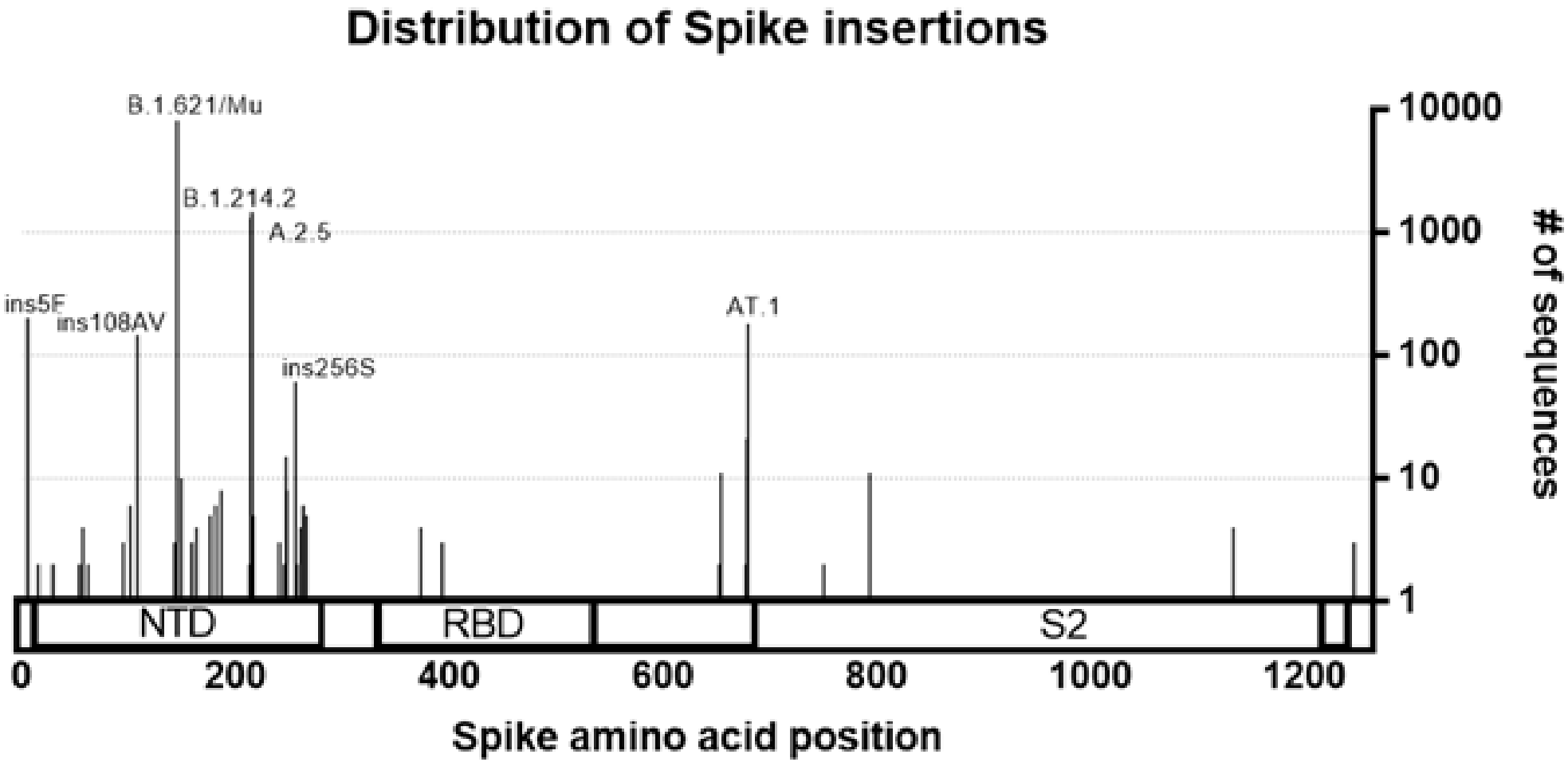
They can also be copied from a complementary strand of the virus, with C swapped for T and A for G.

Others are copied from [segments of human/host RNA]().

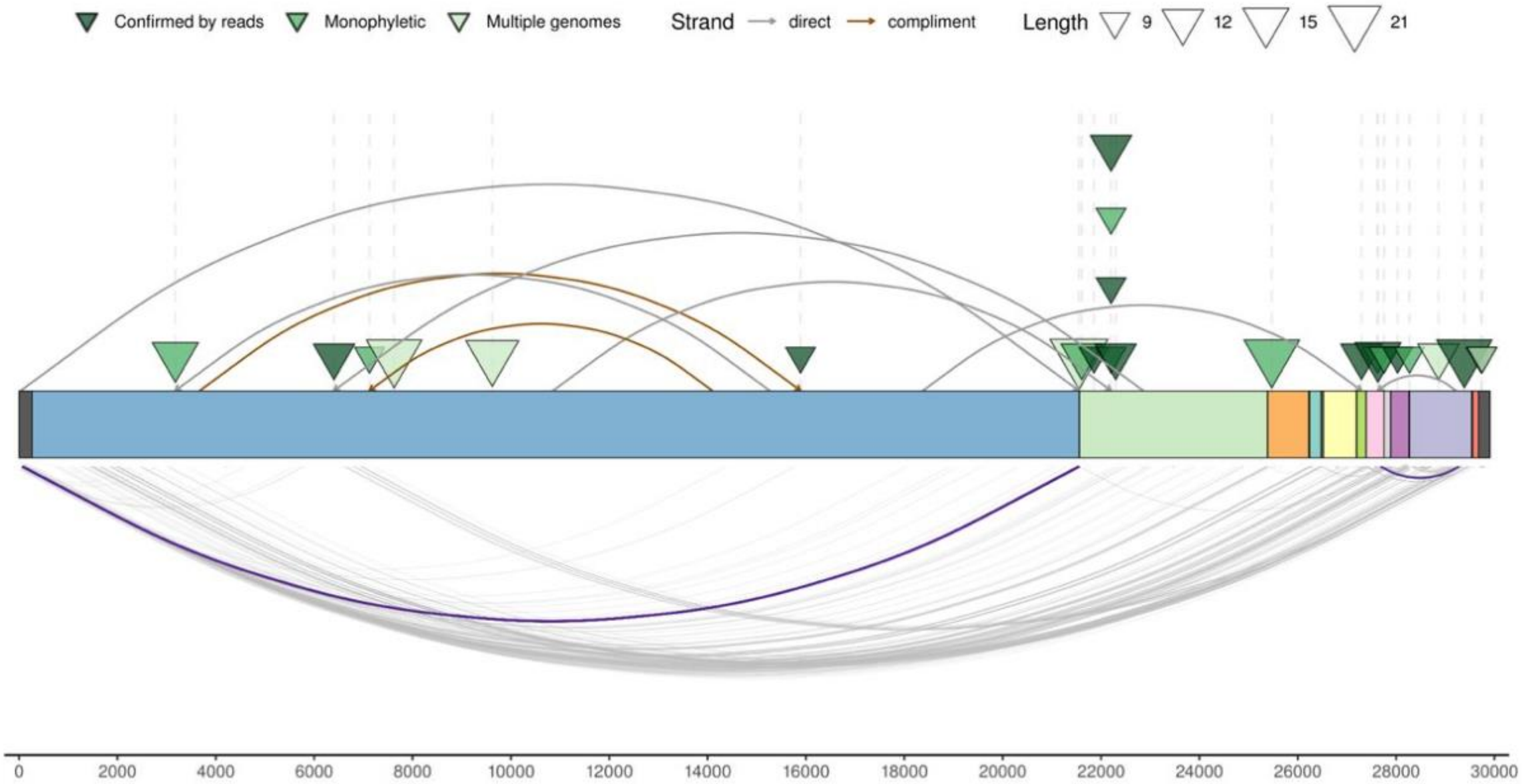
Others have no proven source.

Insertions into the spike show up most often in specific places.

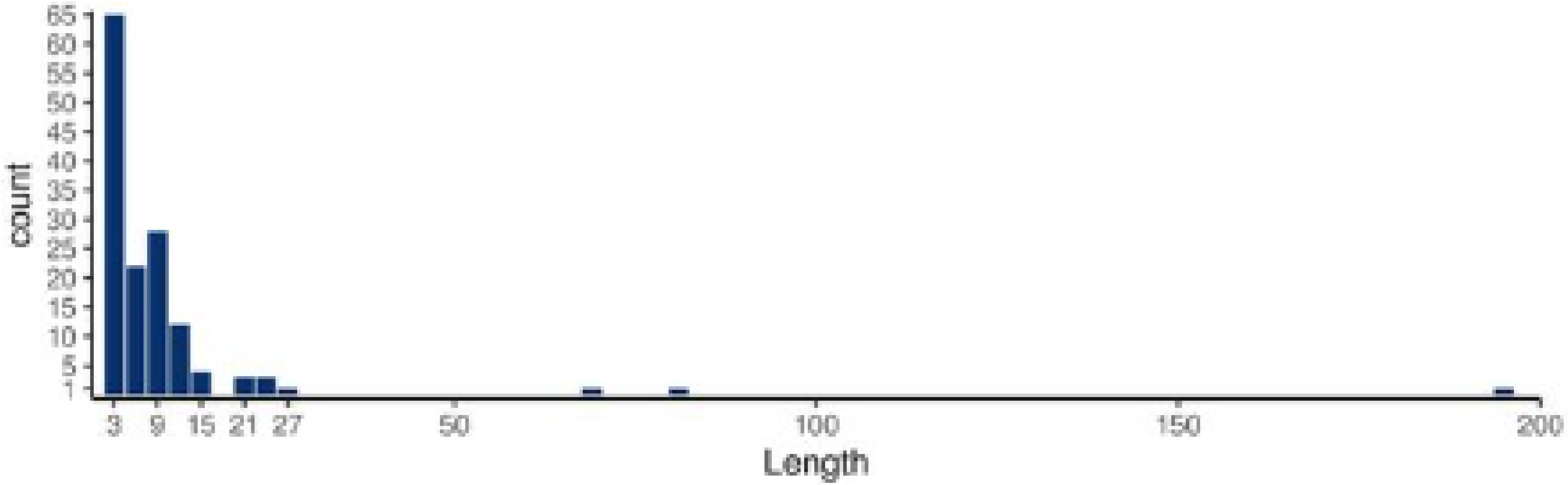
The S1/S2 site is one spot we've seen spike insertions (that's marked AT.1 here):



Another study looked across the whole genome and found most insertions happen in the final third:
The study looked at 141 insertions, with 25 longer than 9 nucleotides.
Seven of the 25 long insertions were located in the spike gene, significantly higher than expected by chance (p = 0.0165)



That study made a histogram of the length of various SARS-CoV-2 insertions.
12 nucleotide insertions are not that uncommon:



We can do some rough math on frequency:
498,000 covid genomes in GISAID. 4,468 had insertions. Odds around 1 in 100 of an insertion.

Of those 4,468 inserts, only 296 were unique. 12 unique inserts were 12 nucleotides long.

The odds are about 1 in 2,700 if you pick a random strain it will have some 12 nucleotide insertion.

You could also say it's 1 in 40,000 you'll find a unique insertion.

It's hard to know the underlying rate of mutation – all we see is the "frequency of insertions that worked".

Many people have tried to come up with an explanation for where this insert came from

In 2020, William Gallaher found a match for the insert in another bat virus, called HKU9:

HKU9 gcatttgta **caga------cctcggcgggc** ctctgt
CoV-2 tatcagact **cagac** ttgct **cctcggcgggc** acgtagt

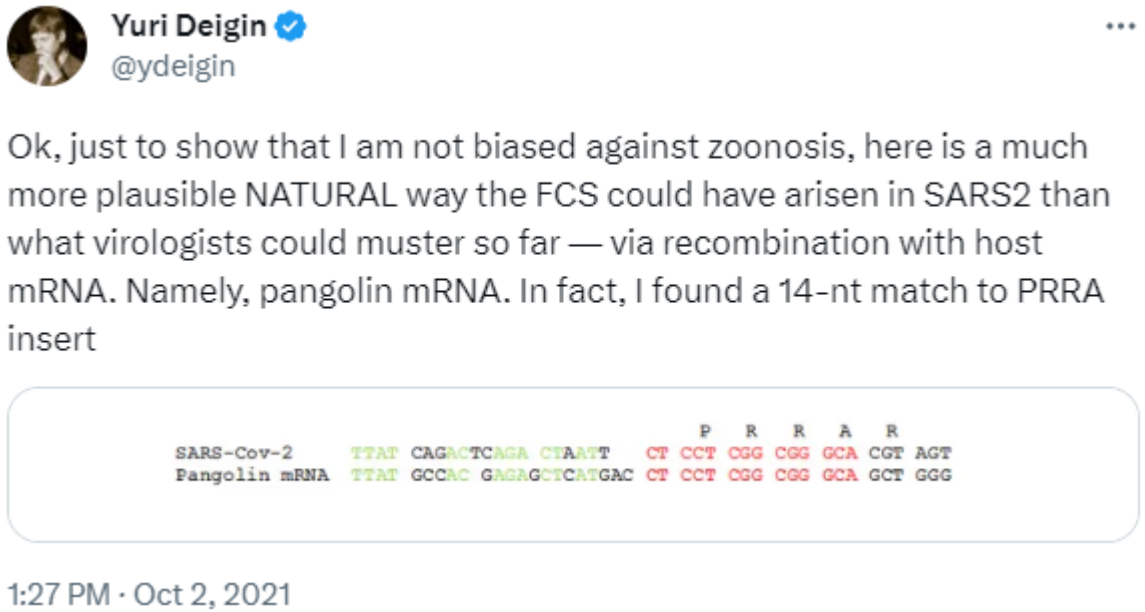
That's only 10 of 12 matched, and unlikely to be the source, because these viruses infect two different bat species.

The insertion could come from viral RNA or from host RNA

You can search for the insert sequence (CTCCTCGGCGGG) in BLAST (restrict it by species)

There are dozens of matches for raccoon dogs
There are > 100 matches for humans

Yuri Deigin found a 14 nucleotide match in pangolin mRNA:



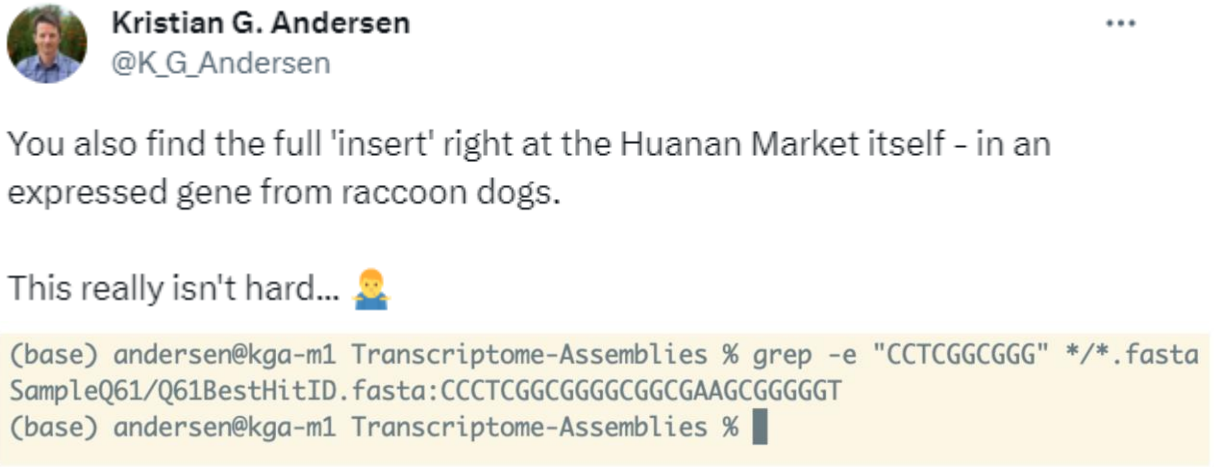
Yuri Deigin ✓
@ydeigin

Ok, just to show that I am not biased against zoonosis, here is a much more plausible NATURAL way the FCS could have arisen in SARS2 than what virologists could muster so far — via recombination with host mRNA. Namely, pangolin mRNA. In fact, I found a 14-nt match to PRRA insert

```
                                                P   R   R   A   R
SARS-Cov-2      TTAT CAGACTCAGA CTAATT    CT CCT CGG CGG GCA CGT AGT
Pangolin mRNA   TTAT GCCAC GAGAGCTCATGAC CT CCT CGG CGG GCA GCT GGG
```

1:27 PM · Oct 2, 2021

Kristian Andersen found 10 out of 12 nucleotides of the insertion, including the CGGCGG, in raccoon dog sample Q61, at the Huanan market:



Kristian G. Andersen
@K_G_Andersen

You also find the full 'insert' right at the Huanan Market itself - in an expressed gene from raccoon dogs.

This really isn't hard... 🤦

```
(base) andersen@kga-m1 Transcriptome-Assemblies % grep -e "CCTCGGCGGG" */*.fasta
SampleQ61/Q61BestHitID.fasta:CCCTCGGCGGGGCGGCGAAGCGGGGGT
(base) andersen@kga-m1 Transcriptome-Assemblies %
```

In summary:

The cleavage site doesn’t look like something a person would design.

We don’t know how it got their naturally. It could have come from one of the similar bat viruses which already have the Proline and Alanine, through some mutation and recombination.

It could have come as a single 12 nucleotide insert.

It could be a mix of mutation and insertion.

That insert could come from anywhere in another bat virus, or another host virus.

It could be copied from anywhere in the host RNA.

It could be copied with any frame shift.

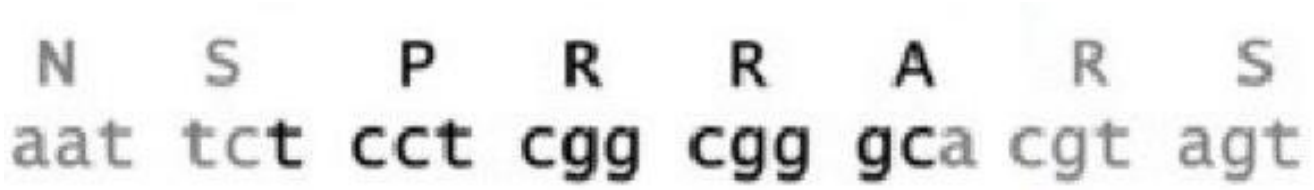
It can also be copied from a complementary RNA strand (with C swapped to T and A swapped to G).

The spike gene sees 33% of the long insertions and the S1/S2 junction is one spot we’ve seen insertions.

Why is it CGG CGG?

One lab leak theory holds that [CGG codons are proof of a lab origin](#).

Why is it CGG CGG?



R (Arginine) has 6 possible encodings.

CGG encoding is rare for R, in bat sarbecoviruses.

But CGG is also rare in human common cold coronaviruses.

SARS-CoV-2:		RATG13		human CoV OC43:	
CGT	21.6%	CGT	21.9%	CGT	30.8%
CGC	9.0%	CGC	9.4%	CGC	11.9%
CGA	7.2%	CGA	7.6%	CGA	8.6%
CGG	4.6%	CGG	3.6%	CGG	5.0%
AGA	42.8%	AGA	41.1%	AGA	31.7%
AGG	14.7%	AGG	16.4%	AGG	11.9%

It's frequently said that the CGG frequency in Covid is 3%, but that appears to be incorrect.
The 5% number can be confirmed in this paper or this one on codon frequency.

This isn’t about humans vs bats, it’s about animals vs viruses.

CGG frequency is 22% in humans and 20% in bats.

But it’s 5% in human viruses and bat viruses.

Human:		Bat:	
CGT	9%	CGT	8%
CGC	15%	CGC	16%
CGA	12%	CGA	12%
CGG	22%	CGG	20%
AGA	20%	AGA	21%
AGG	18%	AGG	23%

CGG is not a good encoding for a virus, because the immune system recognizes it.

William Gallaher, one of the first people to write about the novel coronavirus, wrote the following, on February 7th, 2020:
"One has to consider that the PRRA is an unusual sequence to introduce to generate a furin site – others even among coronaviruses like MHV A59 are so much better. Also that the underlying code CCTCGGCGGGCA introduces an unnecessarily G and C rich region where none otherwise exists. Not likely scenarios for something a gene jockey would do."

Kristian Andersen wrote,
"The CGG codon is rare in viruses because it's an example of an unmethylated "CpG" site that can be bound by TLR9, leading to immune cell activation"

Where did the CGG theory even come from?

It wasn't an obvious addition to the lab leak, at first, but it grew to become a popular part.

Yuri Deigin's first medium post wrote a whole section on "codon usage", where he carefully compared Covid with RATG13, SARS, and several other viruses. Despite all that work, he never noticed the double CGG.

The section concluded:

> So codon analysis also did not reveal any obvious signs of lab origins, but once again confirmed the uniqueness of CoV2 and RaTG13. What does this leave us with? So far, just a number of oddities, which, as scientists like to say, *taken together*, do not allow us to reject the lab origin hypothesis of CoV2.

If the double CGG is important, he certainly couldn't find it.

It appears that Yuri Deigin first learned about it from some random guy on Twitter, in May 2020.

It's not agreed upon by all lab leak theorists. Alina Chan says the CGG CGG "hypothesis seems damning at first" but, "once you think about it, it falls apart".

Today, this is a popular part of most lab leak theories.

Robert Redfield claims that SARS-CoV-2 uses "human codons" for arginine.

In Steven Quay's Bayesian analysis, he wrote (note that he provides no sources for the key claims):

> But it gets worse still for the zoonosis theory. The gene sequence for the amino acids in the furin site in CoV-2 uses a very rare set of two codons, three letter words so six letters in a row, that are rarely used individually and have never been seen together in tandem in any coronaviruses in nature. But these same 'rare in nature' codons turn out to be the very ones that are always used by scientists in the laboratory when researchers want to add the amino acid arginine, the ones that are found in the furin site. When scientists add a dimer of arginine codons to a coronavirus, they invariably use the word, CGG-CGG, but coronaviruses in nature rarely (<1%) use this codon pair. For example, in the 580,000 codons of 58 Sarbecoviruses the only CGG pair is CoV-2; none of the other 57 sarbecoviruses have such a pair.[8]

Is CGG the standard choice for experiments?

Probably not, but most papers don't list the codon choice for their experiments.

One of the few papers I could find that did was a 2014 Dutch study that used RRRRR for a cleavage site.

They used the nucleotide sequence AGA'CGC'CGA'AGG'CGT

That's literally every possible version of R except for CGG.

The Wuhan lab never added furin cleavage sites to viruses, that we know of, so there's no way to know which codons they might have chosen.

The closest thing I've seen is this Yuri Deigin claim:

Shi Zhengli published one paper with Shibo Jiang in 2020.

Shibo Jiang did a separate experiment, in 2013, where he added a furin cleavage site to some DNA.

One of the three arginines added was CGG, the other 2 were AGG and CGC.

The experiment had something to do with bacteria, not coronaviruses, and it was done in Guangzhou, not Wuhan.

But, Shibo Jiang and Shi Zhengli are authors on a different paper together, after the pandemic starts.

Therefore, Yuri concludes, Shi Zhengli was probably putting CGG into coronaviruses, left and right.

In any case, CGG is a rare spelling of R in a coronavirus. What does this mean?

It could be suspicious.

It could be random.

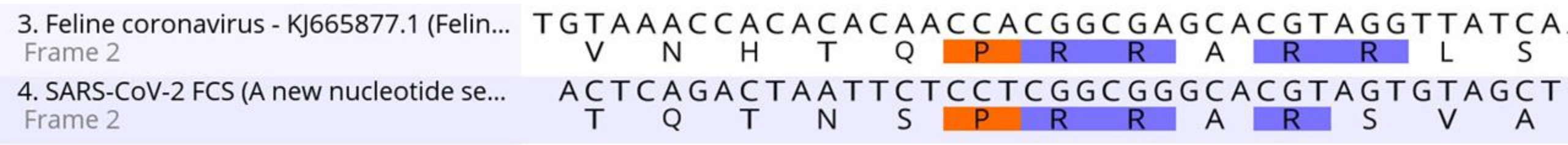
It could mean that the RR was copied from intermediate host RNA, where CGG is common.

It could have been copied out of frame.

It could have come from the complement strand (swapped and reversed, so: CCG CCG)

This feature can occasionally be found in natural viruses. Some MERS strains have a double CGG. Double CGG is found in the bat virus HKU9.

There's a feline coronavirus that uses PRRAR for a furin cleavage site and spells RR as CGGCGA. that's only one mutation away from what SARS-CoV-2 uses.



What are the odds?

Let’s assume that labs have no codon preference, so there are 1 in 36 odds they’ll pick a double CGG.

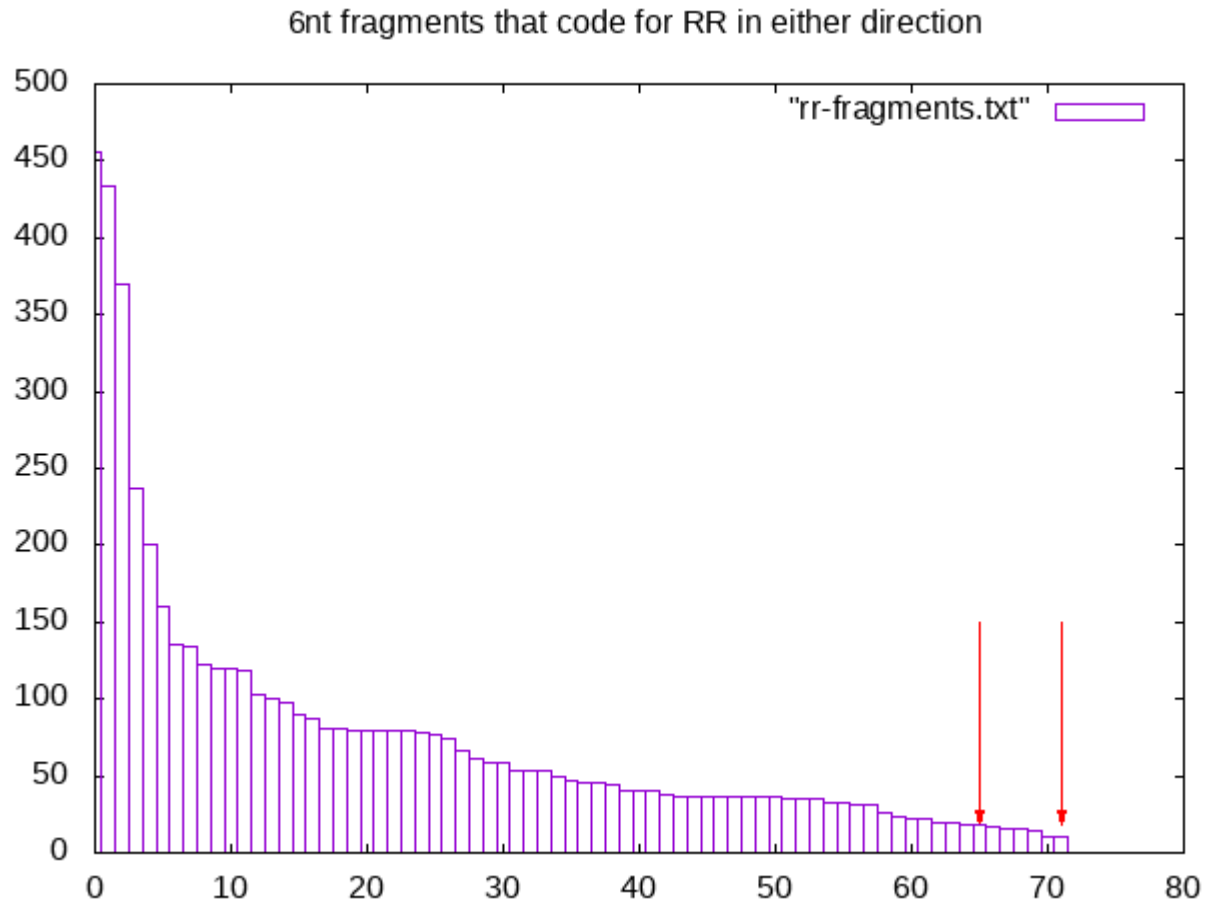
5% of R’s in Covid are CGG, or 1 in 20. Call it 1 in 400 to find 2 in a row.

An analysis from Guy Gadboit looked at 37 bat coronaviruses looking for the RR frequency, counting all possible frame shifts and complement strands.

He came up with 1 in 200 chance of finding the sequence.

Bayes factor between a lab leak and natural origin might be 200/36, or 5.5

Double CGG might be evidence for the lab leak, but it’s not strong evidence.



The insert could also have been copied from the host

Now CGG is around 20%, the odds of two in a row should be around 4%.

While the odds for a lab origin are 1 in 36, or 2.8%.

Another Guy Gadboit analysis scanned a bunch of genomes looking for double CGG's and says the odds are actually 1—2% for various humans and animals, not 4%.

So, the lab leak theory started out with double CGG as 1 in 1,000 odds favoring lab leak, but a proper analysis suggests it could be neutral between both origin theories or perhaps it leans weakly one way (maybe bayes factor of 2).

Human:		Bat:	
CGT	9%	CGT	8%
CGC	15%	CGC	16%
CGA	12%	CGA	12%
CGG	22%	CGG	20%
AGA	20%	AGA	21%
AGG	18%	AGG	23%

Also, “they” didn’t use the “human” choice for Alanine or Proline.

Alanine codon frequency,
SARS-CoV-2:

GCT 55%
GCC 14%
GCA 27%
GCG 3.8%

Alanine codon frequency,
humans:

GCT 27%
GCC 40%
GCA 23% ← SARS-CoV-2 has this one
GCG 11%

Proline codon frequency,
SARS-CoV-2:

CCT 46.3%
CCC 8.1%
CCA 41.3%
CCG 4.3%

Proline codon frequency,
humans:

CCT 30% ← SARS-CoV-2 has this one
CCC 31%
CCA 28%
CCG 11%

This is not the strongest argument, since the proline frequency is close for CCT and CCC.
Also, the insert is out of frame, so the A is already present for GCA. The bigger question is why it’s out of frame.
But you do have to wonder why the lab would optimize two arginines and nothing else.

Finally, it seems possible you could have made up a similar theory for CGA or CGC.

SARS-CoV-2:		Human:	
CGT	21.6%	CGT	9%
CGC	9.0%	CGC	15%
CGA	7.2%	CGA	12%
CGG	4.6%	CGG	22%
AGA	42.8%	AGA	20%
AGG	14.7%	AGG	18%

Remember the Texas Sharpshooter fallacy – it's easy to find patterns in data if you don't pre-specify which patterns you are looking for. You can always find something and make it look unlikely.

Those CGC and CGA theories wouldn't sound quite as good, because those aren't literally the most uncommon codons, but those ones are still about half as likely as in human DNA.

I'm sure that Robert Redfield could still tell Marjorie Taylor Greene that the virus uses "human codons", not "bat codons" and she still wouldn't know the difference.

For some reason, the double CGG works.

If the CGG CGG was unnatural, or nature selected against it, you would think that it would mutate into some other spelling of RR, over time.

You can change only one letter of CGG in 4 different ways and still get R. It could change to CGA, CGT, CGC, or AGG.

We've now had millions of covid cases. As of mid 2021, after 18 months of evolution, CGG CGG was still found in 99.85% of them. It's not mutating away from that. For some reason, this spelling works best.

One experiment tried mutating away the CGG codons. (from CCT CGG CGG to CCA AGG AGG)

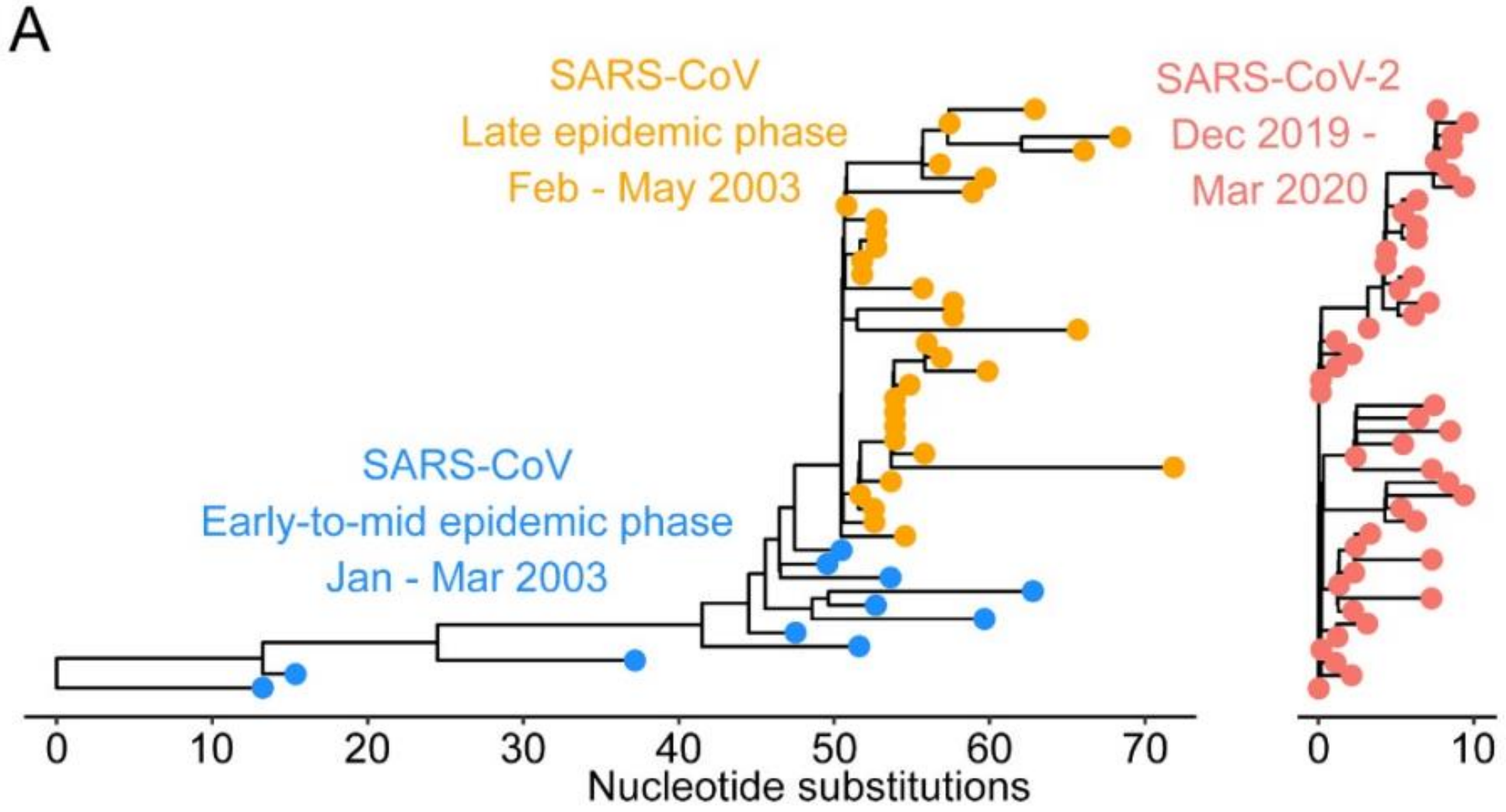
They found that the updated virus was less infectious than the one using CGG.

The issue has to do with protein folding. Their updated version made more spike protein, but the proteins weren't folded as accurately. The CGG codons slow down the protein translation, and that improves folding accuracy.

So that may be the reason. That's our best guess for why evolution came up with this.

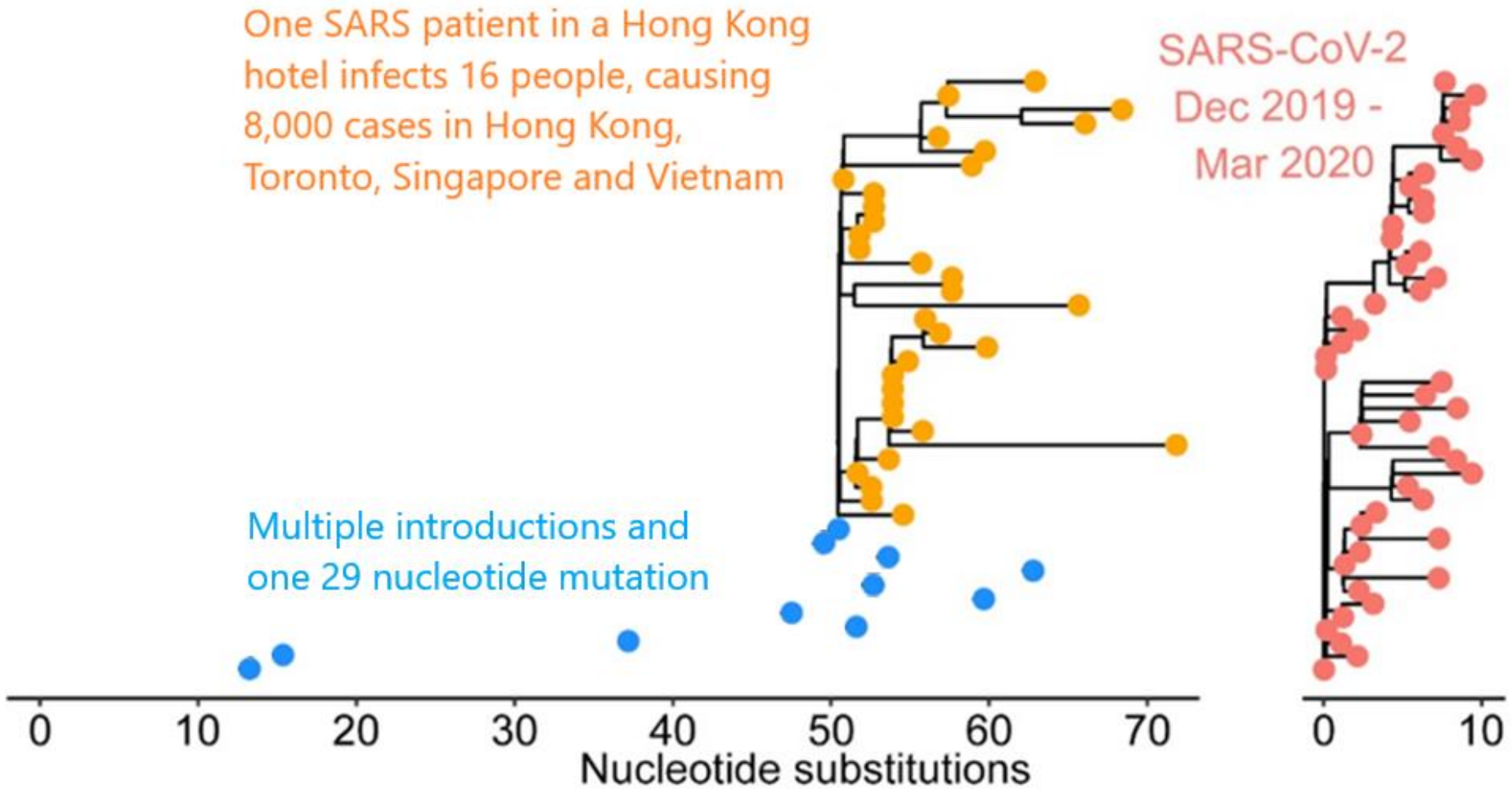
Is SARS-CoV-2 pre-adapted to humans?

A 2020 paper from Alina Chan argued that Covid mutated less than SARS, when it was introduced into humans, suggesting that maybe the virus was pre-adapted in a lab. But this graphic is actually misleading, for several reasons.

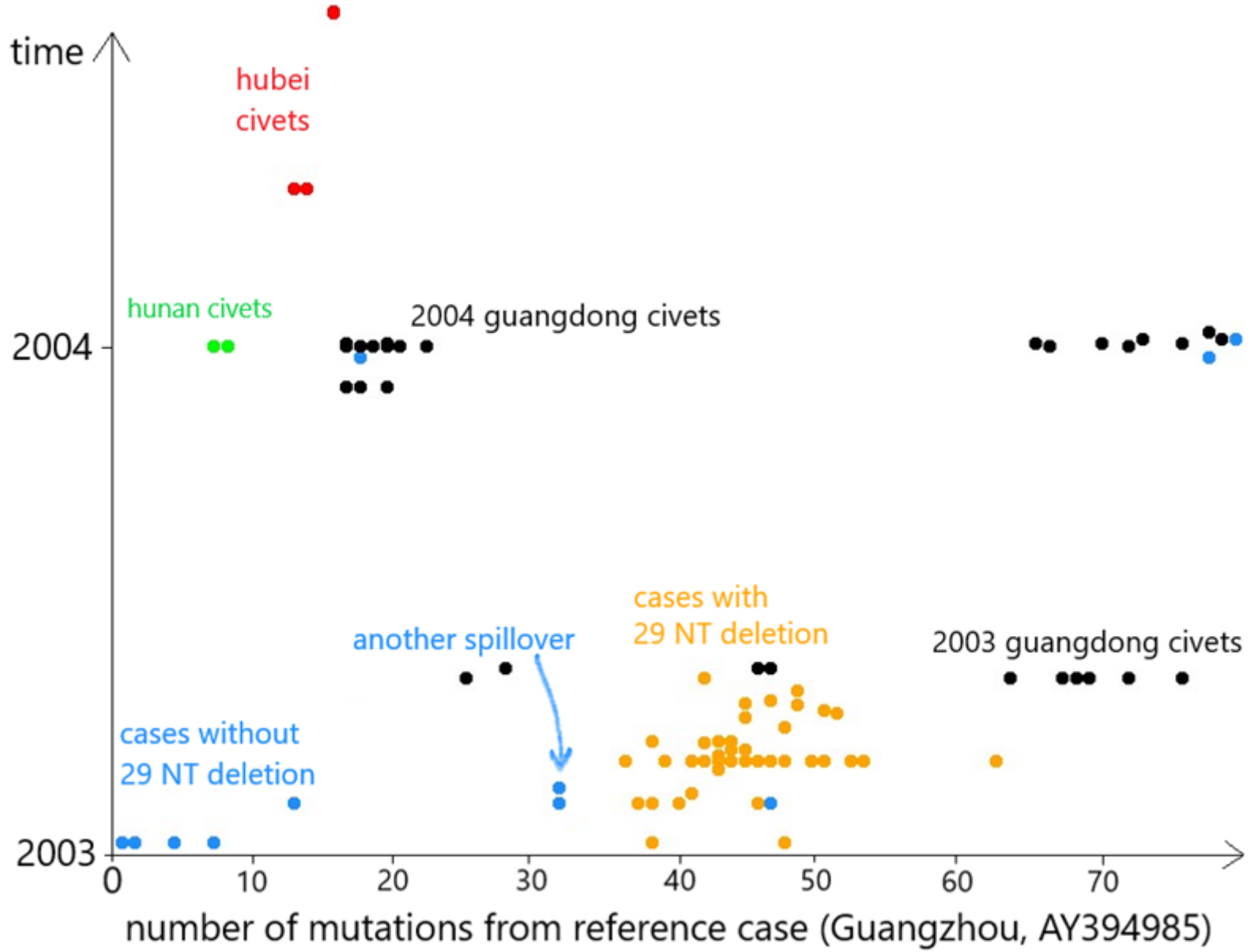


Misleading Figure from Zhan, Deverman, and Chan, 2020

- There were multiple spillovers of SARS. Separate outbreaks in different cities were not epidemiologically linked, some of the blue cases may be linked, others are not. The phylogeny is unclear.
- The SARS animal reservoir already had considerable diversity, so you're drawing the blue cases from a hidden larger tree.
- The orange cluster was a superspreading event which took SARS international, and those are all linked.
- There was one 29 nucleotide deletion sometime before that big cluster, making it look like the whole cluster has had a lot more single mutations from the earliest patient. That mutation was not an adaptation, it was actually detrimental, it just spread because of a founder effect. SARS sometimes sees deletions like this in the ORF8 gene.



I plotted mutations vs time for a set of SARS genomes, both humans and civets.
This should roughly reproduce Alina's diagram.
But you can also see the pre-existing diversity among civets, as well a few separate spillovers.



← Blue dots are another spillover

To be fair, Alina's analysis of SARS is not the worst I've seen.

Robert Redfield takes that prize. He testified to congress that SARS doesn't transmit human to human.

Image shamelessly stolen from slatestarcodex

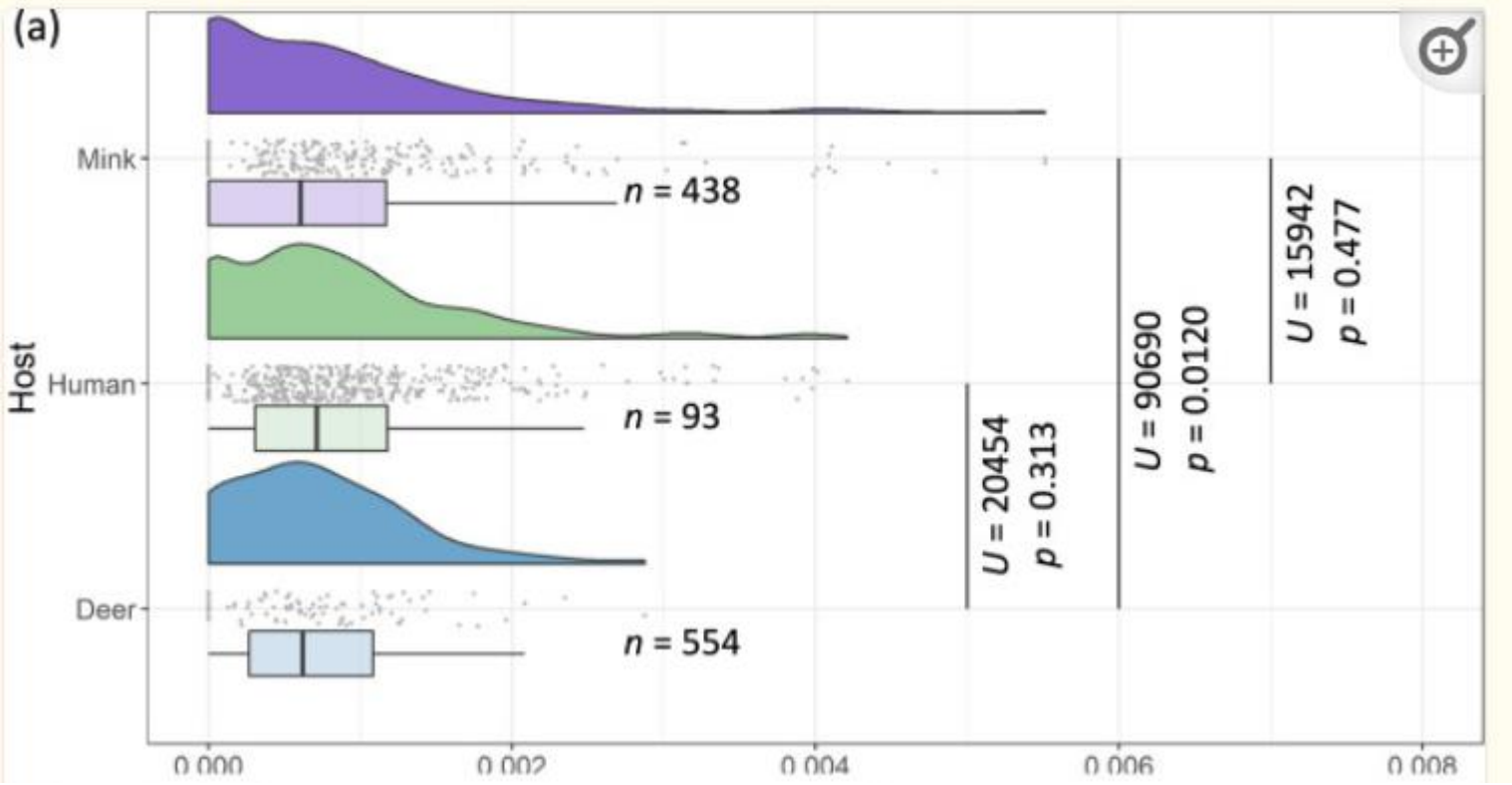
We can also just compare to when covid spilled over into mink or deer

If covid was pre-adapted to humans, then it should mutate faster when it jumped to other species.
Here are some relative mutation rates:

Early phase of Covid in humans: 37 mutations per year.
Alpha and delta strain, in humans: 18 mutations per year.

After covid spilled over into mink: 24 mutations per year.
After covid spilled over into deer: 36 mutations per year.

Mutation rate by species,
figure from Tan et al, 2022



Since Yuri and I disagreed on this, I did a quick review of all the mink evolution literature:

Table. Early evolutionary rates of SARS-CoV-2 in mink vs. humans				
Study	Host	Country	subst/site/year	mutations/year
Lu et al. (2021), *Nature Communications*	Mink	Netherlands (Cluster A)	1.41×10^{-3} (95% HPD of 1.2×10^{-3} to 1.75×10^{-3})	42.2 (35.8 to 52.3)
	Mink	Netherlands (Clusters A-E)	7.9×10^{-4} (95% HPD of 7.2×10^{-4} to 8.4×10^{-4})	23.6 (21.5 to 25.1)
Porter et al. (2023), *Virus Evolution*	Mink	Netherlands	1.83×10^{-3} (95% HPD of 1.3×10^{-3} to 2.41×10^{-3})	54.7 (38.9 to 72.1)
	Mink	Denmark	2.43×10^{-4} [95% HDP of 1.76×10^{-4} to 3.17×10^{-4}]	7.3 (5.3 to 9.5)
Tan et al. (2022), *Nature Communications*	Mink, deer, and humans	Denmark, Latvia, Netherlands, and Poland	~ $6.45 \pm 0.4 \times 10^{-4}$	~ 19.3 ± 1.2
McBride et al. (2023), *Nature Communications*	Human	China	1.3×10^{-3} (95% HPD of 1.1 to 1.6×10^{-3})	38.9 (32.9 to 47.8)
Li et al. (2020), *Journal of Medical Virology*	Human	China	1.19 to 1.31×10^{-3}	35.5 to 39.2
Chaw et al. (2020), *Journal of Biomedical Science*	Human	Worldwide	2.4×10^{-3} (95% HDP of 1.5×10^{-3} to 3.3×10^{-3})	71.7 (44.9 to 98.7)

Yuri's values? 6.59*10^-3 198

Yuri cited the rate from Porter et al 2023, which cites several very different numbers.

I haven't read the paper well enough to understand the range they're giving.

But it's pretty clear that he picked the highest possible value you can find in the literature, which is a clear outlier from the rest of the published research.

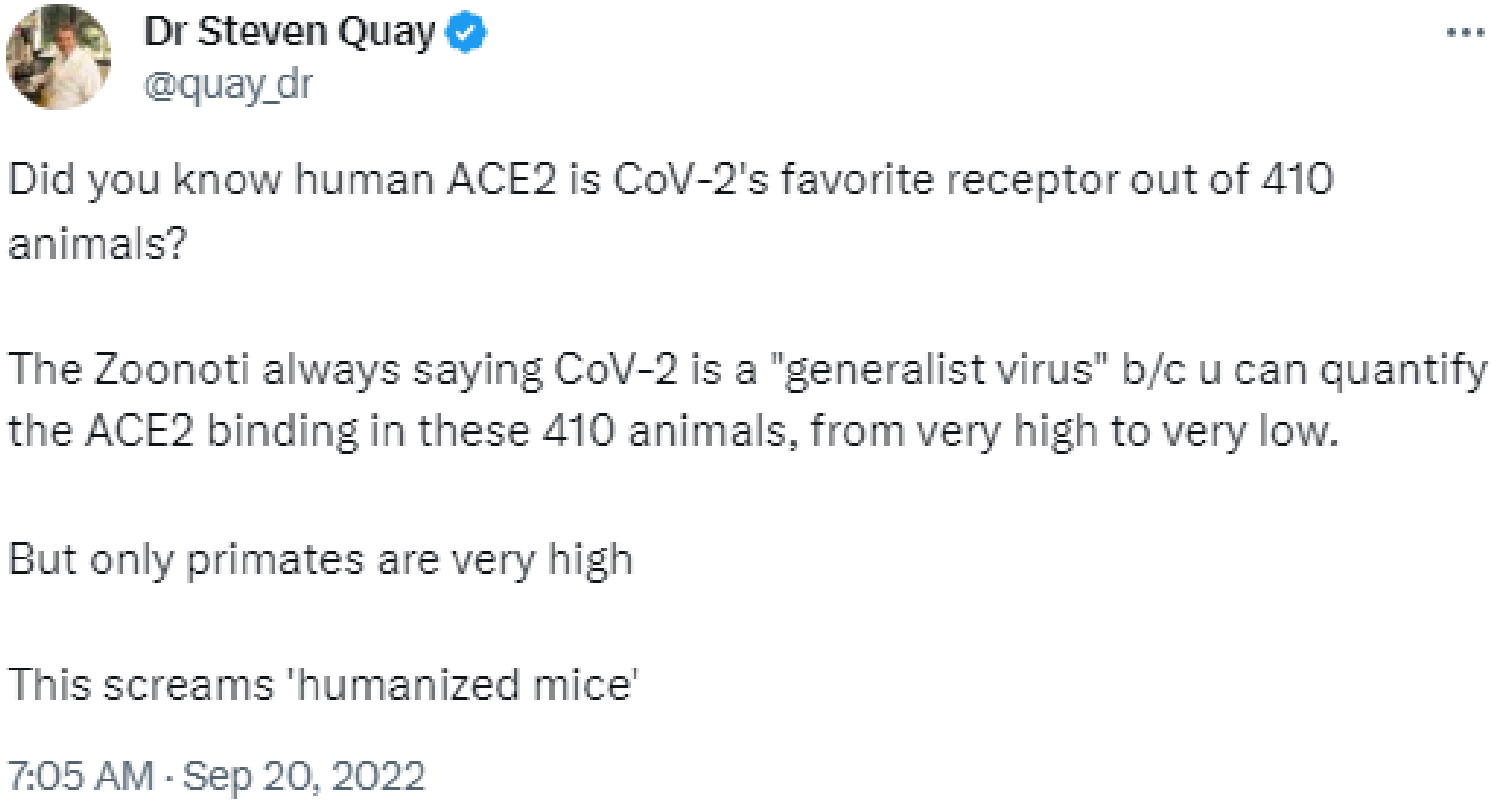
Table 3.

Estimates generated from local clock (FLC) models with a gamma prior on the clock rate.Estimates include the evolutionary rates (substitution/site/year) estimated for the whole phylogeny, and the Netherlands and Denmark foreground branches. The 95 per cent HPD interval is shown in brackets.

Model	Estimated evolutionary rate (mean)	Netherlands evolutionary rate	Denmark evolutionary rate
FLC (stem*)	4.54×10^{-4} [4.13×10^{-4}, 4.93×10^{-4}]	1.83×10^{-3} [1.3×10^{-3}, 2.41×10^{-3}]	2.43×10^{-4} [1.76×10^{-4}, 3.17×10^{-4}]
FLC (shared, stem*)	4.78×10^{-4} [4.36×10^{-4}, 5.2×10^{-4}]	6.59×10^{-3} [3×10^{-3}, 1.05×10^{-2}]	

Open in a separate window

Another theory holds that Covid is optimized for human ACE2:



Dr Steven Quay
@quay_dr

Did you know human ACE2 is CoV-2's favorite receptor out of 410 animals?

The Zoonoti always saying CoV-2 is a "generalist virus" b/c u can quantify the ACE2 binding in these 410 animals, from very high to very low.

But only primates are very high

This screams 'humanized mice'

7:05 AM · Sep 20, 2022

That 2020 study [ranked similarity to human ACE2] based on 25 amino acids in human ACE2 (contact residues).

The rank is how many of these 25 they share with humans.

Humans are going to show up in the top position by design!

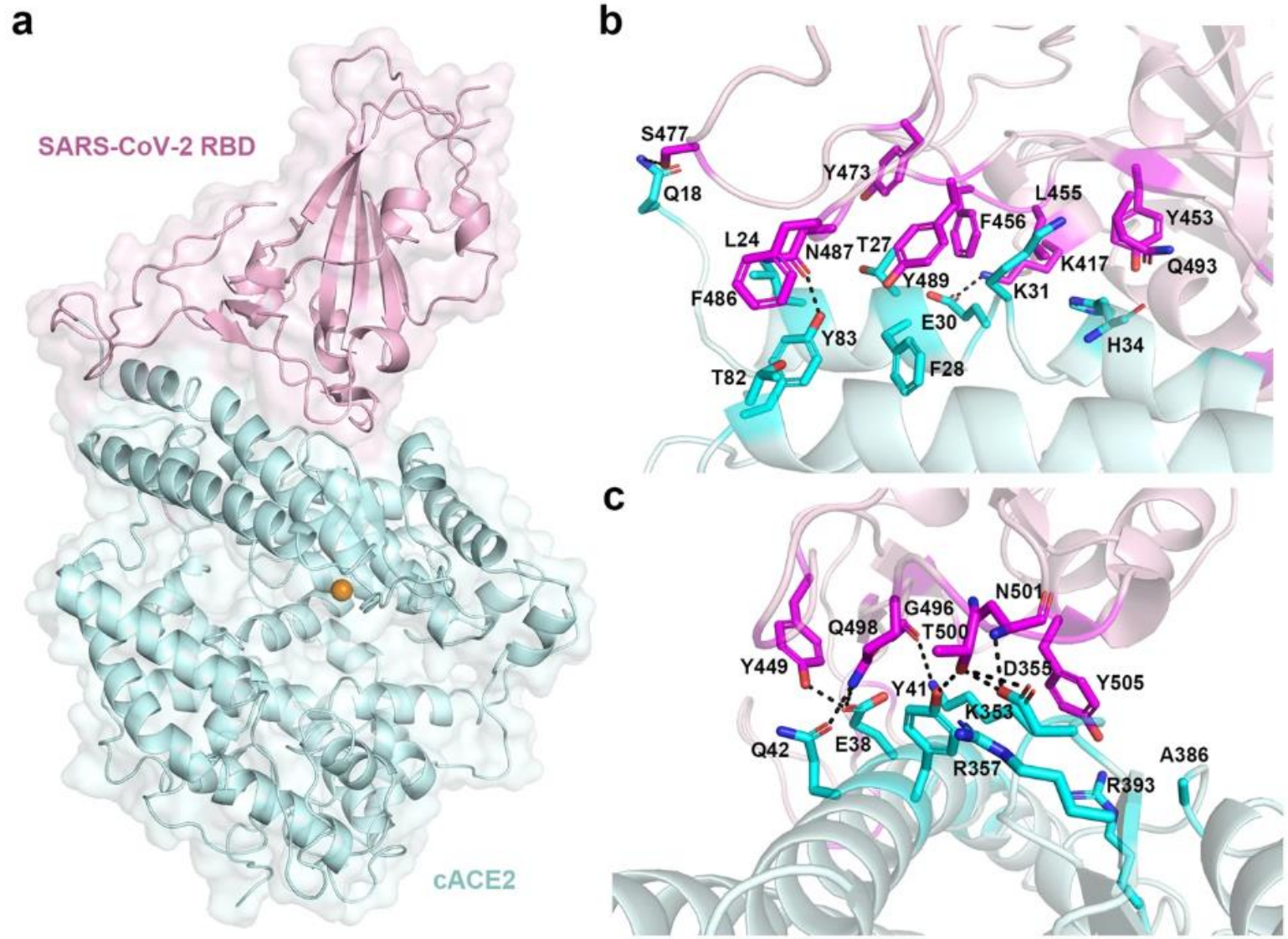
conservation of ACE2 and its potential to be used as a receptor by SARS-CoV-2. We designed a five-category binding score based on the conservation properties of 25 amino acids important for the binding between ACE2 and the SARS-CoV-2 spike protein. Only mammals fell into the medium to very high categories and only catarrhine primates into the very high category, suggesting that they are at high risk for SARS-CoV-2 infection. We employed a protein structural analysis to qualitatively assess whether amino acid changes at variable residues would be likely to disrupt ACE2/SARS-CoV-2 spike protein binding and found the number of predicted unfavorable changes significantly correlated with the binding score. Extending this analysis to human population data, we found only

	ID	S19	Q24	T27	F28	D30	K31*	H34	E35*	E37	D38	Y41	Q42	L45	N53*	L79	M82*	Y83	N90*	N322*	N330	K353*	G354	D355	R357	R393
VERY HIGH																										
Homo sapiens (Human)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Gorilla gorilla gorilla **(Western lowland gorilla)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nomascus leucogenys **(Northern white-cheeked gibbon)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pongo abelii **(Sumatran orangutan)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Macaca fascicularis (Crab-eating macaque)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mandrillus leucophaeus **(Drill)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Nasalis larvatus **(Proboscis monkey)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pan paniscus **(Bonobo)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pan troglodytes **(Chimpanzee)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Piliocolobus tephrosceles **(Ugandan red colobus)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pygathrix nemaeus **(Red-shanked douc)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rhinopithecus roxellana **(Golden snub-nosed monkey)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Chlorocebus sabaeus (Green monkey)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Erythrocebus patas (Patas monkey)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Macaca mulatta (Rhesus macaque)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Papio anubis (Olive baboon)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Theropithecus gelada (Gelada)	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Cercocebus atys **(Sooty mangabey)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Macaca nemestrina **(Southern pig-tailed macaque)**	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HIGH																										
Colobus angolensis **(Angola colobus)**	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.
Propithecus coquereli **(Coquerel's sifaka)**	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
Cricetomys gambianus (Gambian pouched rat)	22	.	.	A	.	.	Q	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.
Cricetulus griseus (Chinese hamster)	22	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	N	.	.	H	.	.	.	.	.	.
Ctenodactylus gundi (Common gundi)	22	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	N	.	T	.	.	.	.	.	.	.
Delphinapterus leucas (Beluga whale)	22	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	I	T	.	.	.	.	.	.	.	.	.
Eulemur flavifrons **(Blue-eyed black lemur)**	22	.	E	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
Indri indri **(Indri)**	22	.	.	.	.	.	N	Q	.	.	.	.	.	.	.	.	N	.	.	.	.	.	.	.	.	.
Monodon monoceros (Narwhal)	22	.	.	.	.	Q	.	.	.	.	.	.	.	.	.	I	T	.	.	.	.	.	.	.	.	.

Contact residues refer to the parts of the spike that interface with ACE2

Fig. 4: The complex structure between cACE2 and SARS-CoV-2 RBD.

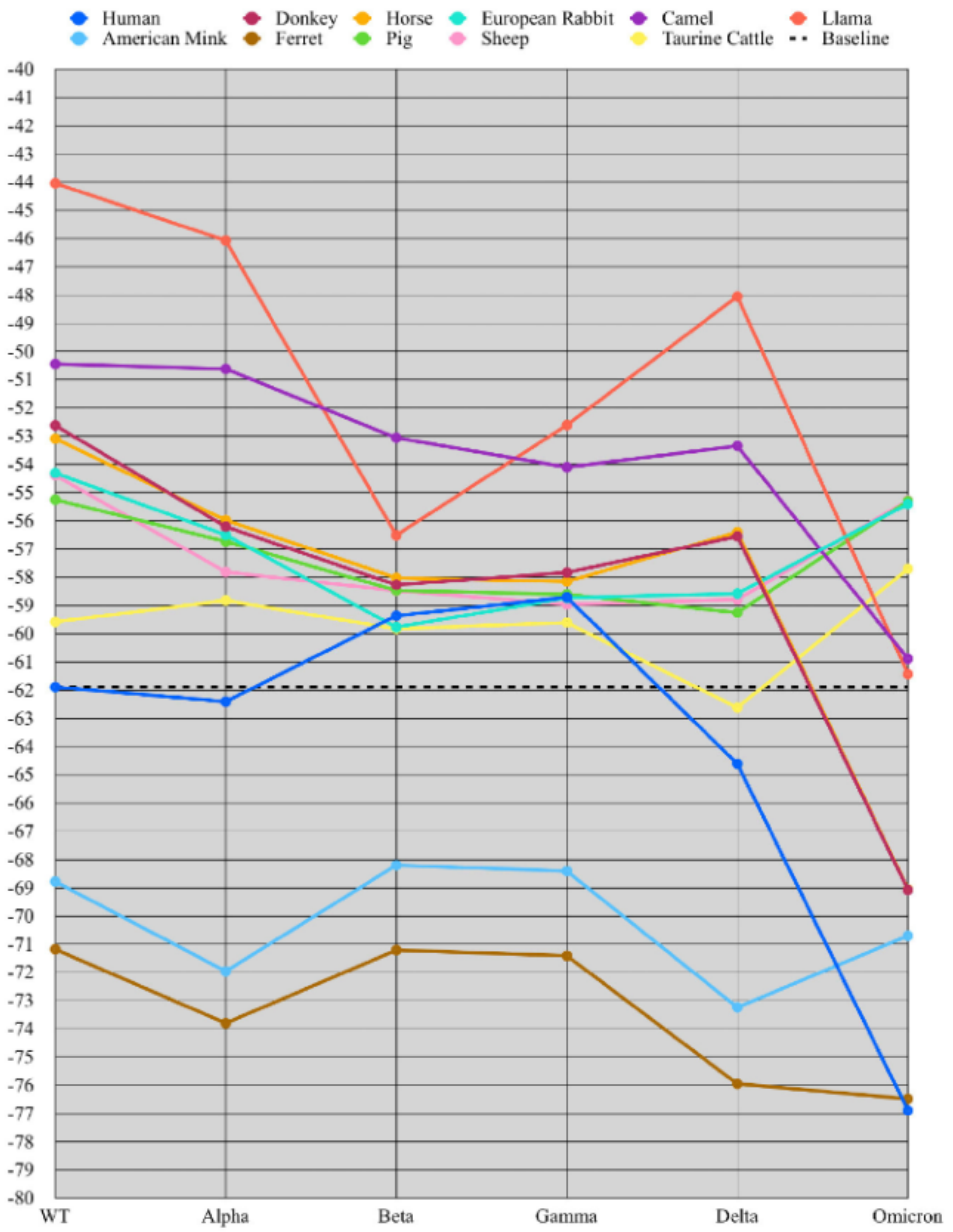
From: Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2



Better studies predict the binding by species, they don't just count similar amino acids.

One study says that Covid binds best to ferret and mink ACE2:

Species		WT	Alpha	Beta	Gamma	Delta	Omicron
Human	*point mut.*	-61.89±0.14	**-62.41±0.44**	-59.36±0.64	-58.71±0.90	**-64.61±0.92**	**-76.89±2.58**
	hom. mod.		**-64.93±0.12**	-61.40±0.03	-60.28±0.38	**-65.74±0.34**	-59.15±1.23
Donkey		-52.62±0.95	-56.21±1.47	-58.27±0.60	-57.83±0.60	-56.55±0.61	**-69.08±1.35**
Horse		-53.09±0.88	-55.98±1.49	-58.02±0.83	-58.15±0.68	-56.4±0.89	**-69.06±1.41**
European rabbit		-54.31±2.2	-56.50±0.62	-59.77±0.73	-58.73±0.27	-58.58±1.93	-55.40±2.12
Camel		-50.44±1.94	-50.62±0.77	-53.06±0.11	-54.1±0.61	-53.35±1.52	-60.89±1.84
Lama		-44.04±0.63	-46.07±1.26	-56.51±1.36	-52.61±1.2	-48.05±1.54	-61.43±1.72
American mink		**-68.78±0.2**	**-71.96±0.96**	**-68.2±0.00**	**-68.41±1.28**	**-73.24±0.60**	**-70.7±1.34**
Ferret		**-71.18±0.29**	**-73.8±1.81**	**-71.21±1.15**	**-71.41±0.79**	**-75.95±0.88**	**-76.48±3.03**
Pig		-55.25±1.04	-56.73±0.99	-58.47±0.78	-58.61±0.38	-59.25±1.14	-55.3±1.93
Sheep		-54.39±1.87	-57.81±1.59	-58.51±0.51	-58.96±0.05	-58.79±0.56	-55.28±1.35
Taurine cattle		-59.58±0.95	-58.82±0.32	-59.83±0.32	-59.61±0.34	**-62.62±0.58**	-57.70±1.95
Chicken		-42.42±0.62	-45.14±1.11	-50.05±0.74	-49.54±1.00	-44.16±0.89	**-64.83±1.5**
Helmeted guineafowl		-44.47±0.11	-47.27±0.35	-50.88±0.33	-51.23±0.83	-48.77±0.01	**-67.98±1.07**
South African ostrich		-44.51±1.17	-42.3±1.15	-47.85±0.89	-47.82±0.78	-47.27±0.02	-59.95±1.51
Mallard		-43.05±1.27	-43.29±1.41	-46.01±0.94	-46.29±1.84	-46.3±1.19	-61.22±1.84



Figures from Peka et al 2023

These computational models aren't all consistent.

Another model says that Covid binds better to bamboo rat ACE2 than human ACE2, but says there's no binding to mink ACE2 (neovison vison)

Table 1

Comparison of binding free energies (kcal/mol) for the ACE2 molecules from pets and wild animals.

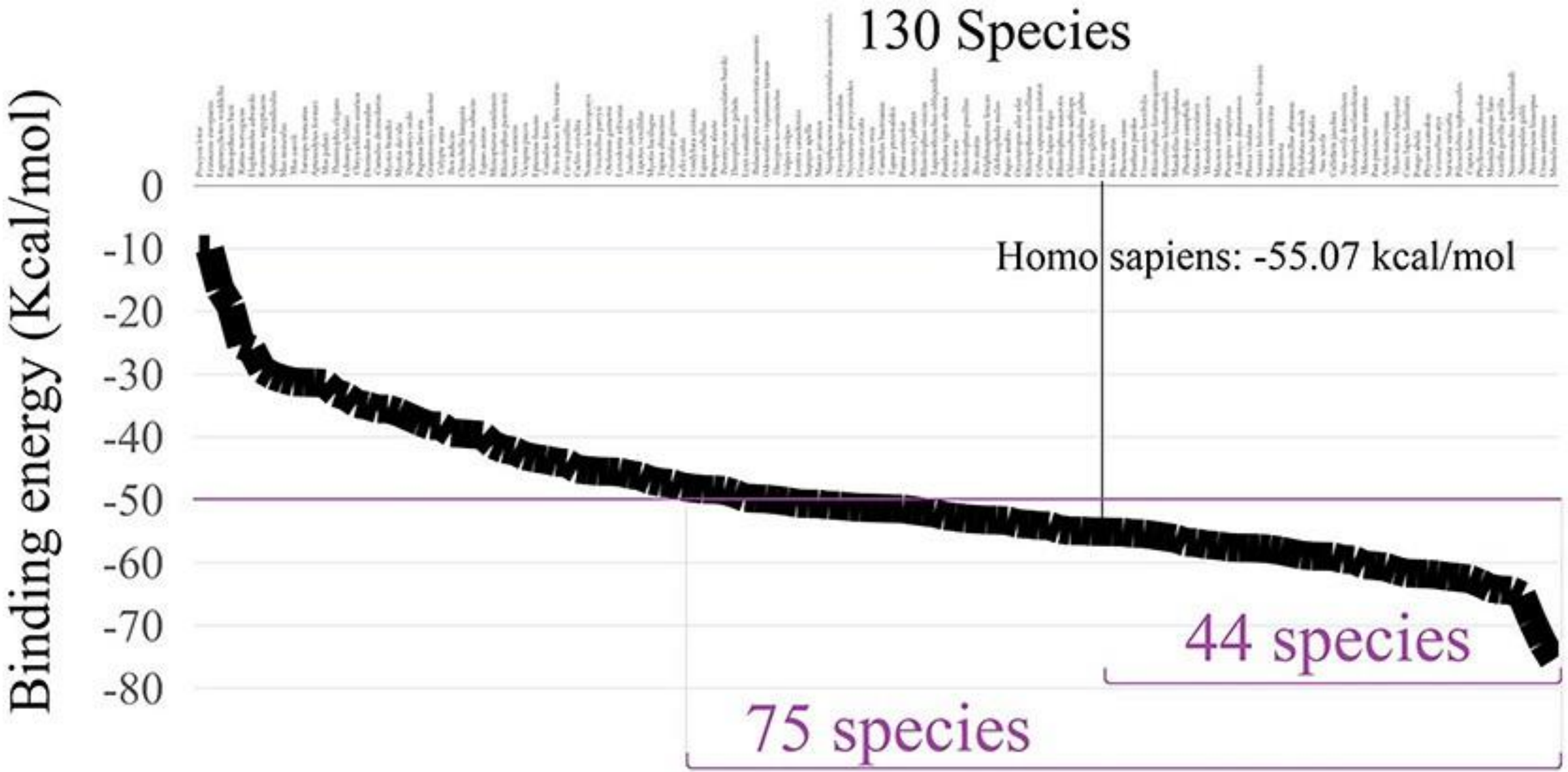
Organism	*Important ACE2 residues at the binding interface with the spike	ΔG (kcal/mol)
Human	^{19}STIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNT^{52}	-60.64 ± 3.10
Bamboo rat	^{19}STTEEQAKTFLDKFNQEAEELSYQSALASWNYNT^{52}	-65.04 ± 3.65
Mole	^{19}LTIEEQAKTFLDKFNQEAEDLSYQNSLASWNYNT^{52}	-59.67 ± 2.95
Vole	^{19}SIIEEDAKAFLDKFNQEAEDLSYQSALASWNYNT^{52}	-57.18 ± 2.57
Mus pahari	^{19}SLTEENAKTFLNKFNQEAEDLSYQSSLASWNYNT^{52}	-56.88 ± 3.13
Palm civet	^{19}STTEGQAKTFLEKFNHEAEDLSYQSSLASWNYNT^{52}	-56.11 ± 2.42
Rat	^{19}SLIEEKAESFLNKFNQEAEDLSYQSSLASWNYNT^{52}	-56.01 ± 2.55
Mus musculus	^{19}SLTEENAKTFLNNFNQEAEDLSYQSSLASWNYNT^{52}	-55.59 ± 2.79
Pangolin	^{19}STSDEEAKTFLEKFNSEAEELSYQSSLASWNYNT^{52}	-54.78 ± 2.22
Mus caroli	^{19}SLTEENAKTFLNKFNQEAEDLSYQSSLASWNYNT^{52}	-50.90 ± 2.62
Snake	^{44}QDETKVATKFLEQFDARATDLYYNASIASWDYNT^{77}	> 400
Neovison vison	–	> 800

Open in a separate window

*positions of residues in the ACE2 sequence are denoted with numbers before the first and after the last amino acid residues. Sequence accession numbers are included in Additional file 6: Table S6.

Table from Chen et al 2022

A third model claims that 44 species all have better Covid ACE2 binding than humans



Computational studies can only get you so far.

That actually argues against the artificial design of SARS-CoV-2 – current models are not advanced enough to be able to predict the creation of SARS-CoV-2.

You wouldn’t know how to engineer the best receptor binding domain, based on these conflicting studies. You would instead need to find a natural virus with efficient binding and start working from that.

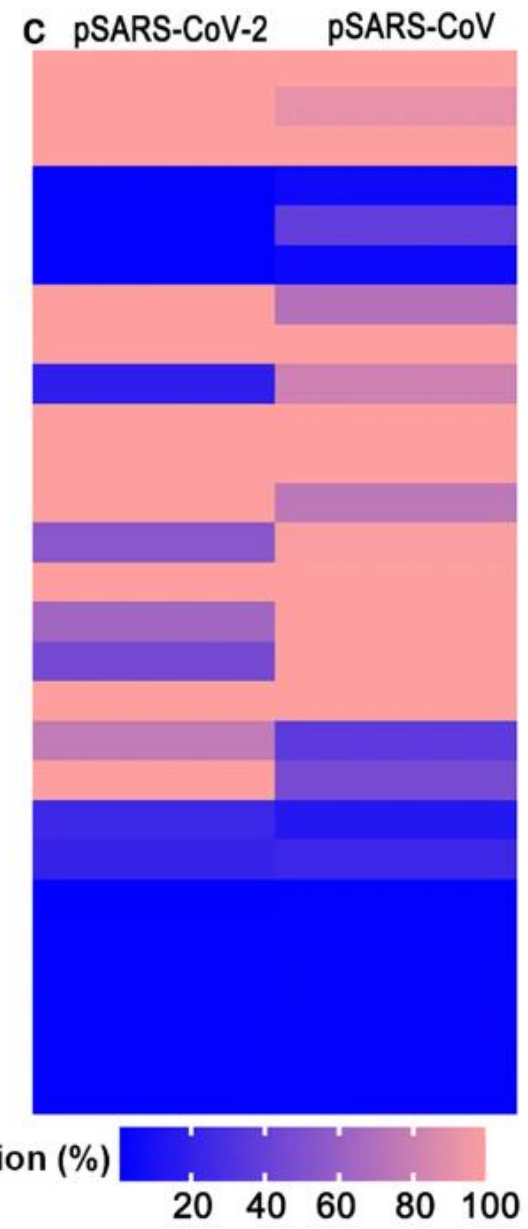
In vitro ACE2 binding studies are also inconsistent:

Wu et al 2020

"We found that the monkey, rabbit, Malayan pangolin, cat, fox, dog, raccoon dog, pig and bovine ACE2s supported pseudotyped SARS-CoV-2 transduction as good as hACE2… Consistent with the binding affinities with SARS-CoV-2 RBD, the bat ACE2s, which could initiate the entry of SARS-CoV-2 pseudoviruses at a low level are from little brown bat and fulvous fruit bat, but not from greater horseshoe bat, Chinese horseshoe bat, or least horseshoe bat. Although the civet ACE2 displays no detectable binding with the SARS-CoV-2 RBD, it could still mediate the transduction of pseudotyped SARS-CoV-2."

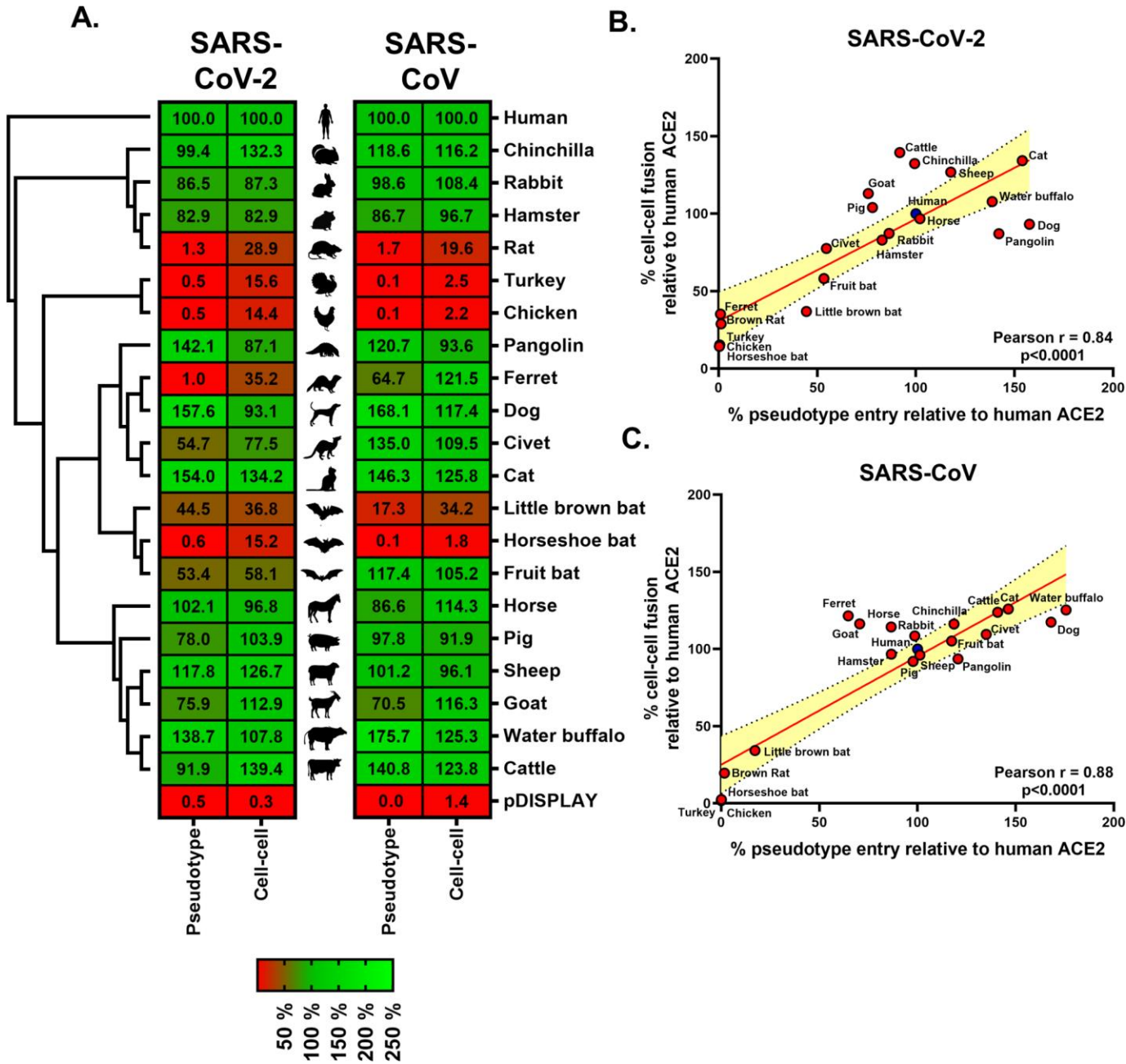
b

	SARS-CoV-2 RBD	SARS-CoV RBD
Human	21.73 ± 1.54	43.27 ± 6.43
Monkey	21.73 ± 1.97	69.07 ± 4.31
Rabbit	76.20 ± 20.96	74.83 ± 3.06
Guinea pig	--	--
Mouse	--	--
Rat	--	--
Malayan pangolin	66.80 ± 8.22	226.67 ± 54.28
Cat	85.70 ± 19.16	69.90 ± 0.75
Civet	--	121.00 ± 16.37
Fox	40.63 ± 2.63	17.85 ± 1.63
Dog	98.87 ± 25.67	19.97 ± 0.67
Raccoon dog	96.40 ± 18.16	178.87 ± 175.05
Horse	132.63 ± 31.05	141.33 ± 20.26
Pig	47.63 ± 8.10	35.83 ± 9.05
Wild Bactrian camel	272.33 ± 161.85	69.10 ± 3.82
Alpaca	16520 ± 19147	54.90 ± 2.15
Bovine	73.67 ± 53.27	89.77 ± 65.86
Goat	157.00 ± 23.58	308.00 ± 174.26
Sheep	137.03 ± 44.25	157.33 ± 26.50
Little brown bat	312.33 ± 90.79	12380 ± 17046
Fulvous fruit bat	1132 ± 338	412.33 ± 58.5
Greater horseshoe bat	--	--
Chinese horseshoe bat	--	--
Least horseshoe bat	--	--
European hedgehog	--	--
Lesser hedgehog tenrec	--	--
Chicken	--	--



[Conceicao et al 2020]

"we observed that pangolin, dog, cat, horse, sheep, and water buffalo all sustained higher levels of entry than was seen with an equivalent human ACE2 construct… In contrast, all 3 bat ACE2 proteins we analysed (fruit bat, little brown bat, and horseshoe bat) sustained lower levels of fusion than was seen with human ACE2…"



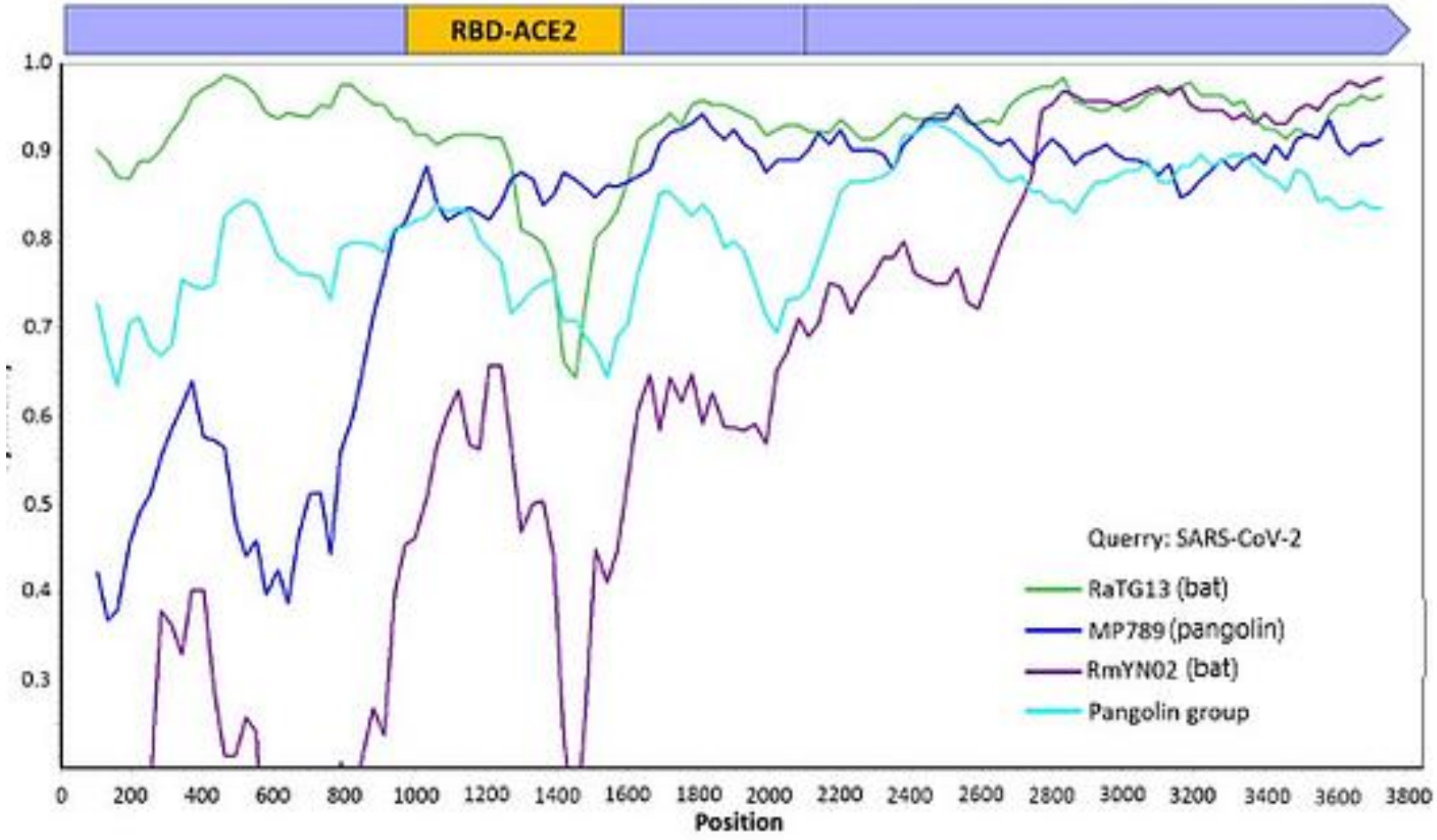
If even in vitro experiments can’t give consistent results, that also questions the artificial design of SARS-CoV-2.

You could sample a lot of different viruses and grow them all in vitro, maybe that would show that one worked well to infect people.

But it would be hard to design a virus that binds “optimally for humans”.

ACE2 binding and the pangolin viruses

SARS-CoV-2 has a spike protein 90+% similar to RATG13, except for a drop in similarity at the Receptor Binding Domain.
SARS-CoV-2 binds well to human ACE2, because of that RBD.
At first, some people also thought that was a sign that SARS-CoV-2 was engineered to optimize this.
In early 2020, a pangolin virus was found with higher similarity in the RBD, but poor similarity elsewhere in the spike.
This created the “pangolin chimera” theory.



Similarity across a portion of the genome. Figure from [Flores-Alanis et al, 2020]

That lead to the lab chimera theory:

In his [2020 medium post], Yuri suggested that the lab had combined bat and pangolin viruses:

> So it was then, in pursuit of arguments against the virus's lab-madeness, that I got infected by the virus of doubt. What was the source of my doubts? The fact that the deeper you dive into the research activities of coronavirologists over the past 15–20 years, the more you realize that creating chimeras like CoV2 was commonplace in their labs. And **CoV2 is an obvious chimera (though not nesessarily a lab-made one), which is based on the ancestral bat strain RaTG13, in which the receptor binding motif (RBM) in its spike protein is replaced by the RBM from a pangolin strain, and in addition, a small but very special stretch of 4 amino acids is inserted, which creates a furin cleavage site that, as virologists have previously established,**

> The authors go on to put forth a conjecture that this may be the result of convergent evolution, in other words, that CoV2 and the pangolin strain came to possess identical RBMs each in their own way, rather than through recombination between common ancestors. Because it would have required a rather unique recombination event — as if someone cut out a precise RBM segment from a pangolin strain and used it to replace the RBM in RaTG13. Talk about Intelligent Design!

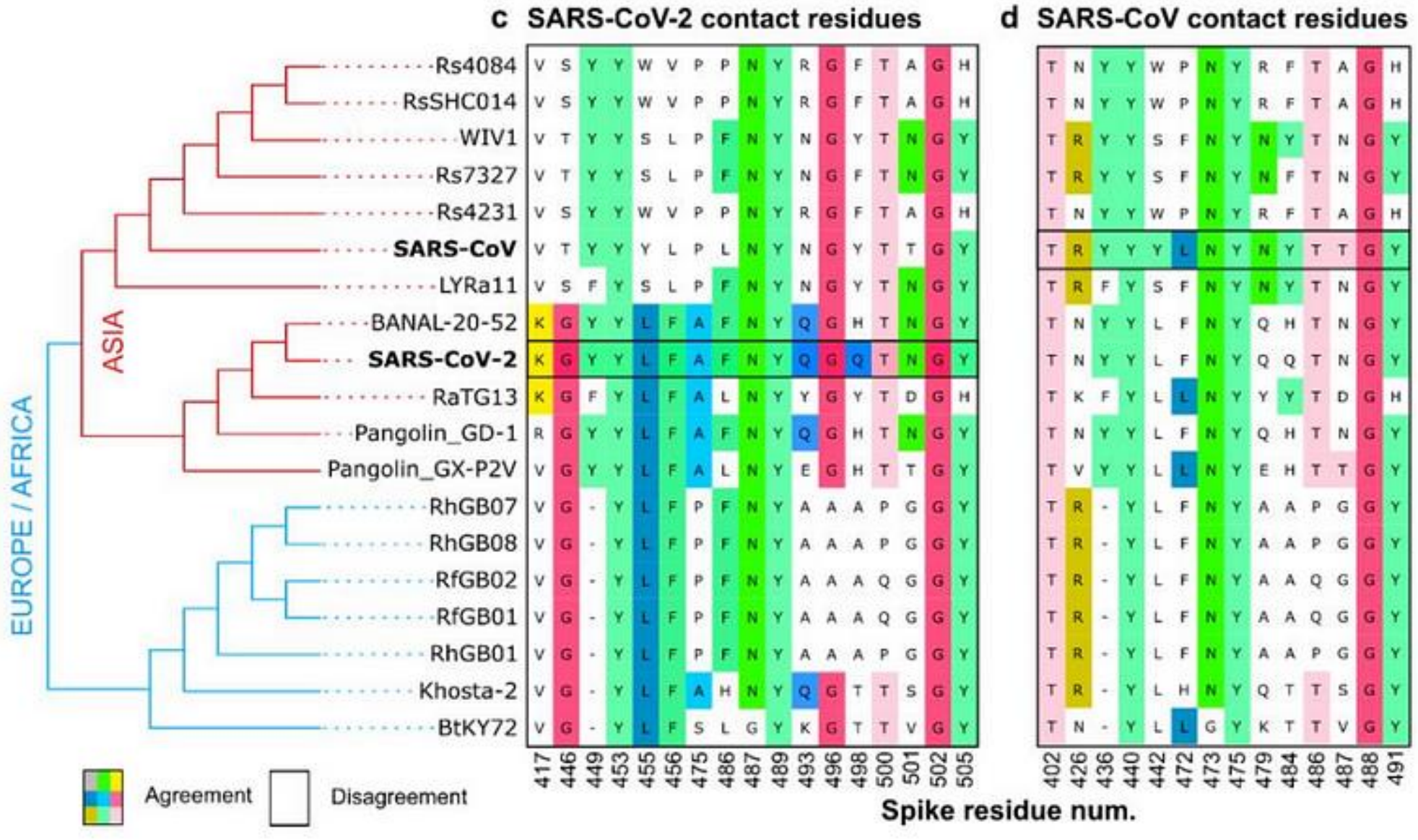
> Could researchers, having received coronavirus samples from pangolins that were intercepted by customs in March 2019, then want to check whether the RBM in pangolin strains can bind to the human ACE2 receptor? And could such researchers also decide to throw an extra furin site in the mix?

In 2021, we found a bat virus with a near identical receptor binding domain to SARS-CoV-2

BANAL-20-52 also binds well to human ACE2

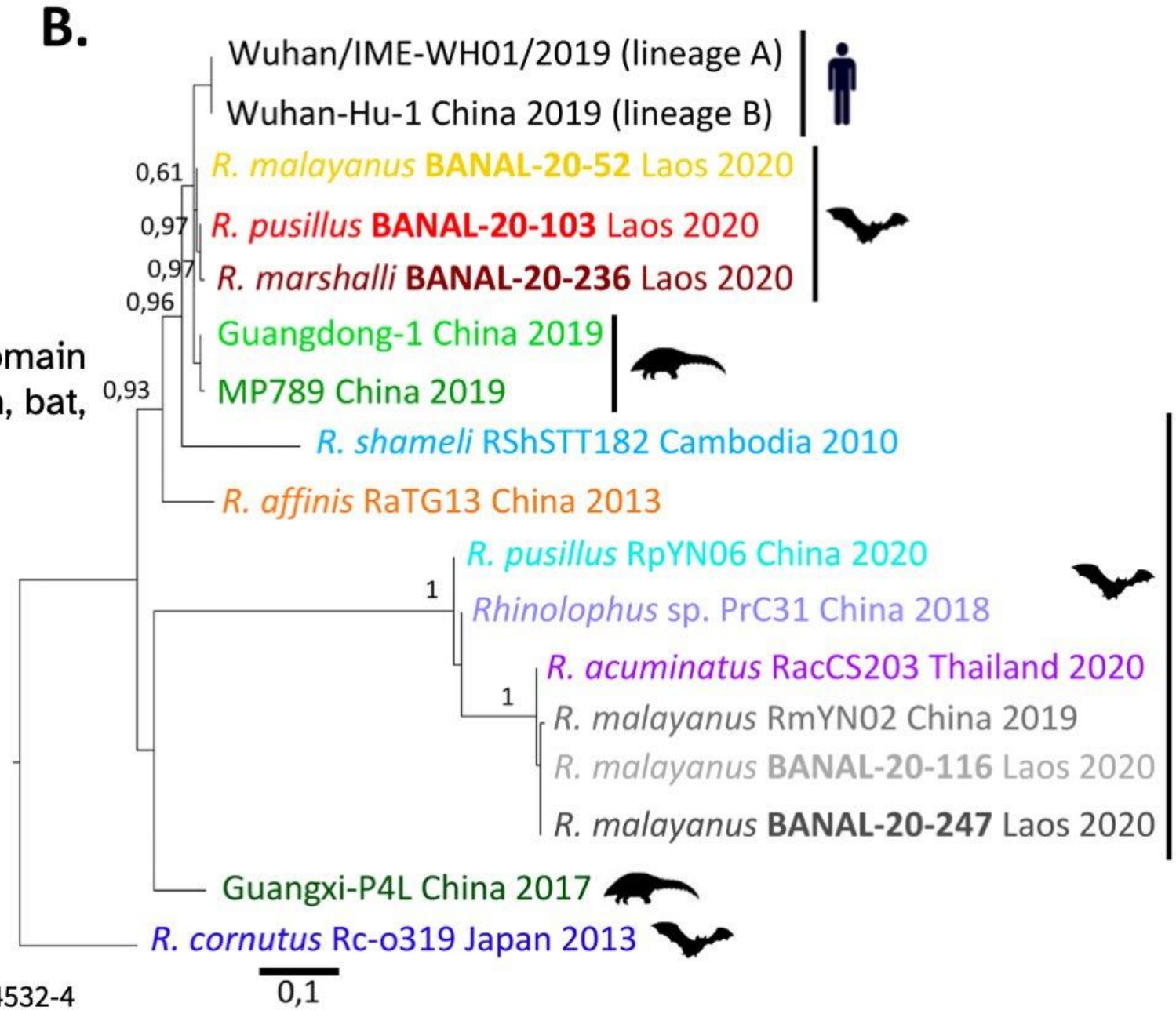
Contact residues [are nearly identical](#) between BANAL-20–52 and SARS-CoV-2.

It's even closer than the Pangolin viruses were.



The Receptor binding domain of SARS-CoV-2 is closer to the Laotian bat viruses than it is to the Pangolin viruses:

Phylogenetic analysis of the protein sequence of the receptor-binding domain of Laotian and representative human, bat, and pangolin sarbecoviruses.



https://www.nature.com/articles/s41586-022-04532-4

Two bat viruses have now been found with an RBD more than 90% similar to SARS-CoV-2.
Both should bind well to human ACE2, without any modification.
One was found in Laos, the other in Yunnan province.
The virus found in Yunnan had only “five amino acid differences between its receptor-binding domain sequence and that of the earliest sequences of SARS-CoV-2”.

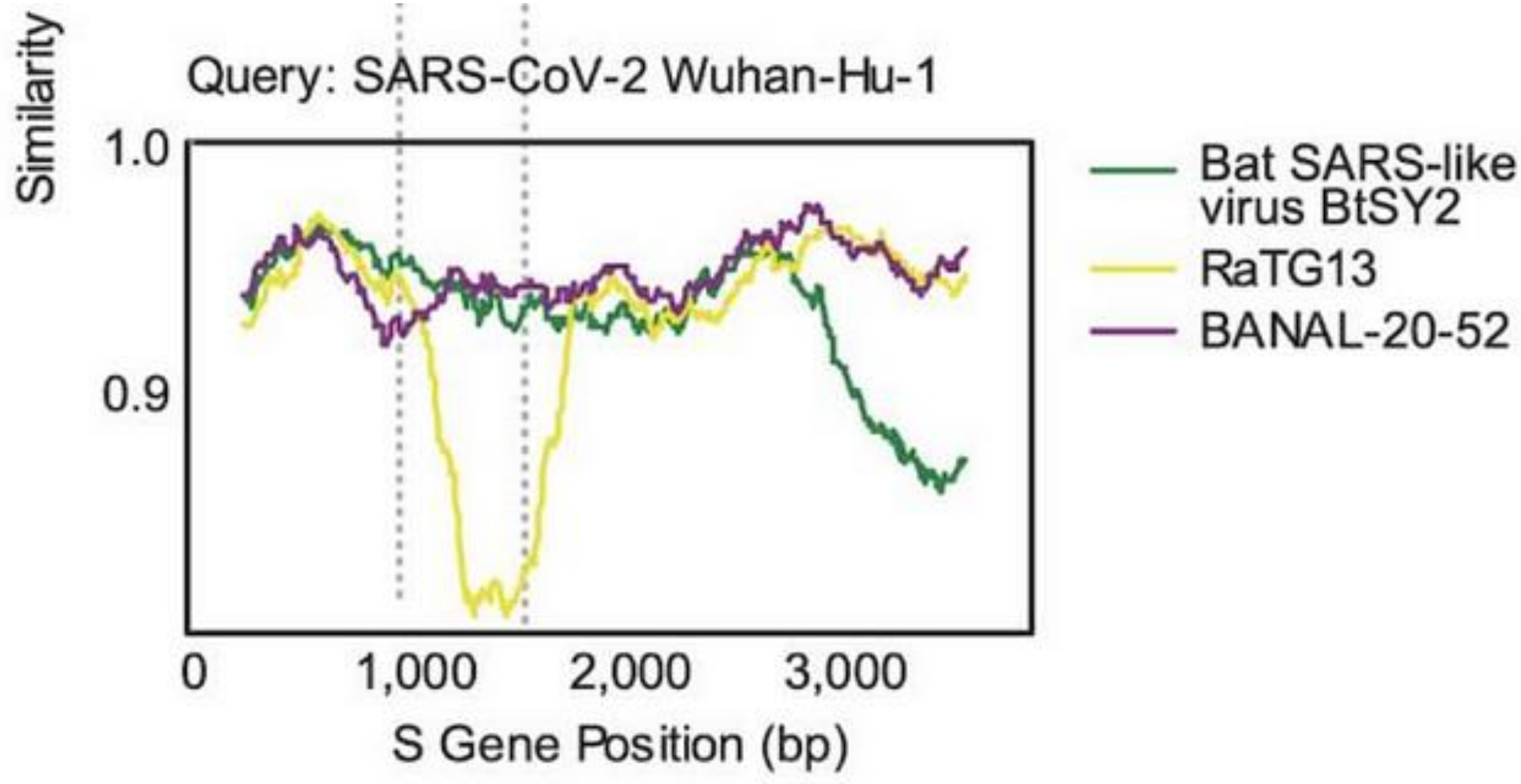


Figure from Wang et al, 2022

You'd think that would make people shift their beliefs towards a natural origin.

Instead, this was the lab leak response to the discovery:

They said that just proves the Wuhan lab must have secretly had a virus like this.

Matt Ridley speculated that the Wuhan researchers must have previously gone to Laos earlier and found this.

This is standard conspiratorial thinking:

Absence of evidence is proof of the conspiracy, but presence of evidence is also proof of the conspiracy.
Lab leak theorists believe that:

- If the virus binds well to ACE2, it must be designed.
- If a pangolin virus looks similar, that just means the lab had the pangolin virus and combined it.
- If a bat virus binds better still, that just means the lab must have had that bat virus.
- If we find an insertion similar to the furin cleavage site in a bat virus, that means the lab had that, too.

Yuri has also moved on to arguing that the pangolin viruses never existed.

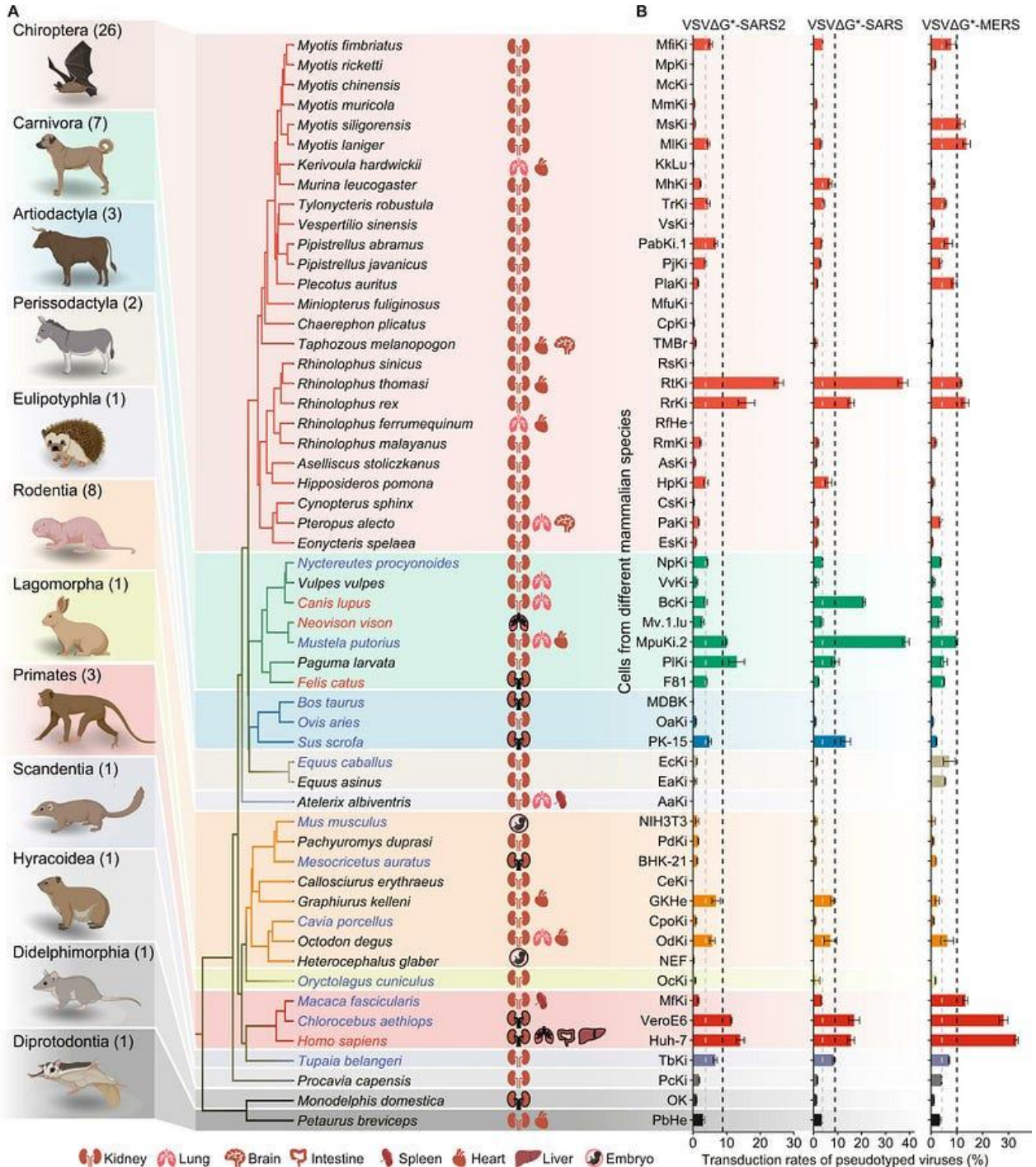
Here's a quote from a 2023 [Matt Ridley article]:

> Dr Zhang works not in Wuhan but in Guangdong at the very institute that announced in 2020 it had found traces of a similar virus in a smuggled pangolin, confiscated before the pandemic began in 2019. This led to brief excitement in early 2020 that pangolins might be the intermediate species that transmitted Covid to people. The lack of pangolins in the Wuhan market and the genetic distance of the 'pangolin' virus from Sars-CoV-2 soon scotched that theory.
>
> But it remains an enigma: how did a pangolin in Guangdong pick up a bat virus from distant Yunnan? Yuri Deigin, a Russian-Canadian biotech expert, thinks the answer is staring us in the face. The same lab that sequenced the samples from pangolins in 2019 was also, we know, sequencing samples from Dr Zhang's malayanus bats around the same time. Maybe the pangolin never had a coronavirus (it had much higher doses of sendai virus), but the sequencing machine was contaminated by malayanus bat samples. Such contamination occurs frequently in such machines and is difficult to prevent.

SARS-CoV-2 is not optimized for humans

We don’t know the intermediate host, yet.

But SARS-CoV-2 grows well in cells from many species, including 2 bats, civets, raccoon dogs, and mustelids.



Many proposed lab methods to create SARS-CoV-2 would fail

On page 35 of the DEFUSE grant proposal, they write:

"We will analyze all SARSr-CoV S gene sequences for appropriately proteolytic cleavage sites in S2 and for the presence of potential furin cleavage sites. SARSr-CoV S with mismatches in proteolytic cleavage sites can be activated by exogenous trypsin or cathepsin L. Where clear mismatches occur, we will introduce appropriate human-specific cleavage sites and evaluate growth potential in Vero cells and HAE cultures. In SARS-CoV we will ablate several of these sites based on pseudotyped particle studies and evaluate the impact of select SARSr-CoV S changes on virus replication and pathogenesis."

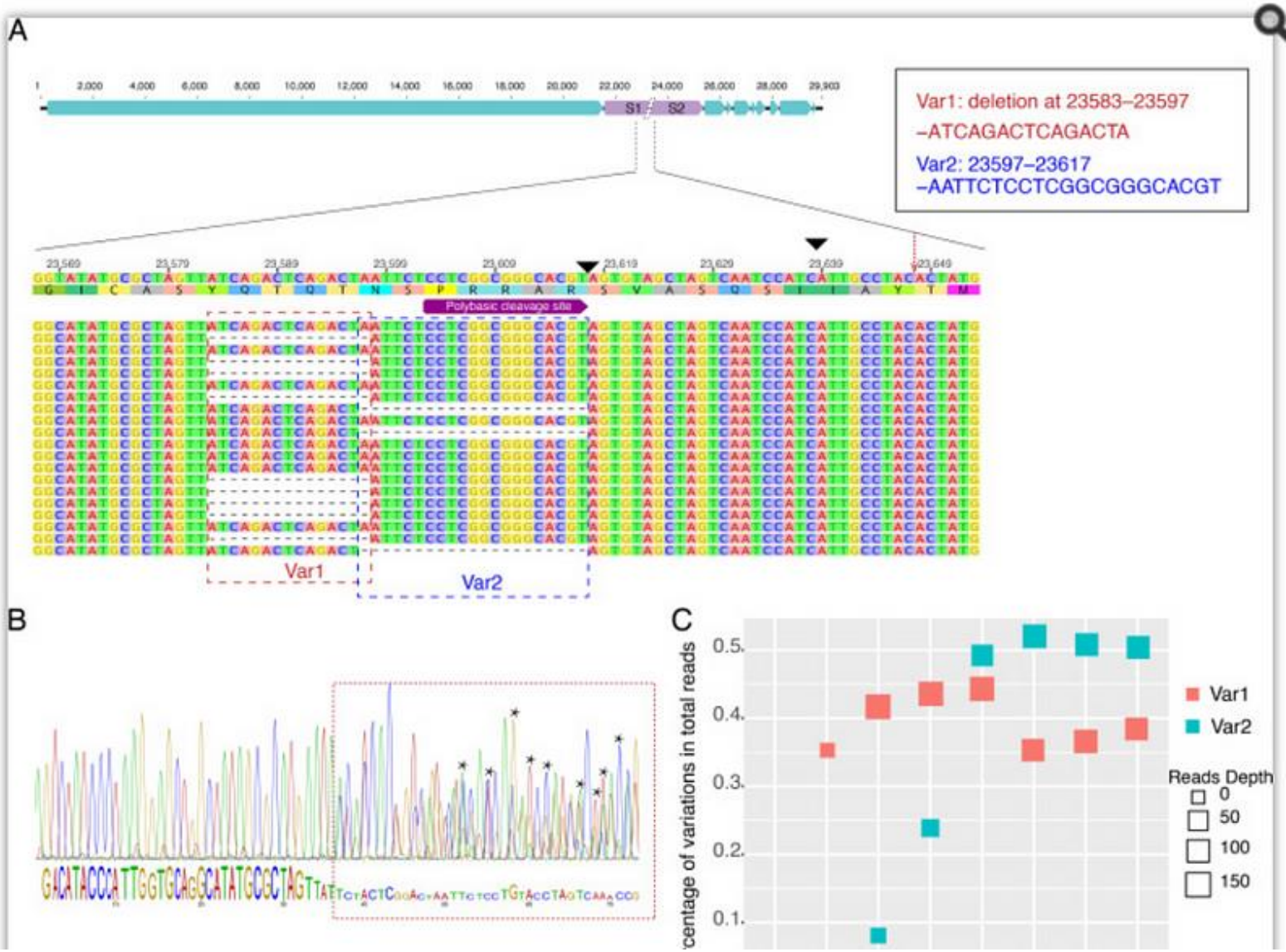
Vero cells are a kind of monkey kidney cells that are frequently used in labs.

If you culture SARS-CoV-2 in Vero cells, it rapidly loses the furin cleavage site.

One common mutation deletes QTQTN, another NSPRRAR.

You could not create SARS-CoV-2, with the same experimental setup.

FIG 1



What about other kinds of cell cultures?

2 other cell lines keep the furin cleavage site.

Calu-3 (airway) cells also induce subtle mutations. Another paper finds unique mutations in the E protein.

Another study found SARS-CoV-2 grows 3x more slowly in Calu-3 cells, making those an unlikely choice.

Vero/TMPRSS2 cells still cause mutations, but those cause the least.

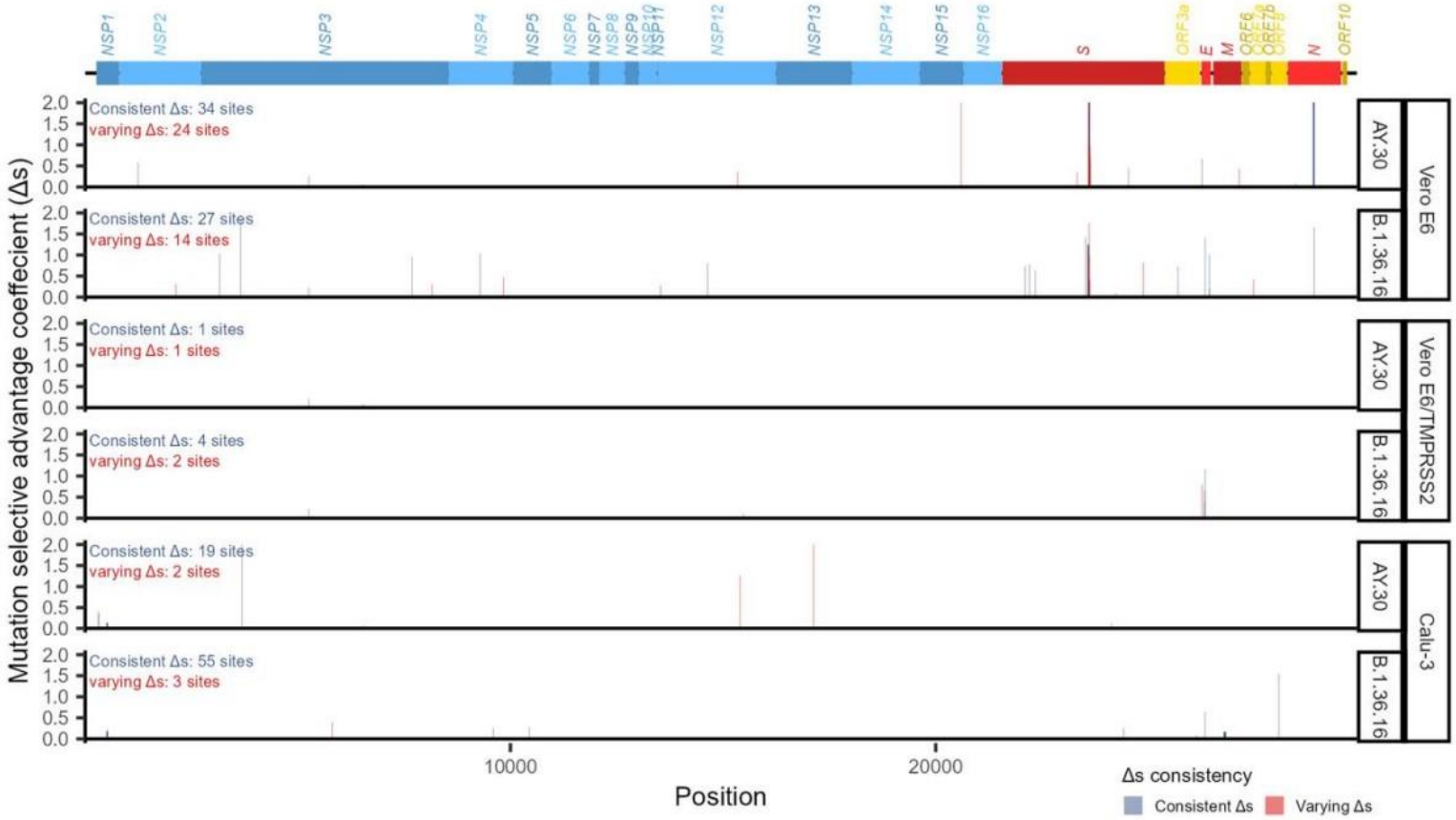


FIG. 4.—Sites detected as potentially harboring adaptation mutations. Site-wise mutation selective advantage coefficients, Δs, are shown, truncated at the maximum value of 2. The numbers of sites showing consistent Δs values (i.e., best described by the model *M1* in which Δs values are assumed to vary insignificantly among viral samples and experimental replicates, see Methods and supplementary notes, Supplementary Material online) are in blue, and those showing varying Δs values (i.e., best described by the model *M2* in which Δs values are allowed to vary significantly among viral samples and/or experimental replicates, see Methods and supplementary notes, Supplementary Material online) are in red. SARS-CoV-2 genome structure is shown at the top, with coding region and gene names (blue: coding regions of NSPs; red: structural protein coding genes; yellow: accessory genes).

Figure from Aiewsakun et al, 2023

What about HAE cells?

Zou et al 2021 cultured SARS2 in "human airway epithelia (HAE) cultured at an air-liquid interface (ALI)".

They used HAE cells from 11 donors. The FCS was preserved in 9 out of 11 of these cultures, although all of the cultures saw periodic FCS deletion mutations, among the sequenced samples:

> "While we found overall the viral transcriptome is similar to that generated from infected Vero cells, we identified a high percentage of mutated viral genome and transcripts in HAE-ALI. Two highly frequent deletions were found at the FCS region: a 12 amino acid deletion (^{678}TNSPRRAR↓SVAS^{689}) that contains the underlined FCS and a 5 amino acid deletion (^{675}QTQTN^{679}) that is two amino acids upstream of the FCS."

But they also found signs of adaptation to cell culture – frequent deletions across the M coding region:

> "In addition to these deletions in the S gene, we identified about 50 different in-frame or frameshift deletions in the M encoding region that appeared in all six samples of both MOI groups, and there were even more deletions in the M coding region that appeared in only a part of the six RNA samples (Data Set S1). Although the ratio of single deletion was low, the 50 deletion patterns that appeared in all six RNA samples had the ratios of 2.39% and 3.18% in MOI 0.2 and MOI 2 groups, respectively, which is similar or even higher than the identified canonical junction-spanning reads related to M sgRNAs (Fig. 3). Notably, most of these identified deletion patterns of the M gene also appeared in SARS-CoV-2-infected Vero cells"

The Wuhan Institute of Virology didn’t use HAE cells or at least never published any studies using HAE cells.

This is evident with only 2 (irrelevant) hits in PubMed using the search terms: (human airway epithelial cells OR ("HAE" AND cells)) AND ("wuhan institute of virology")

UNC did use HAE cells in some experiments (Menachery et al., 2015).

So, you can hypothesize that the WIV got HAE cells from some other institution and used these, just like you can hypothesize that WIV secretly did the UNC portion of the DEFUSE grant, the WIV possessed secret viruses, and the WIV created live viral chimeras with previously unknown backbones.

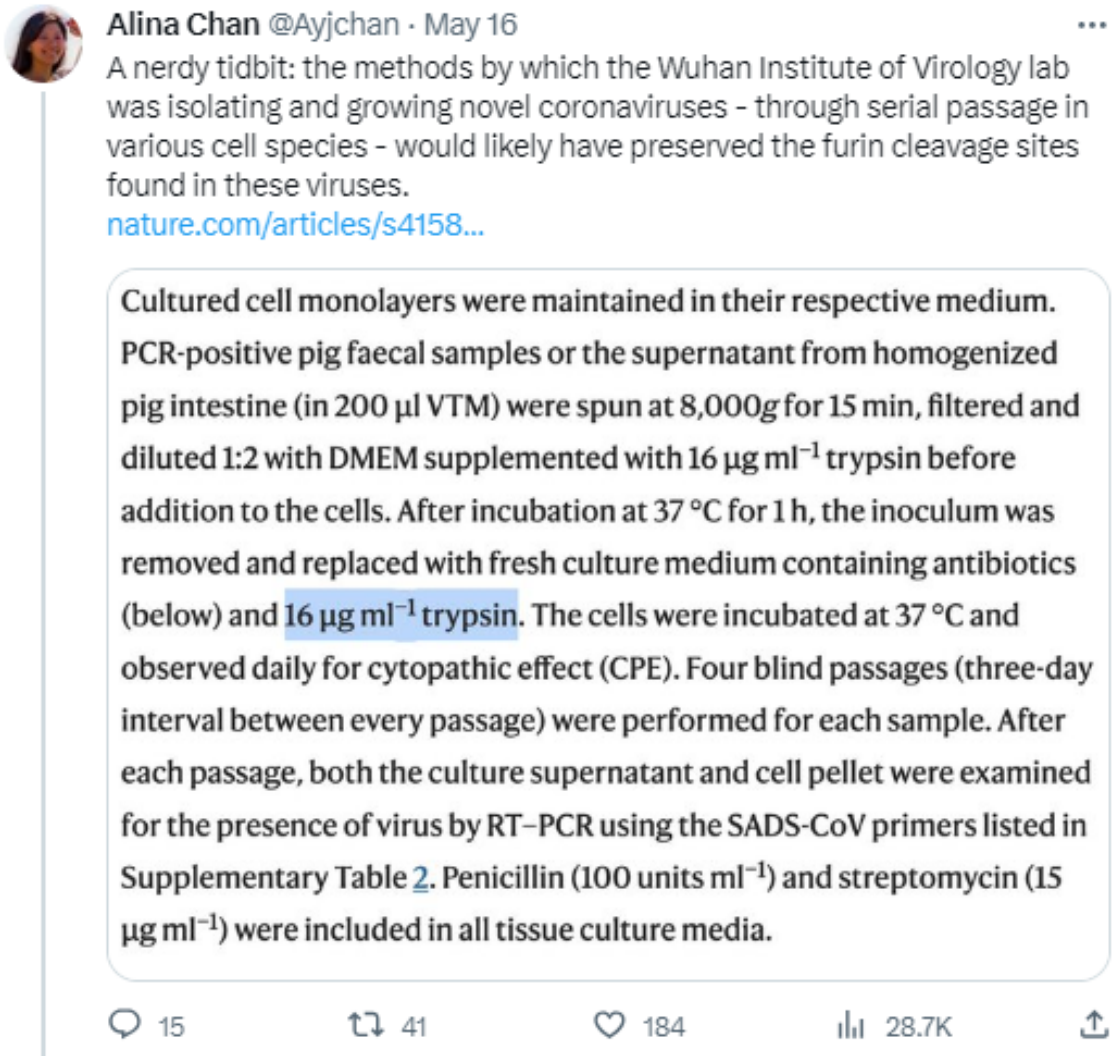
But you need to put some probability on each of those hypotheses, you can’t just treat them all as 100%.

What about trypsin?

Trypsin is an enzyme that can be added to cell cultures, it can have a similar cleavage effect to TMPRSS2, in the case of cells which do not express TMPRSS2.

We know that the WIV has used trypsin in some cell culture experiments.

Could you create SARS-CoV-2 in vero cells if trypsin was added?



Alina Chan @Ayjchan · May 16

A nerdy tidbit: the methods by which the Wuhan Institute of Virology lab was isolating and growing novel coronaviruses - through serial passage in various cell species - would likely have preserved the furin cleavage sites found in these viruses.

nature.com/articles/s4158...

Cultured cell monolayers were maintained in their respective medium. PCR-positive pig faecal samples or the supernatant from homogenized pig intestine (in 200 µl VTM) were spun at 8,000*g* for 15 min, filtered and diluted 1:2 with DMEM supplemented with 16 µg ml^{-1} trypsin before addition to the cells. After incubation at 37 °C for 1 h, the inoculum was removed and replaced with fresh culture medium containing antibiotics (below) and 16 µg ml^{-1} trypsin. The cells were incubated at 37 °C and observed daily for cytopathic effect (CPE). Four blind passages (three-day interval between every passage) were performed for each sample. After each passage, both the culture supernatant and cell pellet were examined for the presence of virus by RT–PCR using the SADS-CoV primers listed in Supplementary Table 2. Penicillin (100 units ml^{-1}) and streptomycin (15 µg ml^{-1}) were included in all tissue culture media.

15 41 184 28.7K

A 2021 experiment tried this and still observed a mutation that deleted the entire cleavage site:

adaptation (Figure 6D). As TMPRSS2 expression prevented MBCS mutations, we tested whether the addition of trypsin (0.7 µg/ml TPCK-Trypsin) would have a similar effect. Surprisingly, the addition of trypsin to VeroE6 cells, but not VeroE6-TMPRSS2 cells, led to deletion of the entire MBCS (Figure 6—figure supplement 1A,C). This deletion may arise due to the complete cleavage (S1/S2 and S2') of virus particles that are not bound to the cellular membranes, which would inactivate them. Cell surface expressed TMPRSS2 could accelerate TMPRSS2-mediated entry and cell-cell spread, reducing the chance of trypsin cleavage in the supernatant. Additionally, we tested whether the addition of fetal bovine serum (FBS, heat-inactivated, 10% final concentration) affected culture adaptation as this is commonly added when producing viral stocks. FBS had a similar effect to trypsin in the VeroE6, but not the VeroE6-TMPRSS2 culture, indicating that proteases capable of cleaving spike may be present in serum and that FBS should be avoided when propagating SARS-CoV-2 (Figure 6—figure supplement 1B,D).

← MBCS = furin cleavage site

Another group tried and got the same result.

But those groups were using 0.5 μg/mL and 0.7 μg/mL of trypsin.

The Wuhan lab used 16 μg/mL when they first cultured SARS-CoV-2 in 2020.

Maybe there's a dosage dependent effect?

Another group (Kim et al, 2022) tried using up to 10 μg/mL of trypsin and the furin cleavage site was, indeed, preserved.

However, there were still lots of mutations after it had been passaged for a while. Trypsin increased the mutation rate:

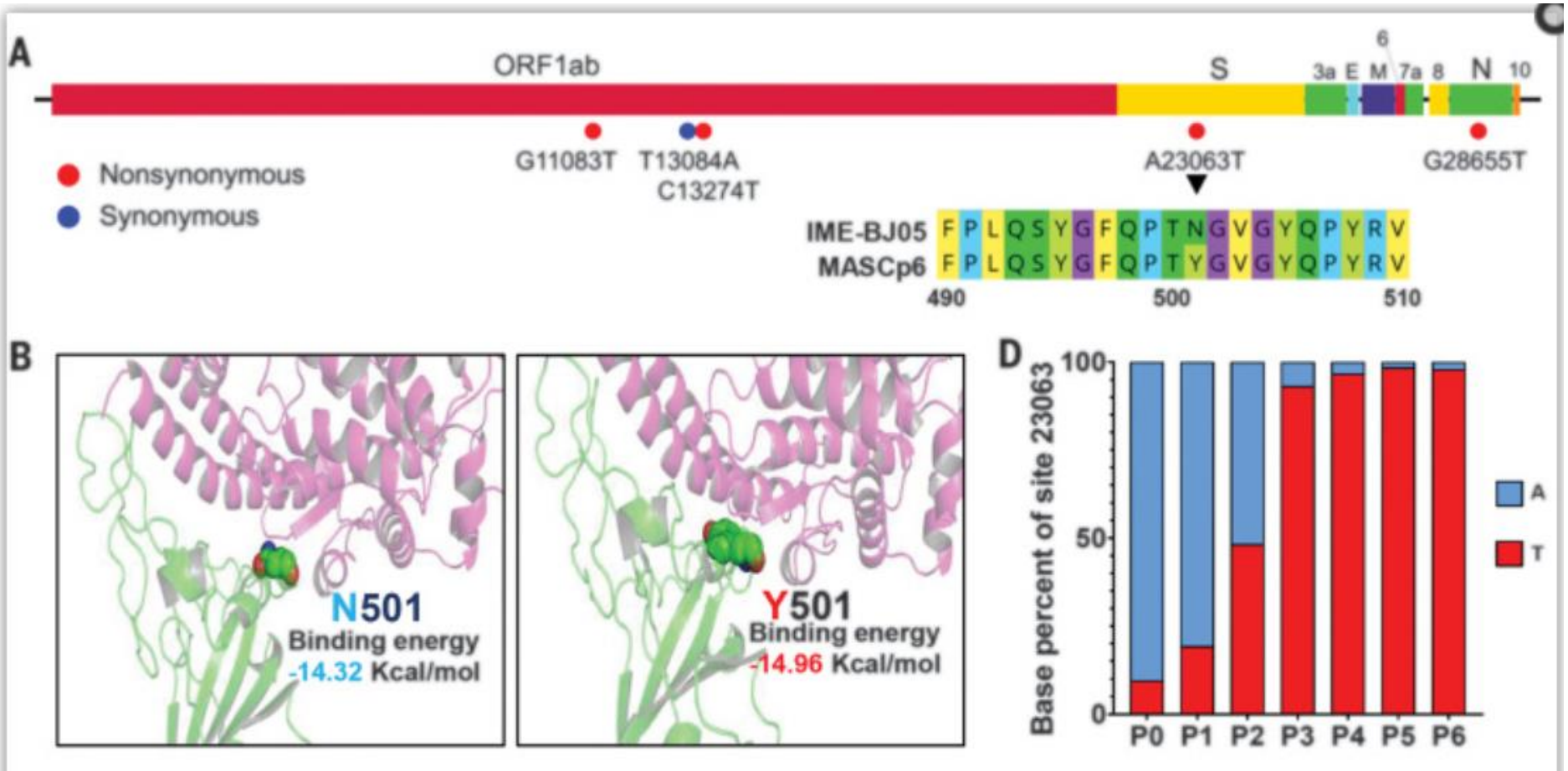
> Interestingly, the number and location of the amino acid (aa) changes differed between the cell-culture-passaged strains in the presence or absence of trypsin (Fig. 10). The 50th-passage strain without trypsin addition contained nine aa mutations, including three aa deletions (DELs), whereas the P50(+) virus had 23 aa variations, including 13 aa DELs. The nine aa

Maybe there's a small window of opportunity for: just the right cells, just the right amount of trypsin, and the virus wasn't in culture for very long before it leaked.

But it's not likely. There would also be a whole lot of publication worthy results along the way, if you did learn all that stuff – scientists around the world have needed several years of experimentation to figure out all of this stuff.

What about using transgenic mice, instead of cell cultures?

Early strains of covid could not infect wild mice.
Scientists have since used transgenic mice with human ACE2 receptors to do covid experiments. When they did this, it caused a mutation in the spike protein, called N501Y.
Because covid didn't already have that mutation, it probably wasn't created in those mice.
That mutation is also beneficial in humans, so it wouldn't revert to N501 on jumping from mice to humans.



What about ferrets?

One theory suggests that SARS-CoV-2 was made through serial passaging in ferrets.

We've passaged the virus through ferrets in lab experiments, it quickly gains another mutation, called N501T, and sometimes one called Y453F.

When the virus spilled over into mink on farms, it frequently gained the mutations N501T, Y453F, and F486L, and L452M.

These would likely mutate back after spill-over into humans, but that wouldn't happen immediately.

This makes mink and ferrets an unlikely intermediate host, whether for a natural or lab origin.

On the other hand, when an experiment infected raccoon dogs with covid, the virus did not mutate.

Raccoon dogs are a likely intermediate host.

What about the DEFUSE grant?

A whistleblower uncovered research planning to put cleavage sites in bat viruses. Ecohealth alliance proposed something called the DEFUSE project, to study viruses potentially emerging from bats.

The grant proposal is from 2018. It outlines the risk that some SARS-like coronavirus will eventually spill over into people, and it proposes various things to stop that from happening.

They plan to come up with drugs that boost bats' immune systems and then spray these drugs inside bat caves.

They also propose genetic engineering, to add furin cleavage sites to viruses.

On page 35 of the DEFUSE grant proposal, they write:

"We will analyze all SARSr-CoV S gene sequences for appropriately proteolytic cleavage sites in S2 and for the presence of potential furin cleavage sites. SARSr-CoV S with mismatches in proteolytic cleavage sites can be activated by exogenous trypsin or cathepsin L. Where clear mismatches occur, we will introduce appropriate human-specific cleavage sites and evaluate growth potential in Vero cells and HAE cultures. In SARS-CoV we will ablate several of these sites based on pseudotyped particle studies and evaluate the impact of select SARSr-CoV S changes on virus replication and pathogenesis."

Is this the smoking gun, showing that scientists made SARS-CoV-2?

No. First, the grant was rejected.

Second, the work adding furin cleavage sites was supposed to be done at the University of North Carolina.

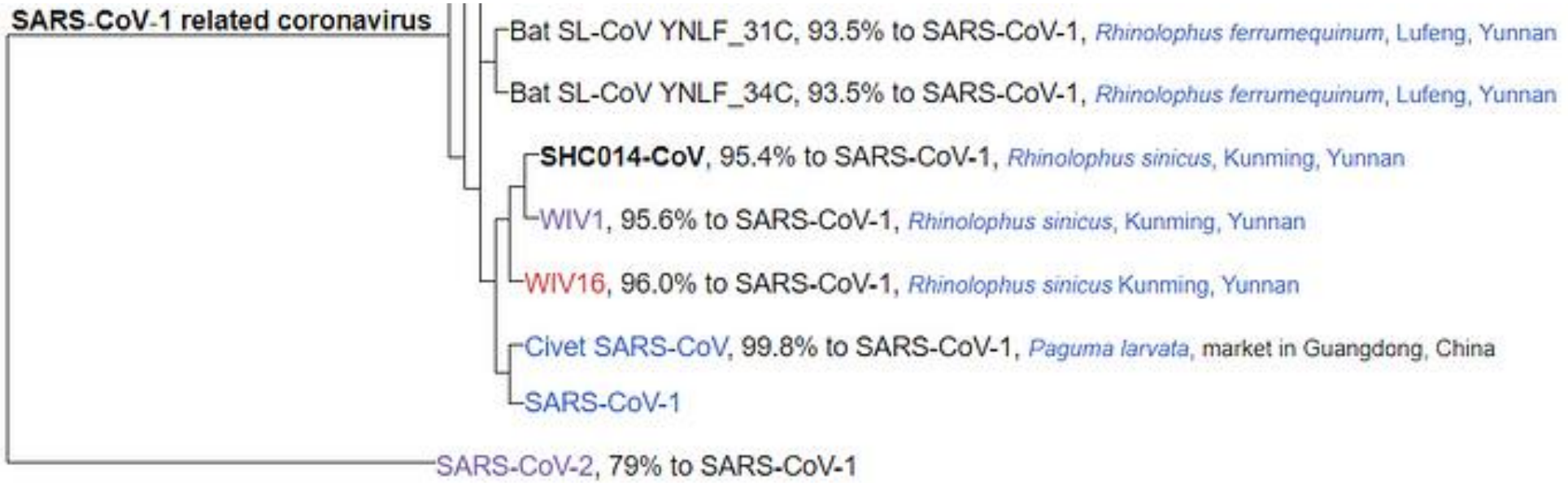
Third, they talk about growing the viruses in Vero cells. As we've seen, vero cells lose the furin cleavage site.

Fourth, it looks like they may be talking about the S2' site, not the S1/S2 site:

> In parallel, we will evaluate whether RBD deletion repair restores the ability of low risk strains to use human ACE2 and grow in human cells. *S2 Proteolytic Cleavage and Glycosylation Sites:* After receptor binding, a variety of cell surface or endosomal proteases[68-71] cleave the SARS-CoV S glycoprotein causing massive changes in S structure [72] and activating fusion-mediated entry[64,73]. We will analyze all SARSr-CoV S gene sequences for appropriately conserved proteolytic cleavage sites in S2 and for the presence of potential furin cleavage sites[74,75]. SARSr-

Fifth, they talk about doing all the work in two known bat virus backbones (WIV1, SHC014) that are closely related to SARS, but only distantly related to covid:

Technical Approach: Our goal is to defuse the potential for spillover of novel bat-origin high-zoonotic risk SARS-related coronaviruses in Asia. **In TA1** we will intensively sample bats at our field sites where we have identified high spillover risk SARSr-CoVs. We will sequence their spike proteins, reverse engineer them to conduct binding assays, and insert them into bat SARSr-CoV (WIV1, SHC014) backbones (these use bat-SARSr-CoV backbones, not SARS-CoV, and are exempt from dual-use and gain of function concerns) to infect humanized mice and assess capacity to cause SARS-like disease. Our modeling team will use these data to build **machine-learning genotype-phenotype models** of viral evolution and spillover risk. We will uniquely



None of these proposed lab leak theories match how researchers usually work

The reason researchers use known backbones like WIV16 isn't just because they're worried about SARS, it's because you want to learn something from an experiment by making one change at a time.

You start with a known backbone, swap in a spike, and see what happens.

Or you start with a known spike, add a cleavage site, and see what happens.

Usually this is done with a pseudovirus before you'd even think about trying it with a live virus, in vitro or in vivo.

The lab leak theories suggest that the Wuhan lab found a novel virus, recognized it was important somehow, made a reverse genetics system for it, inserted a suboptimal furin cleavage site that had never been used before, made that insertion out of frame, tried it as a live virus in some kind of animal or culture.

This was all done by Shi Zhengli's small group, in secret, without publishing any intermediate steps. They kept publishing other research, the whole time.

Then the virus leaked, it migrated across town to find the closest raccoon dog, and started spreading.

After discovering the leak, they kept on working and publishing, including publishing a highly similar virus that made many people question if they'd added the cleavage site.

Yuri agrees that the DEFUSE grant could not create SARS-CoV-2, after Stuart Neil walked through a similar explanation.

In summary:

There’s no evidence the Wuhan lab had a virus that could be turned into SARS-CoV-2.

All known research was working with SARS1 family viruses.

SARS-CoV-2 is only 80% similar to SARS-CoV-1, it’s unlikely anyone would have thought a precursor virus was important, or tried to manipulate it.

The furin cleavage site is suboptimal and not one used for research. It looks natural, not engineered.

It’s not clear how you would successfully culture SARS-CoV-2, most approaches would fail.

Working with transgenic mice or ferrets would cause mutations that SARS-CoV-2 doesn’t have.

There’s no need to optimize the virus for human ACE2 or to make a chimera, because some bat viruses already bind well to human ACE2.

The genetic evidence points towards a natural virus.

Any Bayesian analysis would have to account for the low odds of these many unlikely engineering choices, as well as the low odds that the lab had a relevant virus in the first place.

Probabilities:

Odds DEFUSE grant secretly happened at the WIV **(40% – this is Rootclaim's number, I think it's lower, but I'm steelmanning)**
- they had a suitable secret virus * **(1 in 1,000 – based on Latinne FOIA, 2018 paper, sampling rates. This could be lower)**
- they recognized the spike was interesting * **(1 in 10? It's not much like SARS, but maybe they could measure ACE2 binding)**
- they made a reverse genetics system for it, instead of using an existing backbone * **(1 in 100 – no good reason)**
- they inserted a furin cleavage site * **(1 in 1 – probably lower, but I'm steelmanning here, I'll just give lab leak this one)**
- they put the site at S1/S2, not S2' * **(1 in 2 – maybe not a huge deal)**
- they chose RRAR * **(1 in 10 – A is weird, but not highly detrimental. K works much better)**
- they chose PRRAR * **(1 in 20 – This one is really weird and hard to explain)**
- they inserted it out of frame * **(1 in 6 – let's assume that's in the secret virus, 6 different codons for serine)**
- they did the experiments with live virus, not pseudovirus **(1 in 2? Unclear what DEFUSE intended, probably lower)**
- they found some effective way to culture it * **(1 in 10? – most cultures/animals fail to make SARS2, assume they're lucky)**
- they never published any of the work leading up to this * **(1 in 10? Debatable, very hard to say the exact number here)**
- what they created leaked * **(1 in 50 – normally 1 in 500, but adjust generously upwards to steelman – BSL-2, live virus, etc)**
- the leak started an outbreak * **(1 in 3)**
- it only showed up at the market * **(1 in 10,000 – use ratio of Wuhan vendors to Wuhan population, or use traffic analysis)**
- it showed up at the market twice * **(1 in 2,000 – it could look like 2 lineages by chance, but that's very unlikely)**
- this all happened in the same month the SARS outbreak started * **(1 in 6? or 1 in 4, or ignore seasonality, not a big deal)**
- the most positive samples happened to be in a shop selling susceptible animals * **(1 in 68)**
- that shop was one of the only three (in town) previously fined for selling illegal wildlife * **(3 in 10)**
- the cover-up was so good that neither DRASTIC nor the US government has solved this **(1 in 10 – could be lower or higher)**

Total odds against lab leak: 1 in $5*10^{25}$

Supplemental information

Chronology of all sampling trips done by the lab

1959: The Wuhan Institute of Microbiology was founded

1972: The institute was renamed to Wuhan Institute of Virology (WIV)

2004: First field trip to collect samples (n = 328) from bats in China (i.e., Nanning, Guangxi, Maoming, Guangdong, and Tianjin) between March to December 2004.
- SARSr-CoV identified: Rf1 from *R. ferrumequinum*, Rm1 from *R. macrotis*, and Rp1-3 from *R. pearsoni.*
- Authors: Li et al. (2005)

2004-2005: First appearance of Shi Zhengli and Yan Zhu as co-authors of studies (unrelated to coronaviruses) published from the WIV (Zeng et al., 2004; Huang et al., 2005).

2004-2014: Field trip to collect samples (n = 2,061) from bats from 19 provinces in China between November 2004-2014.
- SARSr-CoV identified: Mi-BatCoV 1, Mi-BatCoV HKU8, BtRf-AlphaCoV/HuB2013, SARSr-CoV, HKU2-related CoV, and novel BtCoV/Rh/YN2012.
- Authors: Wang et al. (2019)

2006: Field trip to collect samples (n = 24) from bats in China (i.e., Hubei, Guangxi, Hong Kong, Guizhou, and Yunnan) in September 2006.
- SARSr-CoV identified: Rp3/Rs672 and HKU3/Rs806 from *R. sinicus.*
- Authors: Yuan et al. (2009)

2006-2016: Screened HKU10 against archived bat fecal samples (n = 8,004) from 25 provinces of China and one province of Laos collected from September 2006 to June 2016.
- SARSr-CoV identified: n/a. Results demonstrated diverse gene pool of HKU10 in bats in Yunnan.
- Authors: Wang et al. (2021)

2009-2016: Field trip to collect samples (n = 555) from bats in 4 locations in Yunnan (Chuxiong, Mojiang, Jinghong, and Mengla), China from 2009-2016.
- SARSr-CoV identified: HKU9-2202 from *R. leschenaultia*.
- Authors: Luo et al. (2018)

2010-2015: Field trip to collect samples (n = n/a) from bats in in numerous Chinese provinces (Anhui, Beijing, Guangdong, Guangxi, Guizhou, Hainan, Henan, Hubei, Hunan, Jiangxi, Macau, Shanxi, Sichuan, Yunnan, and Zhejiang) from 2010 to 2015.
- SARSr-CoV identified: n/a.
- Phylogeny tree was constructed with 202 RdRp sequences from sarbecoviruses and results showed that host switching of coronaviruses occur frequently among bats.
- Authors: Latinne et al. (2020)

2011-2012: Field trip to collect samples (n = 117) from bats in Kunming, Yunnan province, China from April 2011 to September 2012.
- SARSr-CoV identified: RsSHC014 and Rs3367/WIV1 from *R. sinicus.*
- Authors: Ge et al. (2013)

2011-2014: Field trip to collect samples (n = 431) from bats in Yunnan province, China from 2011-2014.
- SARSr-CoV identified: n/a.
- Results: 57 samples were positive for SARSr-CoV, with higher virus levels noted from late summer vs. autumn months.
- Authors: Wang et al. (2016)

2011-2015: Field trip to collect samples (n = 602) from a single habitat in Kunming, Yunnan, China from April 2011 to October 2015.
- SARSr-CoV identifed (11 new strains): 11 new SARSr-CoV strains: Rs4081, Rs4084, Rs4231, Rs4237, Rs4247, Rs4255, Rs4874, Rs7327, Rs9401, Rf4092 and As6526.
- Recombination analyses showed that all building blocks of SARS-CoV are present in bat SARSr-CoVs from this single location in Yunnan.
- Authors: Hu et al. (2017)

2012-2013: Field trip to collect samples (n = 276) from bats in Mojiang County, Yunnan, China.
- Alpha-CoV identified: HKU2, HKU8 and BtCoV1, and novel species HKU7 from *M. schreibersii* and HKU10 from *H. pomona*.
- Beta-CoV identified: novel RaBtCoV/4991 from *R. affinis* and novel HpBtCoV/3740-2 from *H. Pomona.*
- Authors: Ge et al. (2016)

2012-2015: Field trip to collect samples (n = 1,059) in China (Guangdong, GuangXi, and Sichuan) from 2012 to 2015.
- MERS-CoV-related bat coronavirus identified: BtCoV/li/GD/2013-845 and BtCoV/li/GD/2014-422 from *I. io*.
- Authors: Luo et al. (2018)

2012-2019: 18 bat fecal samples were selected from the WIV biobank, collected during longitudinal survey from 2012 to 2019
- SARSr-CoV identified (14 viruses): RsYN2012, RsYN2016A, RsHuB2019A/B, RsYN2016B, RsYN2016C, RsYN2013, RsGZ2015, RsYN2016D, RsGD2014A, RsGD2014B, RsYN2014, RsYN2018, RstYN2015, and RaTG15.
- Authors: Guo et al. (2023)

2013: WIV16 was isolated from one bat fecal sample that was collected in July 2013 in Kunming, Yunnan, China.
- Authors: Yang et al. (2016)

2015: Field trip to collect samples from bats in Mojiang County, Yunnan, China in May 2015.
- SARSr-CoV identified (14 viruses): RaTG15 (from *R. affinis*), Rst7924, Rst7921, Rst7907, Rst7896, Rst7931, Rst7905, and Rst7952 (from *R. stheno*).
- Authors: Guo et al. (2021)

2013-2016: Swine acute diarrhoea syndrome coronavirus (SADS-CoV) was screened against bat samples (n = 591) collected from 7 Guangdong locations from 2013-2016.
- Results: SADS-CoV is a HKU2-related coronavirus that is 98.48% identical to a bat (Rhinolophus) coronavirus detected in 2016 in a bat cave close to the index pig farm.
- Authors: Zhou et al. (2018)

2015: Field trip to collect samples from bats in Mojiang County, Yunnan, China in May 2015.
- SARSr-CoV identified (14 viruses): RaTG15 (from *R. affinis*), Rst7924, Rst7921, Rst7907, Rst7896, Rst7931, Rst7905, and Rst7952 (from *R. stheno*).
- Authors: Guo et al. (2021)

2019: Field trip to collect samples (n = 133) from cave nectar bats from Daoba and Tianshengqiao caves in Mengla County, Yunnan, China from January to December 2019.
- SARSr-CoV identified: n/a. Results found 13% of samples were positive for GCCDC1-CoV, and none were positive for HKU9-CoV and Mengla virus (MLAV).
- Authors: Zhao et al. (2022)

Chronology of all gain of function experiments done by the lab

List of experiments done in WIV from 2005-2019.

- 2005: Published 2 papers on CoV: RT-PCR test for SARS (Hu et al., 2005); detection of SARS-related CoV in horseshoe bats (Li et al., 2005).
- 2006: Published 1 paper on SARS-related CoV from horseshoe bats (Ren et al., 2006).
- 2007: Published 3 papers on CoV – 1st gain-of-function research swapping the S protein of SARS-related bat CoV for that of SARS-CoV-1 enabled it to bind to human and civet ACE2 receptors (Ren et al., 2007); DNA vaccine development (Hu et al., 2007; Bai et al., 2007).
- 2008: Published 4 papers on CoV – review (Shi and Hu 2008); construction of non-infectious SARS-CoV (Wang et al., 2008); vaccines (Gai et al., 2008; Bai et al., 2008).
- 2009: Published 8 papers on CoV – ACE2 binding efficiency of SARS-CoV vs. SARS-related CoVs (Xu et al., 2009); S protein immunogenicity (Zhou et al., 2009); vaccine development (Lu et al., 2008; Hu et al., 2009a; Hu et al., 2009b; Lu et al., 2009); SARS-related CoV identification in bats (Yuan et al., 2009).
- 2010: Published 2 papers on CoV – 2nd gain-of-function research using mutagenesis to identify key residues in bat ACE2 that enhance binding efficiency by SARS-CoV-1 (Hou et al., 2010a); DNA vaccine study (Hou et al., 2010a).
- 2011: Published 3 papers on CoV – inactivated vaccine study (Gai et al., 2011); function of ORF3b (Zhou et al., 2011); antiviral development (Li et al., 2011).
- 2013: Published 3 papers on CoV –S protein immunogenicity (Zhou et al., 2013); a review (Wang and Hu, 2013); potential progenitor of SARS-CoV (Ge et al., 2013).
- 2015: Published 4 papers on CoV – 3rd gain-of-function research showing that HKU4 can be activated by protease by inducing two mutations in its S protein (Yang et al., 2015); 4th gain-of-function research showing that chimeric virus expressing S protein of bat SHC014 in SARS1 backbone can use human ACE2 receptor to infect human airway cells and cause disease in mice (Menachery et al., 2015); function of 2'-O-MTase (Wang et al., 2015); a review (Hu et al., 2015).
- 2016: Published 4 papers on CoV – ORFX function in WIV1 and WIV16 (Zeng et al., 2016); surveillance of bat SARS-related CoV (Wang et al., 2016; Ge et al., 2016); a study identifying novel SARS-related CoV that is genomically closest to SARS-CoV in Yunnan Province in July 2013 (Yang et al., 2016).
- 2017: Published 4 papers on CoV: 5th gain-of-function research showing that chimeric viruses expressing Rs4231 and Rs7327 S proteins in WIV backbone can infect human ACE2-expressing cells (Hu et al., 2017). antiviral study (Sun et al., 2017); antibody study (Zeng et al., 2017); CoV screening in children (Liu et al., 2017); CoV surveillance in rats in the Yunnan (Ge et al., 2017).
- 2018: Published 6 papers on CoV – MERS-CoV surveillance (Omneh et al., 2018; Zohaib et al., 2018); bat CoV surveillance in Yunnan (Wang et al., 2018; Luo et al., 2018); Bat CoV related to MERS-CoV (Luo et al., 2018); investigation on origin of fatal SADS outbreak in pig farm in Guangdong (Zhou et al., 2018).
- 2019: Published 8 papers on CoV – RT-PCR test of MERS-CoV (Zhou et al., 2019); reviews (Fan et al., 2019; Yu et al., 2019; Cui et al., 2019); antiviral study (Xia et al., 2019); novel bat CoVs (Lim et al., 2019; Wang et al., 2019) surveillance on bat CoV spillover among rural residents (Li et al., 2019).

2005-2019 total papers published: 52, only 5 may qualify as gain-of-function, for 2 of those the GoF steps were done at other labs