

Covid origins debate, day 3: Probability

$A = \pi r^2$
 $C = 2\pi r$

$V = \frac{1}{3} \pi r^2 h$

$V = \pi r^2 h$

	30°	45°	60°
sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$

$\int \sin x dx = -\cos x + C$
 $\int \frac{dx}{\cos^2 x} = \tan x + C$
 $\int \tan x dx = -\ln|\cos x| + C$
 $\int \frac{dx}{\sin x} = \ln\left|\frac{x}{2}\right| + C$
 $\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \arctg \frac{x}{a} + C$
 $\int \frac{dx}{x^2 - a^2} = \frac{1}{2a} \ln\left|\frac{x-a}{x+a}\right| + C$

$\tan(\theta)$
 10
 5
 -5
 θ/rad

$2x$
 60°
 30°
 x
 $x\sqrt{3}$
 45°
 $x\sqrt{2}$

$ax^2 + bx + c = 0$
 $a(x^2 + \frac{b}{a}x + \frac{c}{a}) = 0$
 $x^2 + 2\frac{b}{2a}x + (\frac{b}{2a})^2 - (\frac{b}{2a})^2 + \frac{c}{a} = 0$
 $(x + \frac{b}{2a})^2 - \frac{b^2 - 4ac}{4a^2} = 0$

This has been a long and complicated debate.

I think we've already agreed on a few things:

“3 sick researchers at the Wuhan institute of virology” claims can't be trusted.

It's normal for pandemics to start in big cities.

A September lab leak is impossible.

→ as a corollary, the Wuhan database going offline on September 12th must be unimportant.

Some things don't seem worth contesting:

We disagreed on whether or not seasonality is important. I think that leans towards a natural origin, but it's not strong evidence either way.

We disagreed on the exact prior odds of a pandemic in Wuhan. We've agreed it's at least 1.5%.

I think it's higher than 1.5% for a few reasons. The most obvious one is the difference between southern and northern China. I can dig up more details about bats, farming, markets, and cuisine, etc, to get the 1.5% up to 3%, if this debate comes down to a factor of 2, but there are some much bigger disagreements we should work out before that.

Where the 2003 SARS pandemic first spilled over is academically interesting, but doesn't matter too much for this debate. Even if I could prove that started in Hubei, it wouldn't necessarily change the odds of SARS2 starting in Hubei.

Some disagreements are very important:

I think the odds that a lab leak would show up across town at the Huanan market are about 1 in 10,000.

I think that the lineage A and lineage B evidence points strongly towards a market origin.

I think it's important that the evidence within the market suggests infected animals in 1 shop (or 2).

I think the lab had 0 secret viruses. Even if the DEFUSE grant secretly happened, I think they would collect ~30 more and be unlikely to find a SARS-CoV-2 precursor.

I think the odds they would create a virus like SARS-CoV-2 are low, even if they did DEFUSE. For instance, because it's a novel backbone, only 80% similar to SARS, and PRRAR out of frame is a very strange furin cleavage site.

Rootclaim thinks the fact that the first cases were found at that market carries no weight, and that location is equally likely under a lab leak or zoonosis.

Rootclaim thinks that evidence points towards the lab.

Rootclaim thinks that evidence carries no weight.

Rootclaim thinks the lab had 180+ secret viruses, one of which would be easy to turn into SARS-CoV-2.

Rootclaim thinks that the existence of the DEFUSE grant gives a 20% likelihood that the WIV created SARS-CoV-2.

There are some things that I think I've proved, but I'm not sure Saar agrees:

The first cluster of cases is at the Huanan Seafood Market.

It's normal for cases from a market outbreak to center on a market (it also happened with Xinfadi).

The data has no serious bias issues that question that it's centered on Huanan market.

The only lab leak scenario that fits the data is someone going straight from the lab to the market, starting a cluster, and the virus not spreading at the lab or anywhere else (that's possible, but very unlikely).

Mostly, I think this conversation has been hard for a reason: the thing I'm arguing against keeps changing.

Objections to the Market Evidence:

- ~~• The first case wasn't at the market~~
- ~~• The first two cases were visitors to the market~~
- ~~• There are actually lots of early cases, besides the market~~
- ~~• The data is biased~~
- ~~• Cases won't be centered on the market, anyways~~
- ~~• It's actually just the mahjong room~~
- The market is actually a super likely place for a cluster to start
- It's the ventilation
- Sure it's centered on the market, but that's because it's the Wuhan CDC
- Even if the raccoon dogs were sick, it's still a lab leak

← I think we made it to here?

An illustration might help:



These were some claims Rootclaim made in week 1, I think I rebutted most of these:

Claim:

- The first case wasn't at the market
- The first two cases were visitors to the market
- There are actually lots of early cases, besides the market
- The data is biased
- Cases won't be centered on the market, anyways
- It's actually just the mahjong room
- The market is actually just a super likely place for a cluster to start
- It's the ventilation
- Sure it's centered on the market, but that's because it's the Wuhan CDC
- Even if the raccoon dogs were sick, it's still a lab leak

Rebuttal:

Dec 8th case was actually Dec 16th

4 of 5 first cases worked there, 5th a repeat visitor

multiple lines of reasoning reject this

unbiased before Dec 29th and after Jan 18th,
and no other claims of bias have been proven.

Xinfadi market proves otherwise

First case didn't play mahjong, and there are
lots of mahjong rooms in Wuhan.

← (I think maybe we made it to here?)

To reiterate briefly, here's why I think the first cluster was at the Huanan market:

The first 5 known cases were there. 4 were employees, the 5th was a repeat visitor.

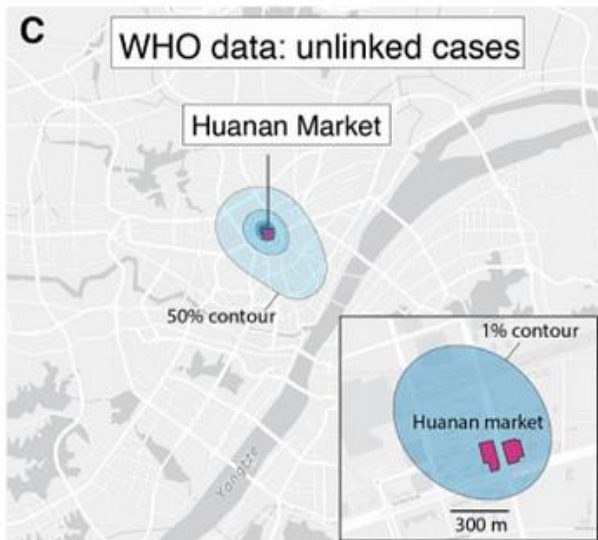
Years of searching have failed to find any earlier cases.

Early hospitalizations were 50%+ market linked at multiple hospitals across town, before the market link was known.

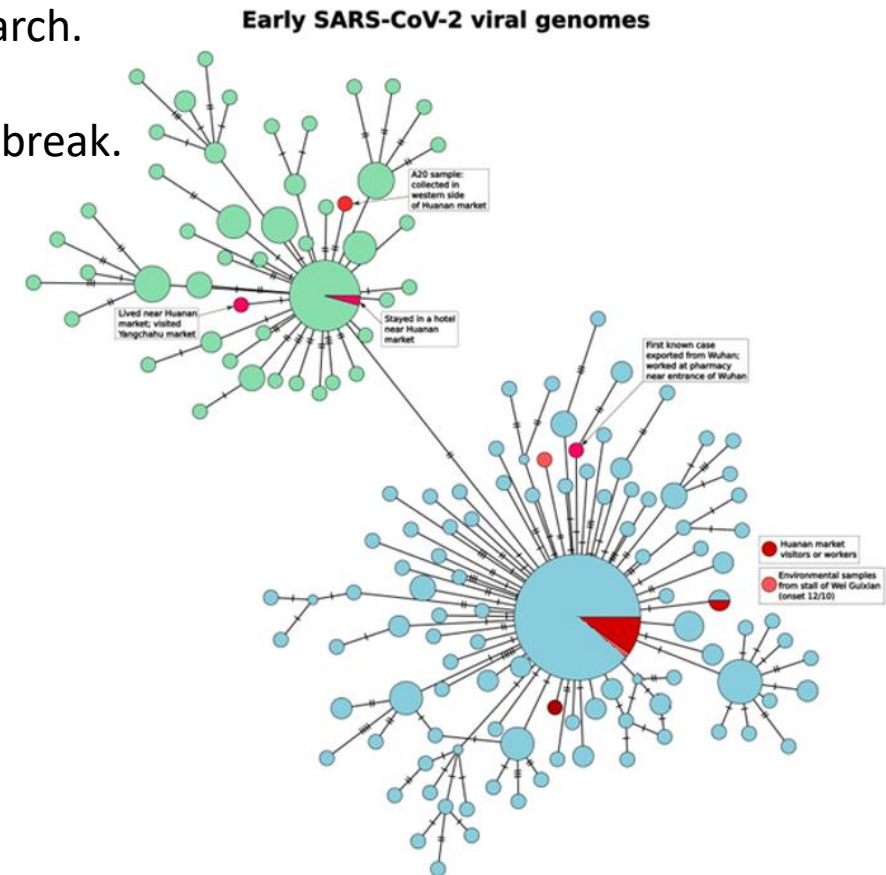
December hospitalizations were still 33% market linked after a retrospective search.

The genetic clock and simple exponential math line up well with the market outbreak.

Even the unlinked cases were centered on the market:



The market is also the center of the pandemic, genetically:



Here's why I said 1 in 10,000:

4 out of the 5 first known cases worked at the market.

~1,000 people work at the market, out of 10 million in Wuhan = 1 in 10,000.

Analysis of traffic in Wuhan said there are ~1,600 places in Wuhan that get more traffic than the market.

Weighted by traffic to each, it gets 1 in 10,000 of the traffic (measured by Sina visitor system check-ins).

Average: 703.0095465	Count: 1676	Sum: 1178244
----------------------	-------------	--------------

$120/1,178,244 =$
(1 in 9,819)

The [full list of 1,600 places can be found here](#).

I can't include this in my slides – it would be 45 pages of places more popular than the Huanan market.

We can filter that list, remove outdoor locations, add more criteria, assume some bias in the dataset, but Huanan is still not going to be anywhere near the top of the list.

Even among supermarkets and shopping malls, Huanan market doesn't stand out.

There are 70 markets more popular than Huanan, in this dataset.

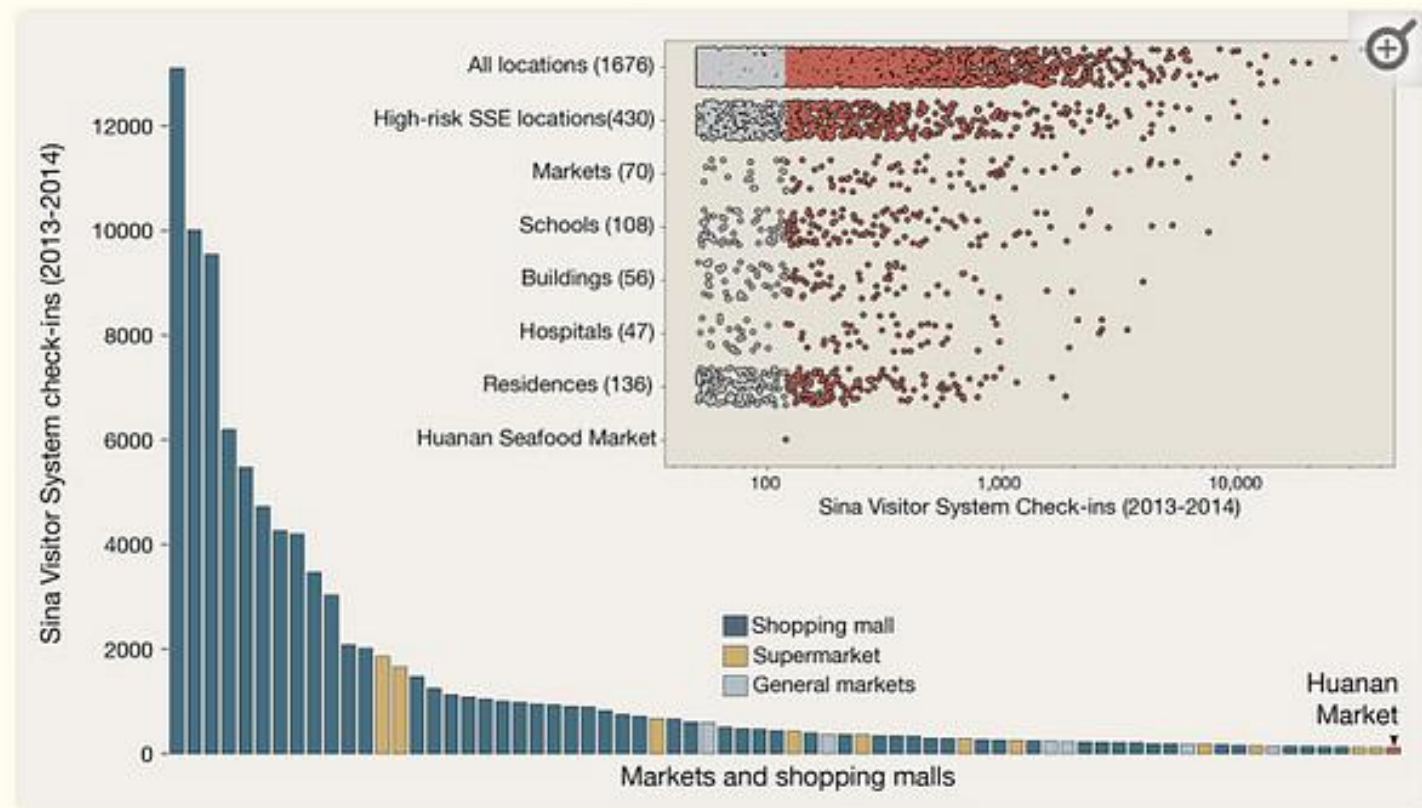


Fig. 3.

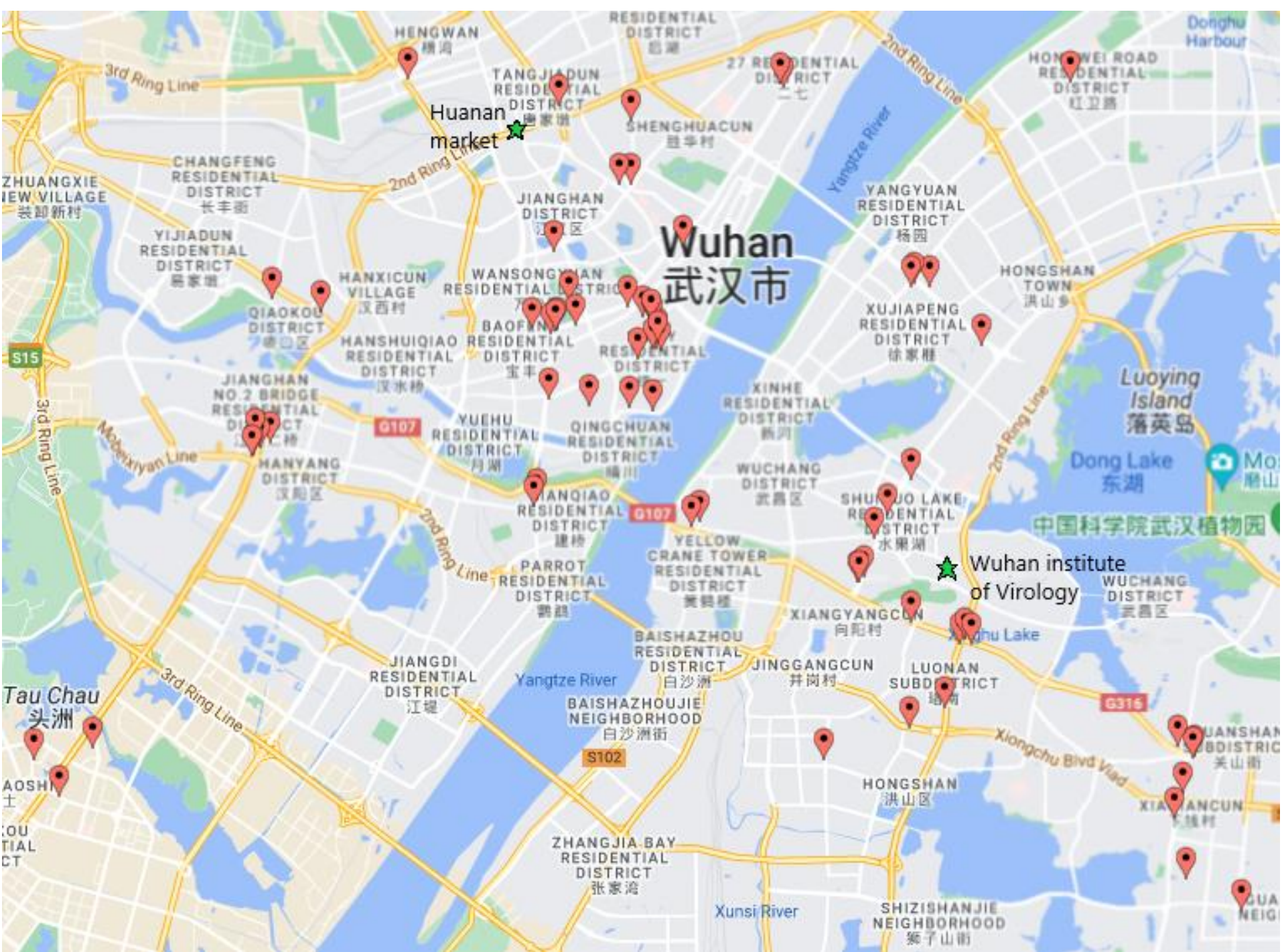
Visitors to locations throughout Wuhan.

Number of social media check-ins in the Sina Visitor System from 2013-2014 as shared by ([33](#)). Number of visitors to individual markets throughout the city are shown in comparison to the Huanan market. Inset: the total number of check-ins to all individual locations across the city of Wuhan, grouped by category. Locations with more than 50 visitor check-ins are shown, and the locations which received more check-ins than the Huanan market in the same period are shown in red.

Here's a map of all those supermarkets and shopping malls.

The Sina visitor system data ranks each of these higher traffic than Huanan market.

And many of these are closer to the WIV than the Huanan market is.



The Worobey paper also looked at other possible superspreading locations:

They flagged 430 locations in Wuhan as high risk superspreading locations after comparing to lists of other superspreading events.

The Huanan market gets 1 / 2,500th of the total traffic to those high risk places.

“Beyond markets, we found at least 1,676 total locations in Wuhan with more visitors than the Huanan Seafood Market. However, some high traffic locations may be less predisposed to COVID-19 superspreader events or substantial spread over a longer period than others: to further quantify this, we utilized a list of known SARS-CoV-2 superspreader locations/events (34) to further subset the following categories of locations that may serve as potential high-risk locations for superspreader events: 'Residential area', 'College', 'Building', 'shopping mall', 'Hospital', 'Middle school', 'supermarket', 'bar', 'Convenience Store/Convenience Store', 'Sports place', 'Comprehensive Stadium', 'church', 'Temple', 'primary school', 'company'.

This subset identified another further 430 locations which may be at higher risk for superspreader events, which received more human visitors than the Huanan Seafood Market. As a fraction of all social media check-ins to the set of 70 markets described above, the Huanan market represented (120/98,146) visits or 0.12%; as a fraction of all social media check-ins to the set of 430 locations similar to those of known superspreader events, the Huanan market represented (120/262,233) or 0.046%. For all four wet markets selling wild animals in Wuhan, these numbers were 206/98,146 (0.21%) and 206/262,233 (0.079%), respectively.

While the potential risk of a location to be the site of an ascertained COVID-19 superspreader event (SSE) undoubtedly depends on many factors beyond number of visitors, there are no reasons to believe the Huanan Seafood Market is at an unusually high risk of a SSE compared with several other locations in Wuhan. COVID-19 cases associated with the Huanan market were not older (and actually leaned slightly younger) than all December 2019 COVID-19 cases on average (7), indicating the market population was not excessively elderly. Further, the main entryways to the market were large and open to the street, indicating a significant degree of airflow through the main thoroughfares.

While the association of social media check-ins and true visitor number likely varies across different types of sites and is likely subject to demographic biases, for the Huanan market to be even a remotely likely random location for a superspreader event within the city of Wuhan would require it to be extremely under-reported in the social media data. The fraction of Huanan market social media visitors out of social media visitors to all markets was 0.12%, or slightly higher than the number of visitors per day officials reported to the WHO mission report (10,000) as a fraction of the general Wuhan population of approximately 11 million (0.09%). Further, the Huanan market specifically received fewer social media visitors than 2 Walmart stores, 2 Carrefour stores, and 1 RT-Mart store, and does not stand out among other large wholesale markets in the city.”

There is nothing unique about the market, other than the wild animals.

- The ventilation isn't abnormally bad:
 - The ceilings are high.
 - Each market street ends with an open entrance, so the building has dozens of open holes for ventilation.
 - Any kind of bar or restaurant is worse than this.
- Any other kind of supermarket is similar – those have more traffic, same or worse ventilation. Those have break rooms where employees socialize (just like the mahjong closet).



The Mahjong room is not some unique place that can incubate covid better than any other place in Wuhan. There are many [Mahjong rooms in town](#):

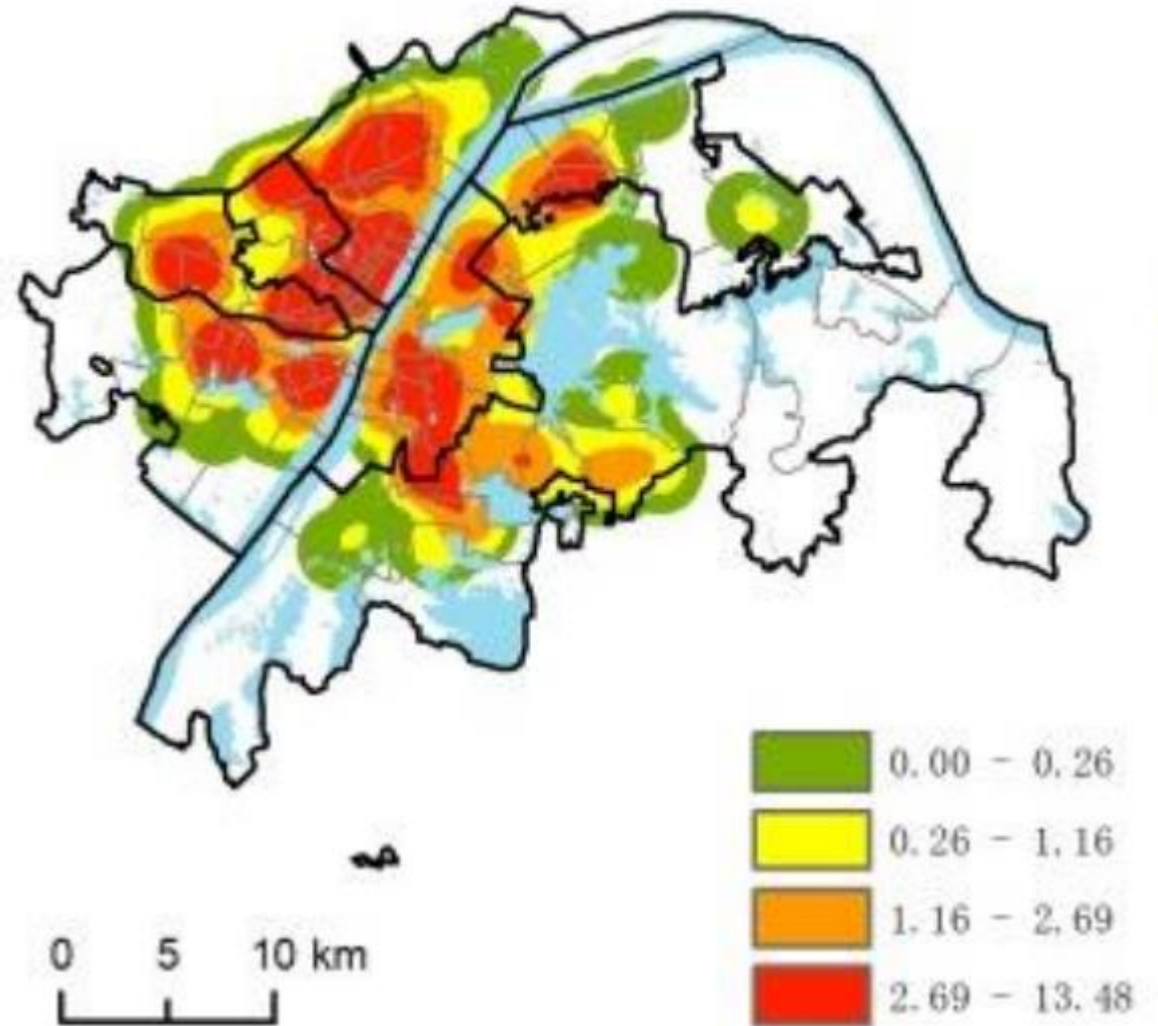
“Chess and card rooms: These facilities provide Mahjong, poker and all types of chess activities. They are popular recreational areas for local residents. With the popularity of board games, chess and card rooms have begun to become popular among youth. There are 838 chess and card rooms in the study area.”

Those places were spread around the city, many were closer to the Wuhan institute of Virology.

Including other “leisure entertainment facilities”, like bars, theatres, cybercafes, karaoke, etc, the study found 2,914 places in the study area.

What are the odds the first super-spreading location just happened to be at the only one of these places that was selling wild animals?

chess and card rooms



What are the odds that the virus would start in this particular mahjong room?

Measured by number of locations: 1 in 2,914.

But, you also need to weight that by total number of visitors, many of those places are likely popular and this room is tiny and most outsiders don't know about it.

Weighted by traffic, it's probably even lower than 1 in 10,000.



I think this may be a [video inside the mahjong room](#)

On the other hand, there are only 4 wet markets in town selling wild animals.

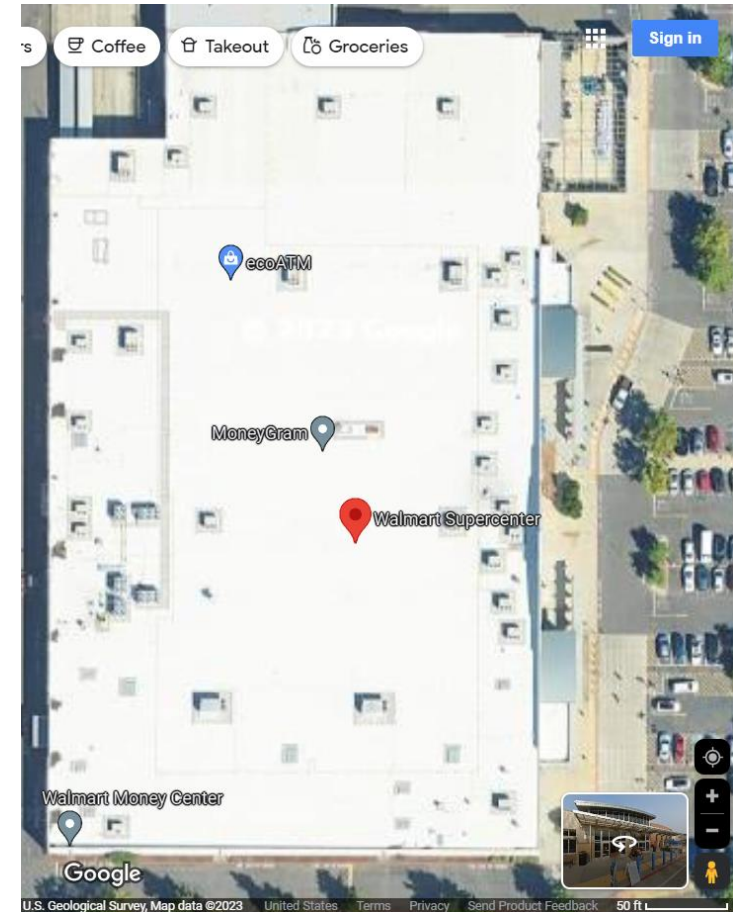
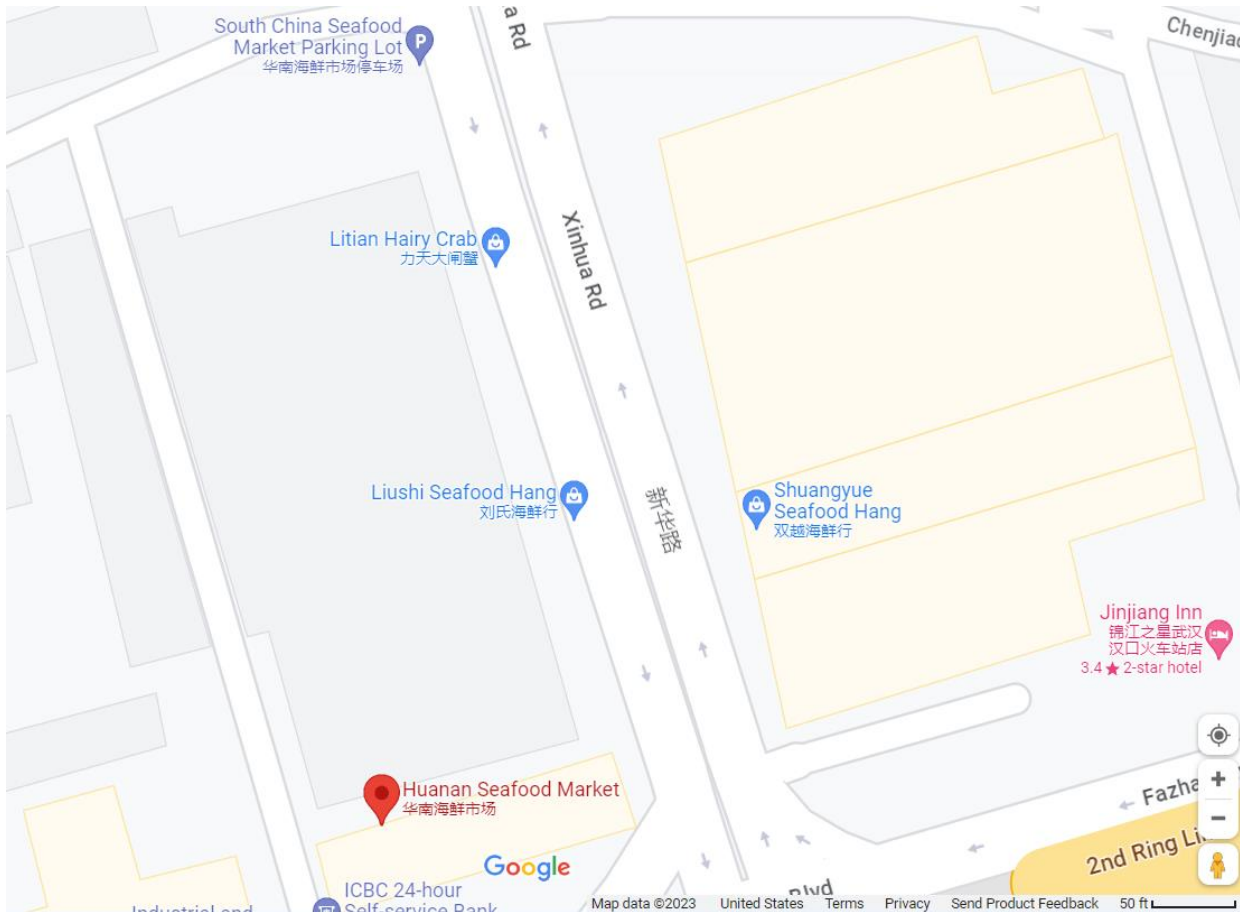
The Huanan market has the most shops selling wildlife of any of those markets.

And wild animals are known to start pandemics.

The Huanan market is just a large shopping area.

The west side of the market is smaller than a Walmart store, the west and east sides combined are slightly larger:

World-wide average customers per Walmart store are about 3,500 per day, but some supercenters get 10,000+/day



Huanan market isn't even the biggest wet market in Wuhan.

Huanan market has 680 shops.

Baishazhou market [has 3,600 shops.](#)

Baishazhou market is also closer to the Wuhan Institute of Virology, it's a [few miles away](#), on the same side of the river.

If there was some unique property of large markets that selectively amplified pandemics, then Baishazhou would still be more likely to be the point of amplification.

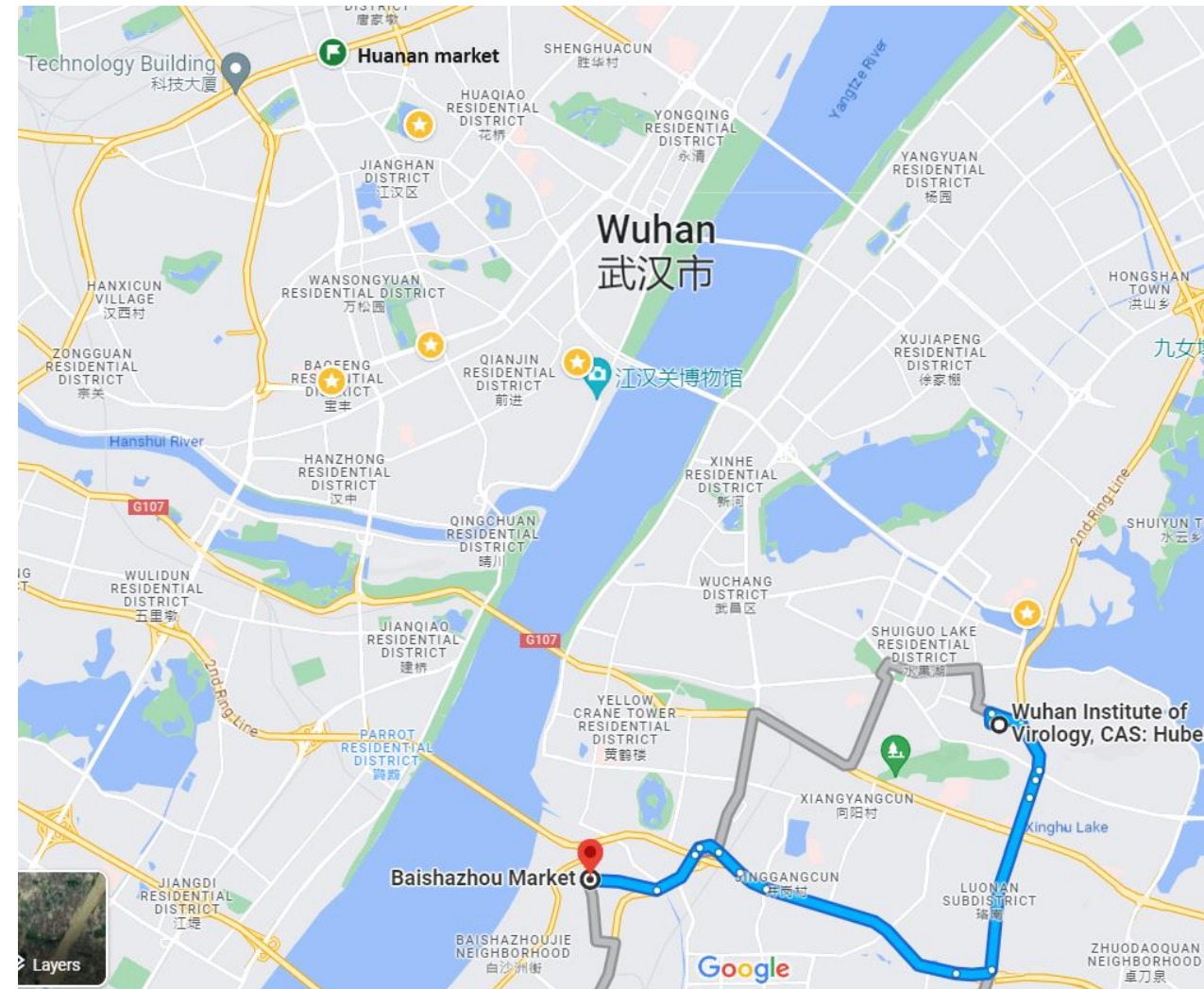
What Huanan market did have is wildlife.

The [Xiao Xiao 2021](#) paper listed 17 shops in Wuhan that were selling wild animals.

Huanan market had 7 of the listed shops.

Baishazhou market had only 2 of those shops.

The other two markets each had 4.



We can make lists of other superspreading events throughout the pandemic, to see where the first cluster is most likely to occur:

Date	Location	Setting	Z
SARS-CoV-2			
Mar., 2020	Washington, US	Choir practice	52
Feb., 2020	Gyeonggi, S. Korea	Social contact	51
Feb., 2020	Chungcheongnam, S. Korea	Social contact	27
Jan., 2020	Ningbo, China	Social contact	25
May 2020	Incheon, S. Korea	Social contact	24
Mar., 2020	Arkansas, US	Religious gathering	17.5*
Feb., 2020	Chongqing, China	Family	13
Feb., 2020	Munich, Germany	Seminar	11
Jan., 2020	Alps, France	Ski resort	11
Feb., 2020	Chicago, US	Family gathering	10

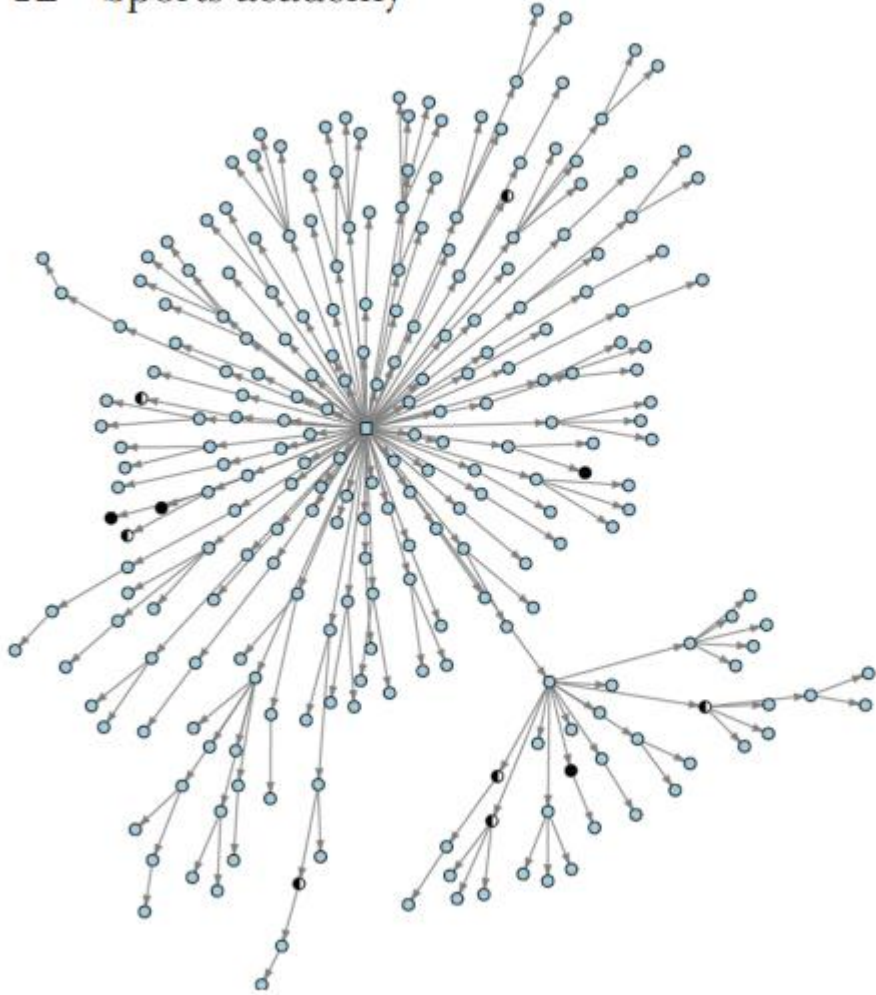
[List source](#)

SARS-CoV-2	January 19, 2020	Ningbo, China	Bus	30
	January 24–28, 2020	France	Resort	11
	February 10, 2020	Daegu, South Korea	Church	38
	Feb 26–27, 2020	Boston, MA	International conference	97
	Monday, March 2, 2020	New York, USA	Hospital (multiple)	90
	Tuesday, March 10, 2020	Washington, USA	Choir practice	52
	June 17-20 2020	Georgia, USA	Summer camp	260 ³
	Monday, August 17, 2020	Paju, South Korea	Coffee shop	56

[List source](#)

A [Korean study](#) (for the 2021 delta wave) found that covid spread well in sports academies, karaoke centers, and schools:

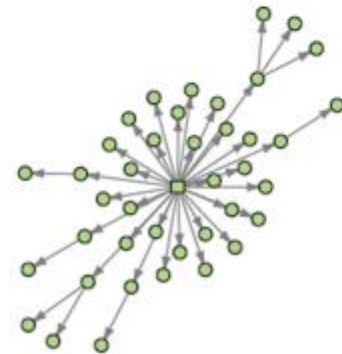
A Sports academy



B Karaoke center



C High school

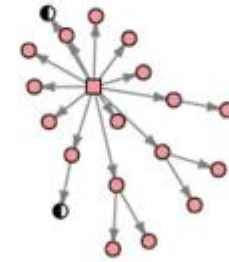


D Other settings

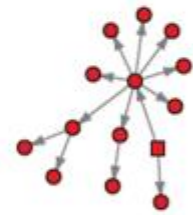
Welfare facility for disabled



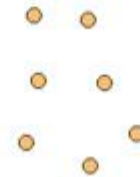
Reading room



Acting school



Sporadic cases



A [Korean study](#) from 2020 showed that religious groups were the largest superspreading events, followed by workplaces, schools, and leisure facilities.

Market was listed as one of the leisure facilities, but it's just one out of 7, and finer data wasn't provided.

Table 1. Characteristics of COVID-19 cluster categories in Seoul, South Korea as of December 4, 2020 excluding “Others” category.

Categories	Infection Locations	Cluster Size	Confirmed Date for the First Case Linked to Cluster	Confirmed Date for the Last Case Linked to Cluster
Religious Groups	Church, bible meeting, temple	1178	25 March 2020	2 December 2020
Convalescent Homes	convalescent facility	215	10 June 2020	22 November 2020
Hospitals	private hospital, university hospital	222	31 August 2020	4 December 2020
Workplaces and Schools	Call center, office, city office, school, academy, distribution center	652	8 March 2020	4 December 2020
Leisure Facilities	Athletic facility, Korean sauna, private meeting, market, teaching center, internet cafe	682	4 March 2020	4 December 2020
Itaewon Clubs	Nightclub	139	8 May 2020	6 June 2020

Covid spread well on ships. It also spread well in nursing homes:

Table 1B. SARS-CoV-2 'isolated' superspreader events.

Setting	Country	Clusters	Total Infected	
Navy Ship	USA	1	1156	29
Navy Ship	France	1	1081	28
Cruise Ship – Ruby Princess	N/A	1	662	71
Cruise Ship - Diamond Princess	N/A	1	620	72
Prison	USA	1	358	73
Elderly Care	USA	1	167	74
Elderly Care	Germany	1	74	75
Elderly Care	France	1	45	75
Elderly Care	Scotland	3	>30	76
Elderly Care	Italy	1	>27	75

[Table source](#)

A [Hong Kong study](#) found that superspreading occurred in bars, weddings, and temples:

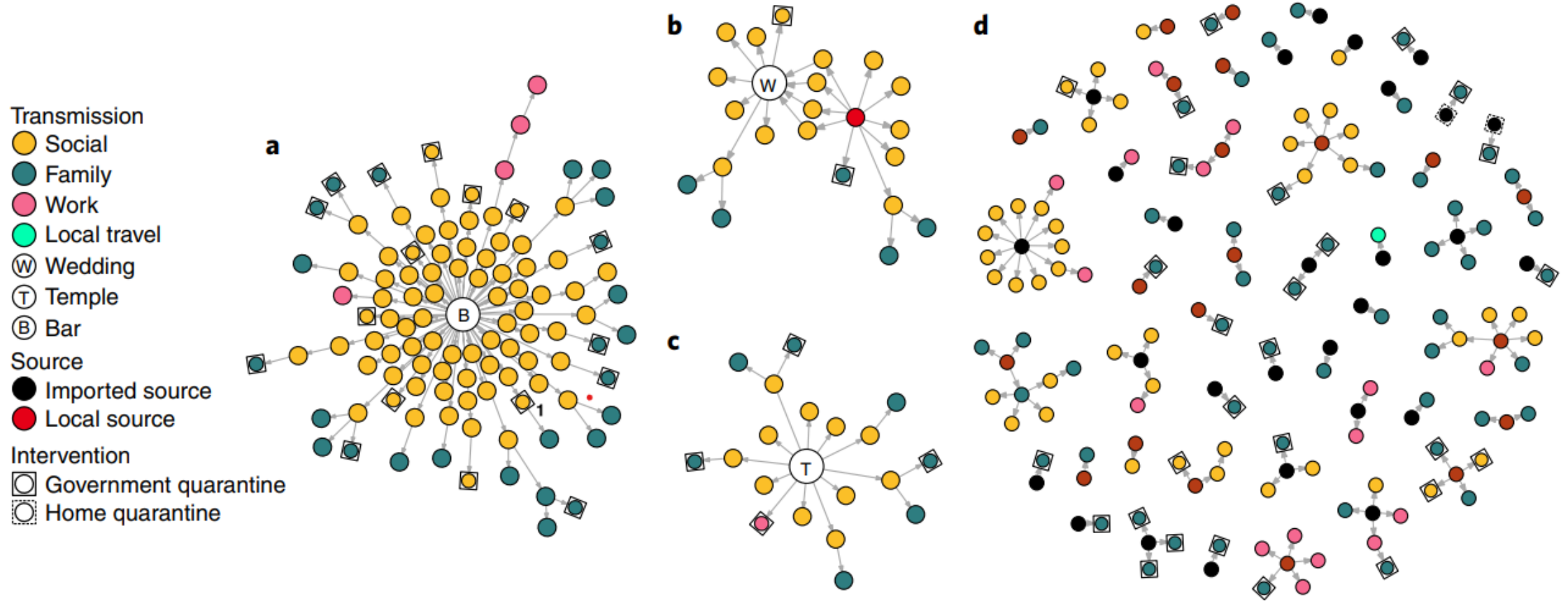
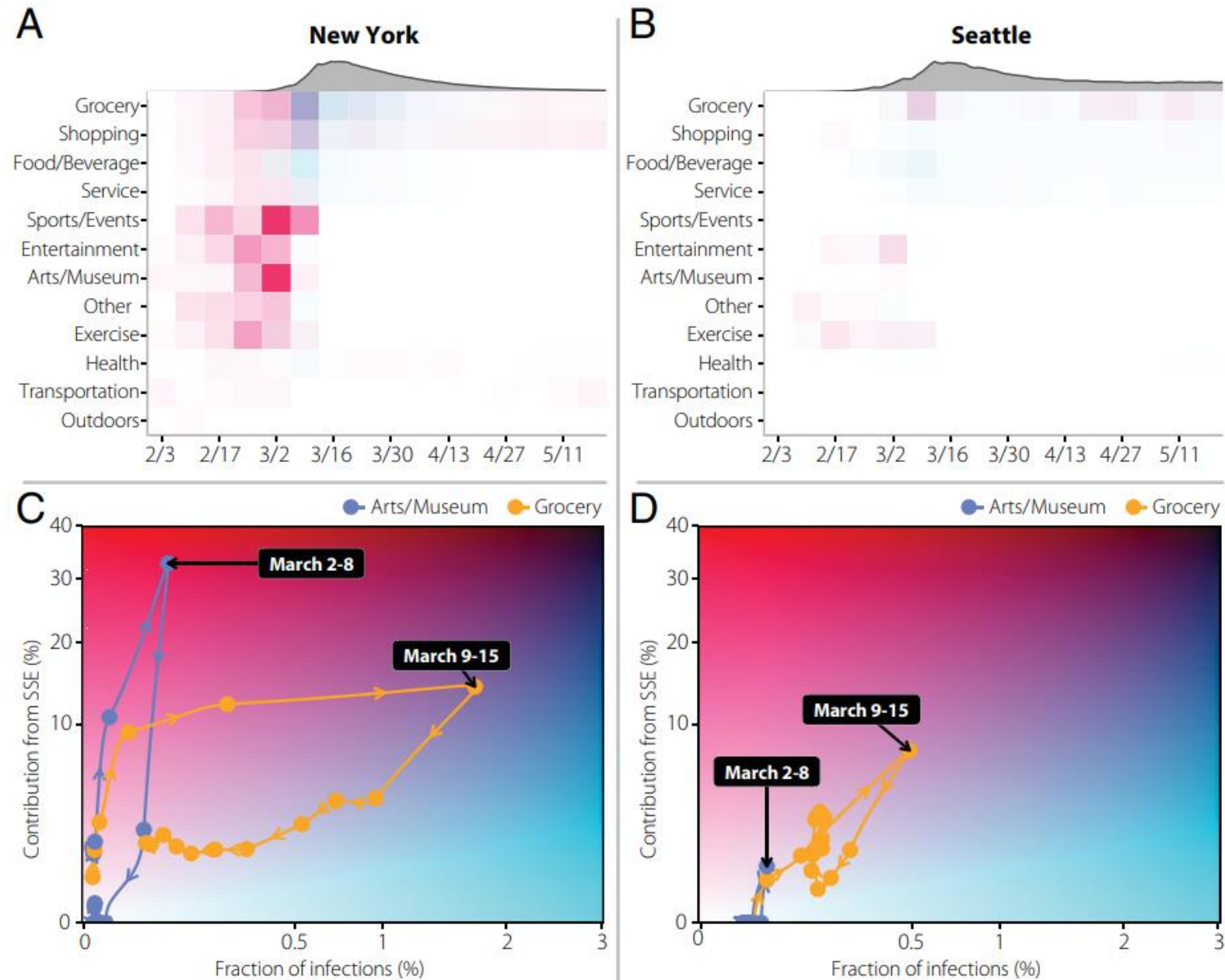


Fig. 2 | Chains of SARS-CoV-2 transmission in Hong Kong initiated by local or imported cases. a, Transmission network of the 'bar and band' cluster of undetermined source ($n=106$). **b**, Transmission network associated with a wedding without clear infector-infectee pairs but linked back to a preceding social gathering and local source ($n=22$). **c**, Transmission network associated with a temple cluster of undetermined source ($n=19$). **d**, All other clusters of SARS-CoV-2 infections where the source and transmission chain could be determined.

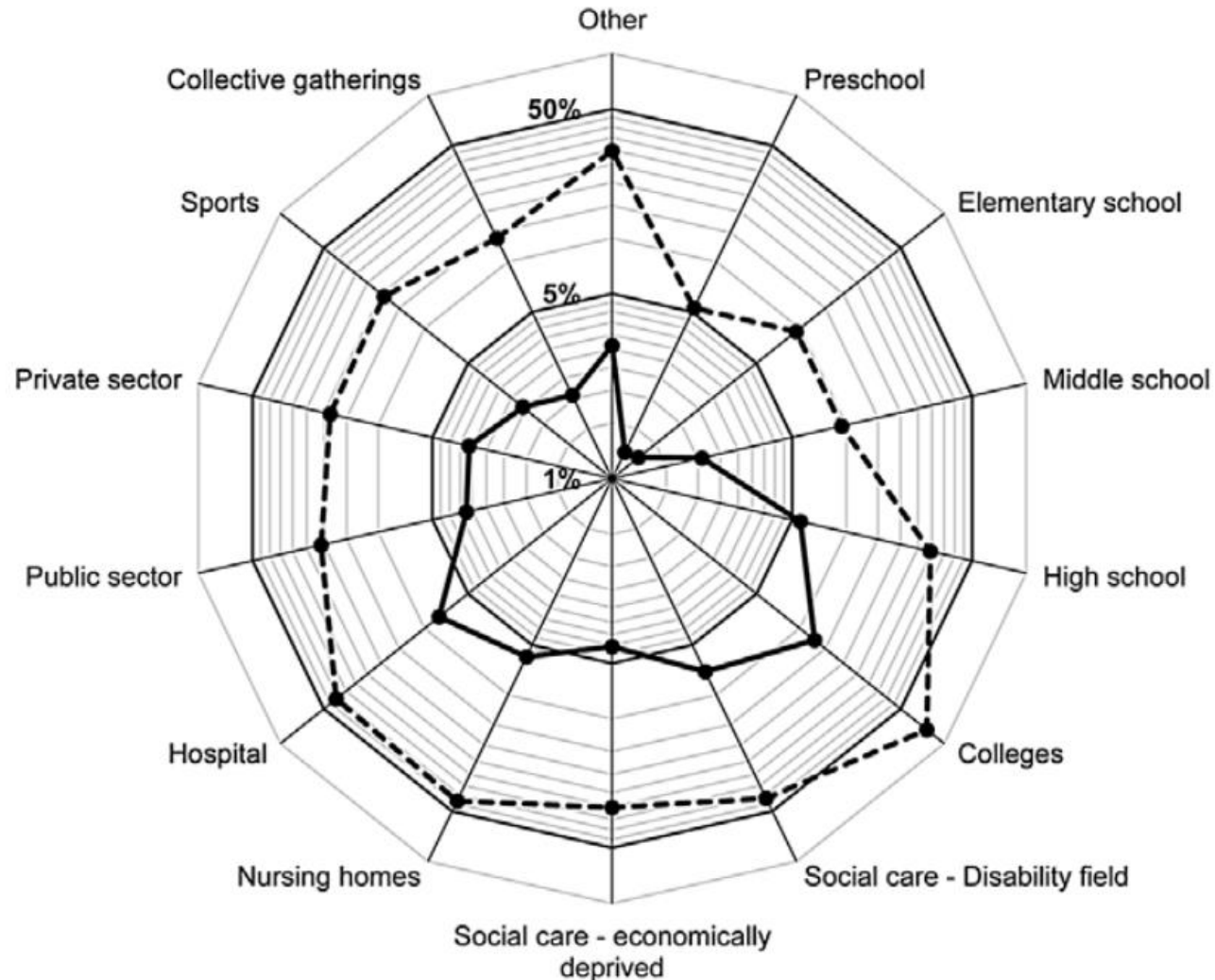
A [US study](#) in New York and Seattle found superspreading at sporting events, museums, and some grocery stores:



A [French study](#) used contact tracing data to identify superspreading locations.

The most common were: colleges, followed by hospitals, disability social care facilities, and high schools.

—●— Proportion of SSE among all events in the setting -●- Proportion of cases occurring in a SSE in the setting



An [Indian study](#) identified super-spreading clusters at several different workplaces, a pharmaceutical company, a convention center, and a steel plant.

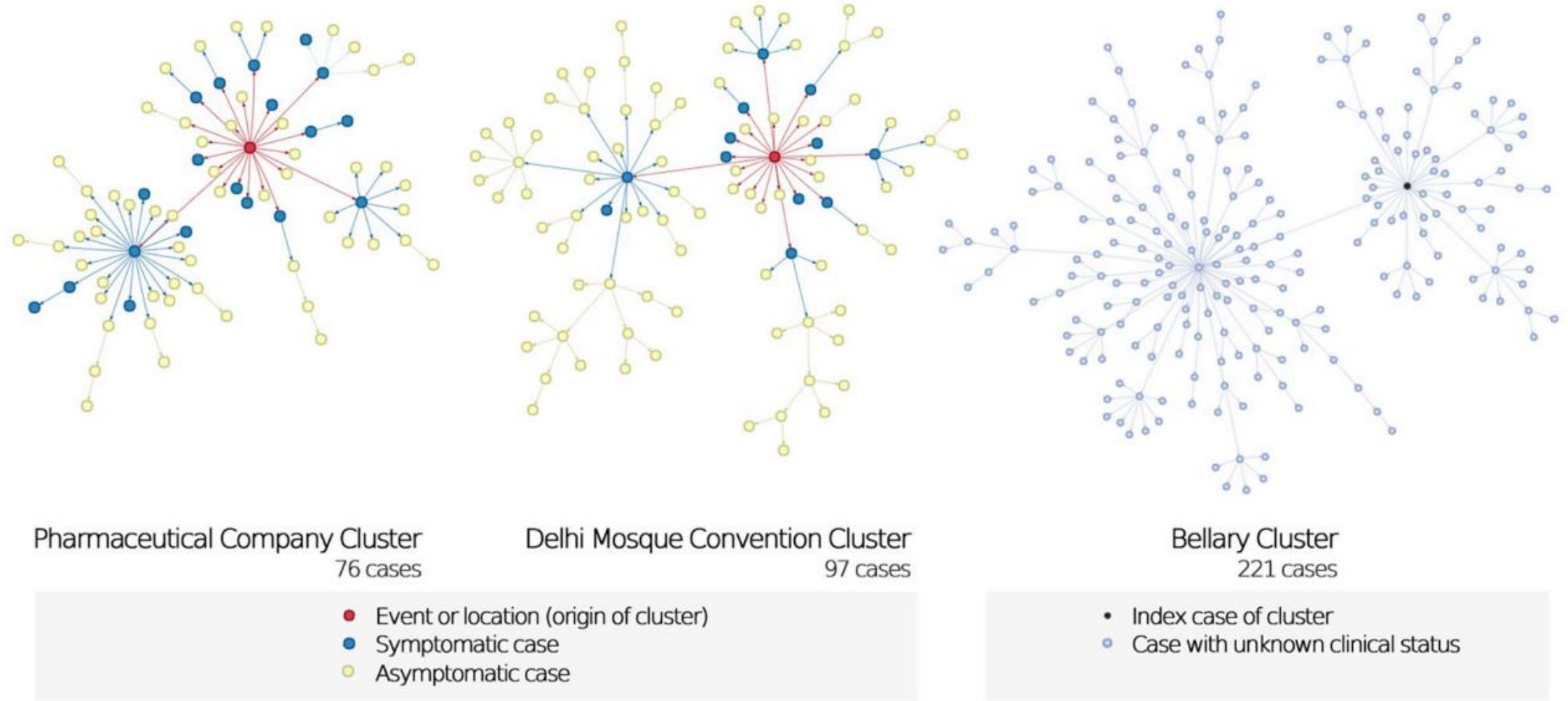
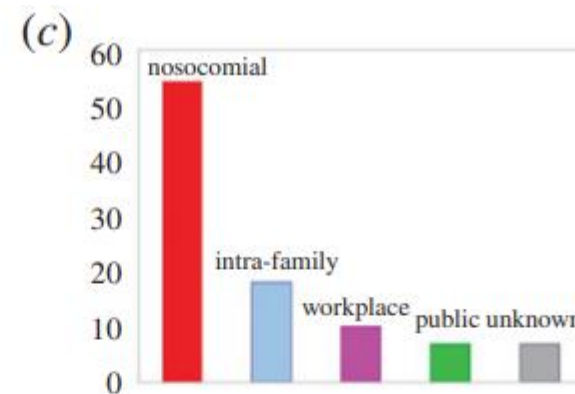
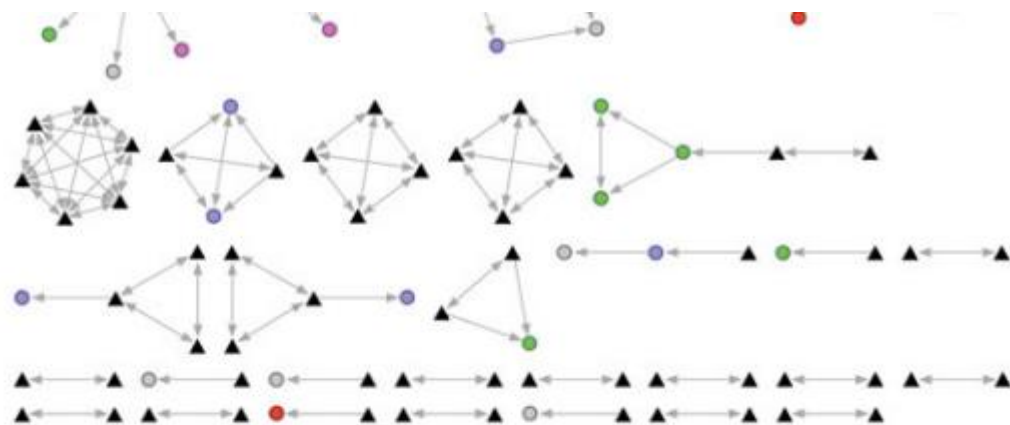
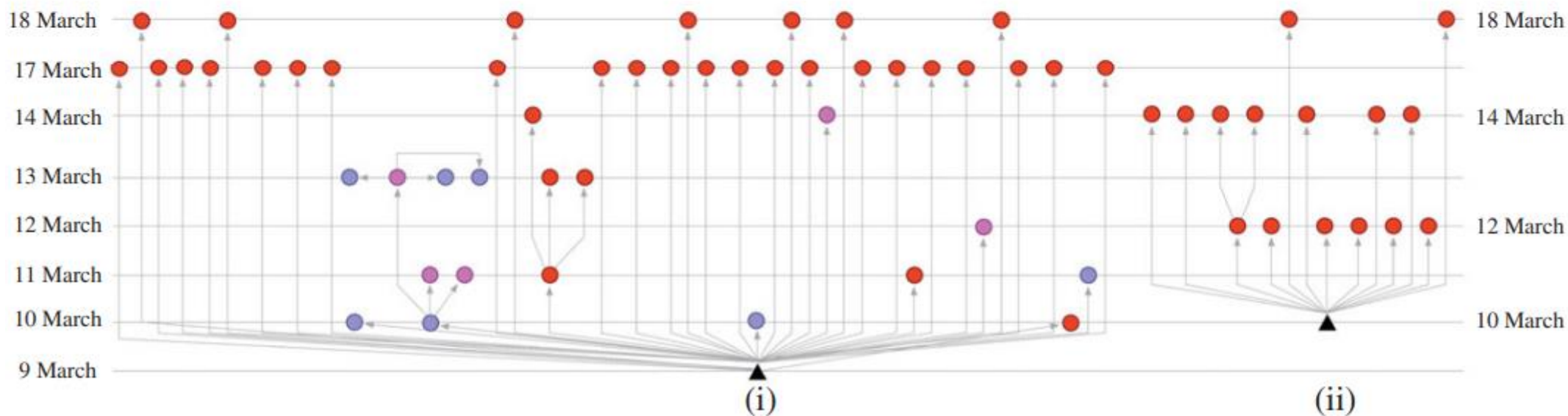


Fig 2. Transmission networks of three largest clusters of SARS-CoV-2 cases in Karnataka, India up to 21 July 2020. The networks indicate the heterogeneity in transmission from infected cases, with a few patients causing most secondary cases. The Bellary cluster occurred in June and July, during which symptomatic status of cases was not available.

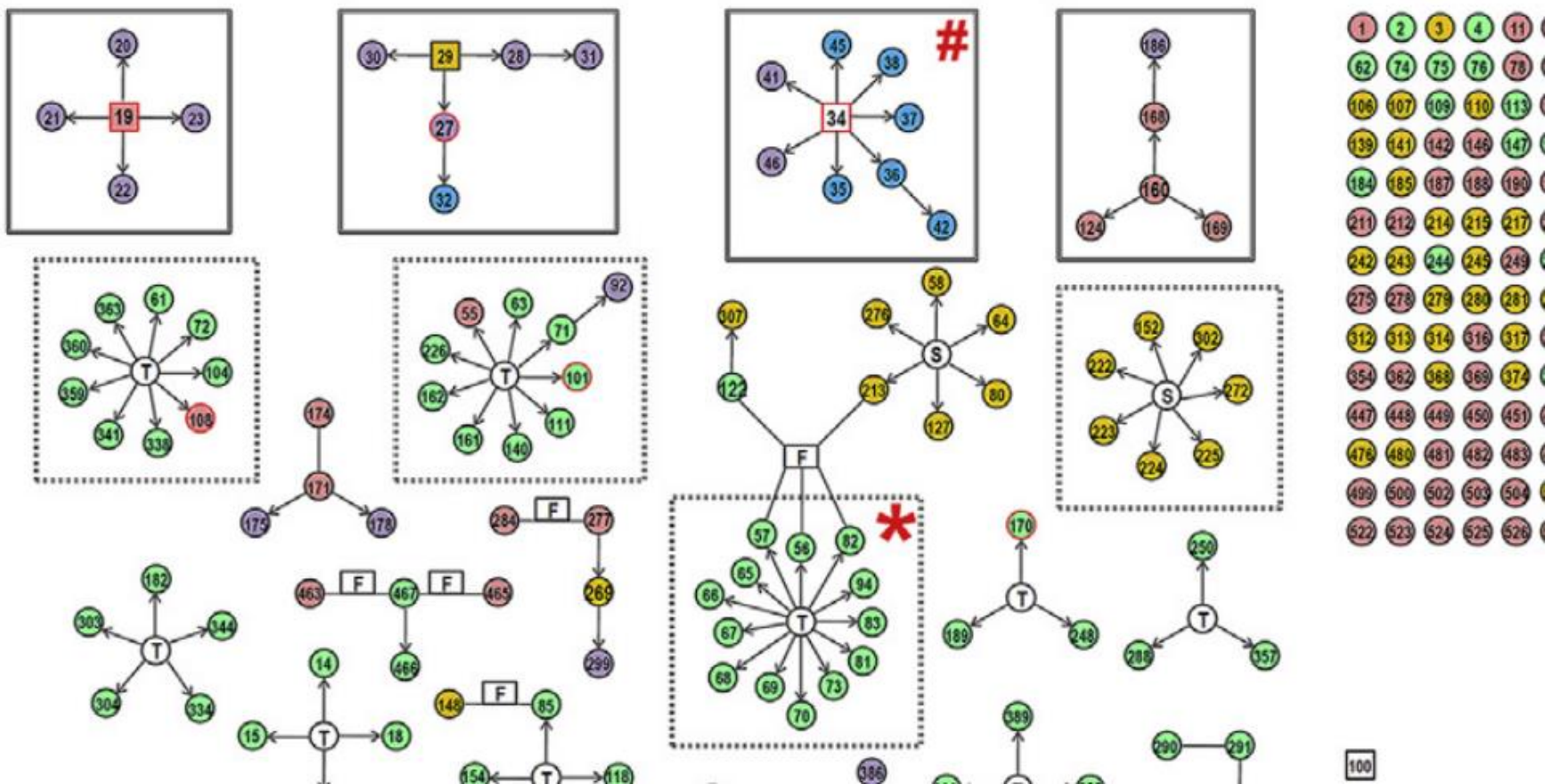
A [Romanian study](#) found lots of super-spreading in hospitals:



(d)



A [Taiwanese study](#) found superspreading events among household and tour group settings



Food processing plants do show up on [one of the lists](#), those sound kind of like a wet market. But the list also includes other shopping locations, schools, hospitals, restaurants, bars, churches, etc.

Table 1A. SARS-CoV-2 'societal' superspreader events.

Setting	Country	Number of Sites	Total Infected
Religious	South Korea	2	4531
Religious	Germany	2	171
Religious	USA	3	150
Religious	Netherlands	1	102
Religious	Singapore	2	33
Religious	Hong Kong	1	19
Worker Dormitories	Singapore	17	1690
Work (Food Processing Plant)	Germany	1	1029
Work (Food Processing Plant)	USA	3	779
Work (Food Processing Plant)	Ghana	1	534
Work (Food Processing Plant)	UK	4	469
Work	Singapore	2	34
Work	China	3	25
School	France	1	133
School	Israel	1	130
School	New Zealand	1	96
School	Singapore	1	26

Shopping	Singapore	3	122
Shopping	China	3	44
Hospital	South Korea	1	118
Hospital	China	1	54
Hospital	Japan	2	20
Bar	Japan	6	100
Bar	Zurich	1	6
Bar	New Zealand	1	77
Bar	Germany	1	16
Bar	Austria	1	15
Bar	Hong Kong	4	106
Building Site	Singapore	3	90
Conference	USA	1	89
Conference	Japan	1	10
Sport	South Korea	1	65
Sport	Japan	2	20
Meal	Singapore	2	55
Meal	China	2	19
Aircraft	Dubai→Hong Kong	1	26
Skiing	France	1	11
Funeral	India	1	16
Wedding	Hong Kong	2	22
Bus riders	China	1	30

Meatpacking plants are colder, more crowded, and often have a single shared airspace:



The Huanan market is just several streets of shops, not a single airspace, not as cold, variable crowding levels, more visitors than residents:



The Huanan market might be in the 50-60 Fahrenheit range, if it matches [outdoor temps in Wuhan](#), or it could be warmer than that:

Wuhan - Average temperatures (1991-2020)						
Month	Min (°C)	Max (°C)	Mean (°C)	Min (°F)	Max (°F)	Mean (°F)
November	7.6	16.8	12.2	46	62	54

Here are some [guidelines for meatpacking plants](#):

IV. PLANT REFRIGERATION

Adequate refrigeration is one of the most important means of controlling the growth of microorganisms. It is imperative that sufficient refrigerated space be provided to properly handle carcasses and product.

All perishable products should be stored in areas with a maximum temperature of 37° F. Ideally the temperature of processing rooms should not exceed 50° F. In those situations where the temperature requirements would be difficult or impossible to obtain, work surfaces must be cleaned every 4 hours during operation.

The people working at the market weren't abnormally old.

Rootclaim says that the median age of people who got sick at the market was 56:

**The median
age of those infected at the
market was 56**

**But if you read the WHO report, you'll find out that
the median age outside the market was also 56:**

Among the 174 cases, 98 were male and 76 were female. The male-to-female ratio was 1:0.78, the median age was 56 years old, the oldest was 92 years old, the youngest was 22 years old, and the interquartile range was from 46 to 67 years old, as shown in Fig. 5.

Compared with the entire population in Wuhan by age and gender (Fig. 6), the "40-", "50-" and "60-" age groups accounted for a higher proportion among the 174 cases (70.7% vs 42.3%).

And the market linked cases are mostly in their 40's and 50's, they're actually younger than the average covid case found in Wuhan.

I calculated a median age from this table and came up with **45 at the market.**

That was assuming that all the 20- age group was 25, the 60- age group was 65, etc.

Even if you assume all the 20- age group are 29, the 60- age group are 69, etc., then the median age at the market would be 49.

Whereas outside the market, the median case is 56.

So the entire Rootclaim case, regarding age, falls apart when you actually look at the data.

source: [WHO report annexes](#), page 177

Most of the market-related cases are occupational population (Table 3), mainly male, mostly between 40 and 60 years old. 94% of the cases related to the Huanan Market were directly exposed. Among them, market vendors and purchasers accounted for 77% of all cases, which was significantly more than passers-by, buyers, deliverymen, visitors, and those who were indirectly exposed to the Huanan Market, indicating that the way and frequency of exposure in the market may be related to the morbidity risk.

Table 3. General information of cases.

		Number of cases (N=55)	Composition ratio (%)
Gender			
	male	37	67
	female	18	33
Age			
	20-	3	5.5
	30-	7	13
	40-	20	36
	50-	16	29
	60-	7	13
	70-	2	3.6
Types of exposure			
Direct exposure			
	Vendor ^a	30	55
	Purchaser ^b	12	22
	Passer-by ^c	5	9.1
	Buyer ^d	3	5.5
	Deliveryman ^e	2	3.6
	Visitors ^f	1	1.8
Indirect exposure			
	Contact of the Huanan Market exposed population	2	3.6

If patient zero was at the lab, the number one place you'd expect to find a cluster is at the lab.

Other than that, superspreading can happen at any crowded location.

The number two place might be at the hospital that the “3 hospitalized WIV researchers” went to.

Beyond that, it gets harder to say. Any one of the lab workers could get their family sick.

Their children could cause an outbreak at a school.

The lab workers could infect any restaurant or bar in town.

They could infect any place they went shopping, and the Huanan market is way down the list of most popular places to shop.

Shi Zhengli [supposedly enjoys karaoke](#), singing events have been good at spreading covid.

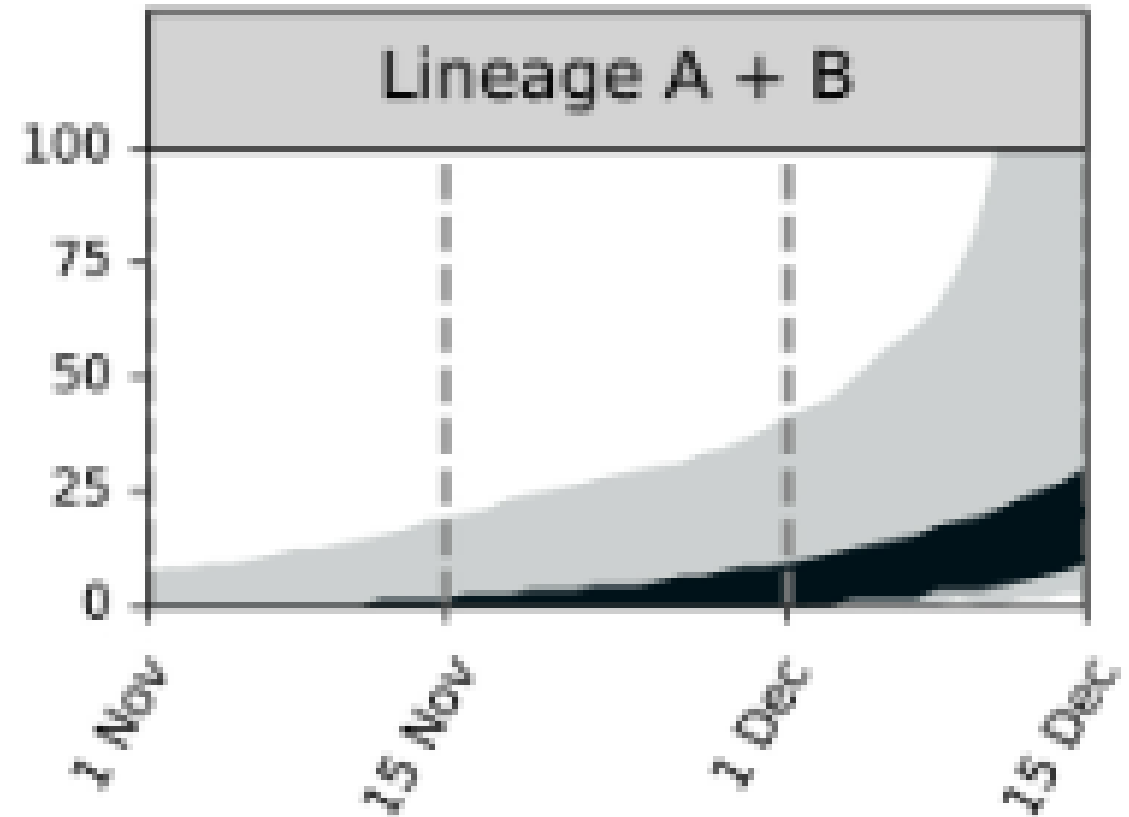
Even if the market was the best place in Wuhan for spreading covid for some unknown reason, you still need lab workers to go to many places to find out which place spreads covid best.

These 5 people at the market got sick Dec 10th to 13th. They got infected maybe a week earlier than that, so the market was infected somewhere around Dec 3rd.

At that time, there might be only 10 people infected with covid in Wuhan (50% confidence) or less than 40 (95% confidence), according to Pekar's models based on genetic diversity.

Simple exponential math of the outbreak gives a similar estimate of the number of early December cases.

So the odds that a lab worker, or one of the first 10-40 cases infected by the lab leak, would randomly go to the market, vs. some other potential cluster location, are still very low.

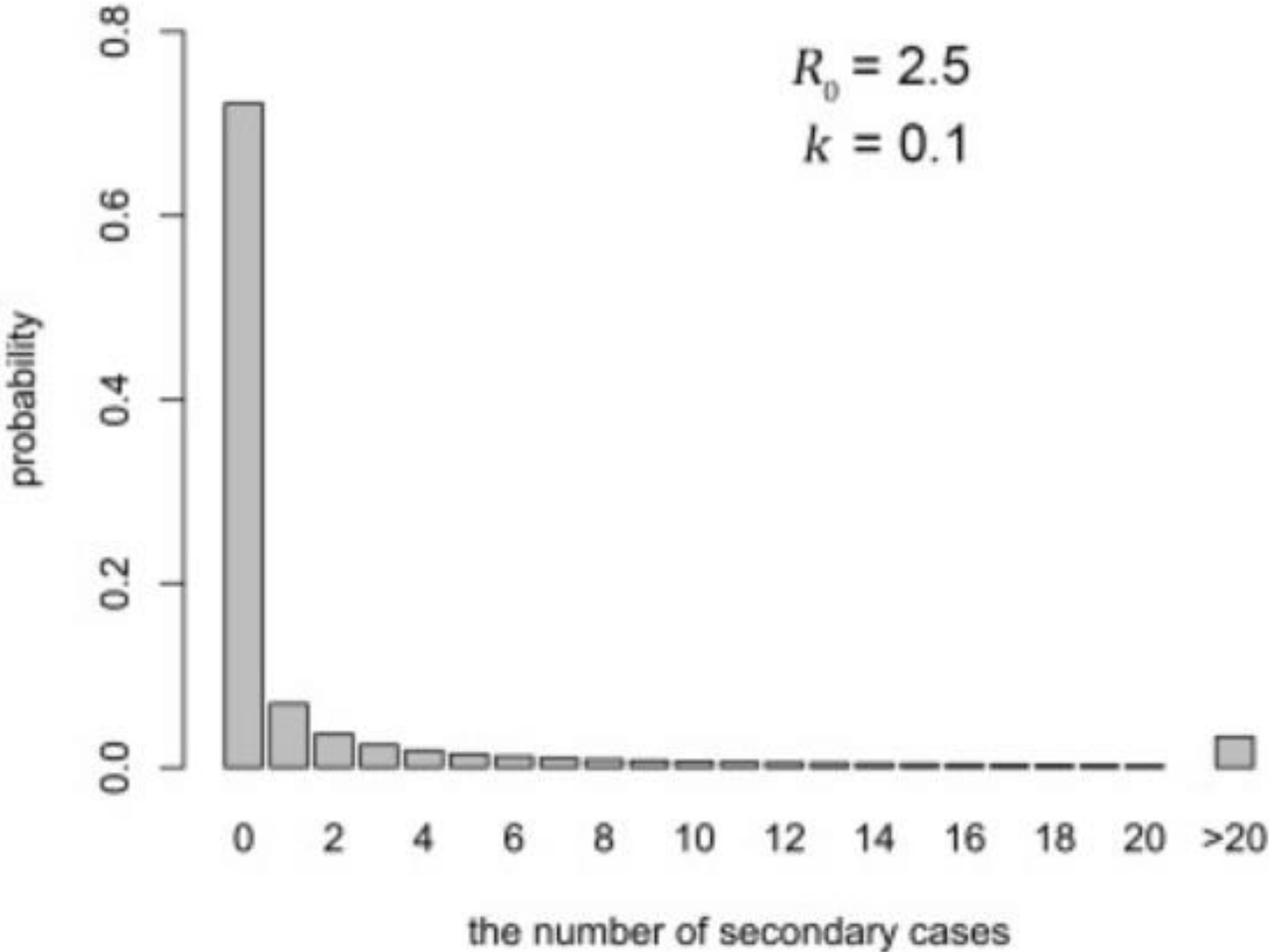


Predicted cumulative number of covid cases. Figure 4 from [Pekar et al 2022](#), light gray is 95% confidence interval, black is 50% confidence interval

Covid does not do a good job at spreading one person at a time.

Most people infect zero others, a few people cause clusters. Covid either grows exponentially or it goes extinct.

It can't just transmit at a low level for a long time before it stumbles upon the market.



[Image source](#)

Rootclaim had a lot of different objections to the market odds.

Another objection was that the traffic to the Huanan market was based on social media check-ins and this was biased against the market.

It is a valid question if this data is representative. I'd point out that people also aren't likely to check-in at the supermarket, a subway station, or many other places. So it's hard to quantify the bias here.

Rootclaim offered an estimate that 10,000 people per day visit the Huanan market. Some other lab leak supporters say [it's 7,000](#).

Since 10 million people live in Wuhan, you might guess the odds are actually 1 in 1,000 of the virus spreading at the Huanan market.

That's possible, though we don't have similar data from other locations to compare against.

Also, the average person visits more than one place in a day, so the denominator is 10 million * (average number of places visited per day).

If each person goes to 5 places per day (maybe a job, 2 subway stations, a market, a restaurant), then the Huanan market is still 1 in 5,000.

I should clarify one more thing:

I'm giving 1 in 10,000 odds that the virus "spreads at the market and nowhere else". That is: no outbreak at the lab, no other outbreak, just the one at the market. I could build a better model based on covid transmission rates to calculate the exact odds here that the lab worker would not infect the lab, their family, etc, and only infect the market.

10,000 could actually be too low, when you consider the fact there's no early outbreak at the lab or elsewhere.

Or it could be somewhat too high, after accounting for social media bias in the traffic data or the relative odds of superspreading at each location in Wuhan.

Those two factors work in opposite directions.

1 in 10,000 is probably about right, as a good estimate for how unlikely the market is for the first outbreak, assuming covid started as a lab leak from the Wuhan institute of virology.

We can do more work to come up with the exact traffic numbers across town, but the most important thing to note is the discrepancy between our two analyses:

I think this is a very unlikely coincidence.

Rootclaim's analysis assumes there is a 100% chance the virus went straight from the lab to the market.

I'm not sure how far we got with discussing Lineage A/Lineage B

These were the claims Yuri made in week 2, I think I rebutted each of these:

Lab leak claim:

- There are actually intermediate genomes
- Lineage B may have started at the market but Lineage A came first
- Actually, proCov2 came even earlier than Lineage A
- The Lineage A sample at the market is mutated/fake/unimportant
- Even if there were 2 lineages, the 2 lineages came from the lab.

My rebuttal:

There's no convincing evidence of any intermediates. Every genome Yuri pointed out came later in the pandemic, and was from a single lab in Sichuan which used a bad software program.

The genetic clock says otherwise.

proCov2 is a random reversion, and not found early.

The lineage A sample is real and important. Both lineages found at the market supports zoonosis.

Very unlikely to come to the market twice, in that case I get to square the 1 in 10,000 odds that the virus came from the lab to the market.

Here's the key evidence I'm claiming for 2 lineages:

Lineage A was found at the market. The first 2 lineage A cases were also found closer to the market than expected by chance, if you think covid was actually all over town. ($p = .001$)

I'm calling that a [bayes factor of 50](#).

One of those 2 cases was found before December 29th – that was actually the first patient diagnosed by Zhang Jixian, before she found any market linked patients. She just thought their symptoms looked like SARS. It was only retrospectively noticed that they lived so close to the market. So that location can't be biased.

The other patient was not linked to the market, so I don't understand how that one could be biased, either.

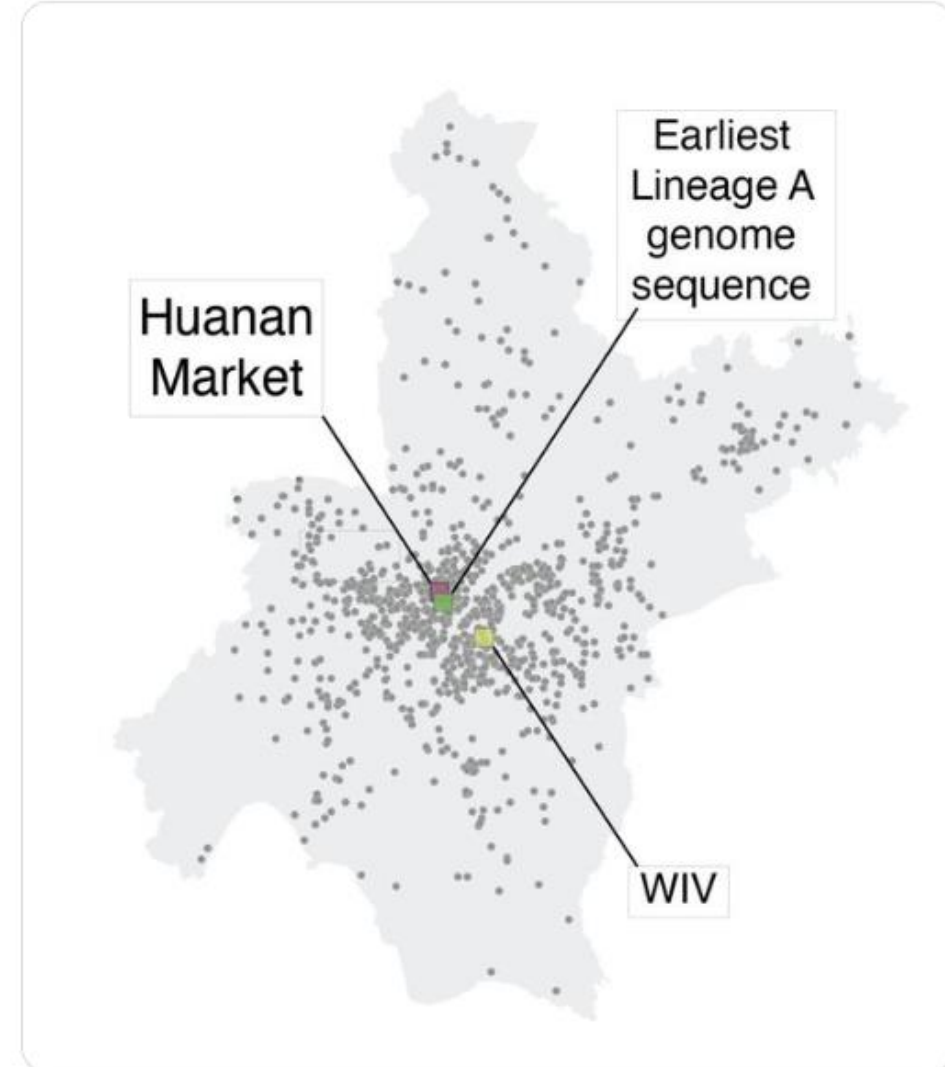
It was retrospectively noticed that they spent 5 days at a hotel very near the market, before they got sick.

The earliest 3 lineage A sequences found anywhere were these 2 patients and the lineage A sample taken from the market.



Here is a map of Wuhan with 1000 residential locations sampled from the 'null' distribution (population density in 100m X 100 x squares across all ~8000 sq km).

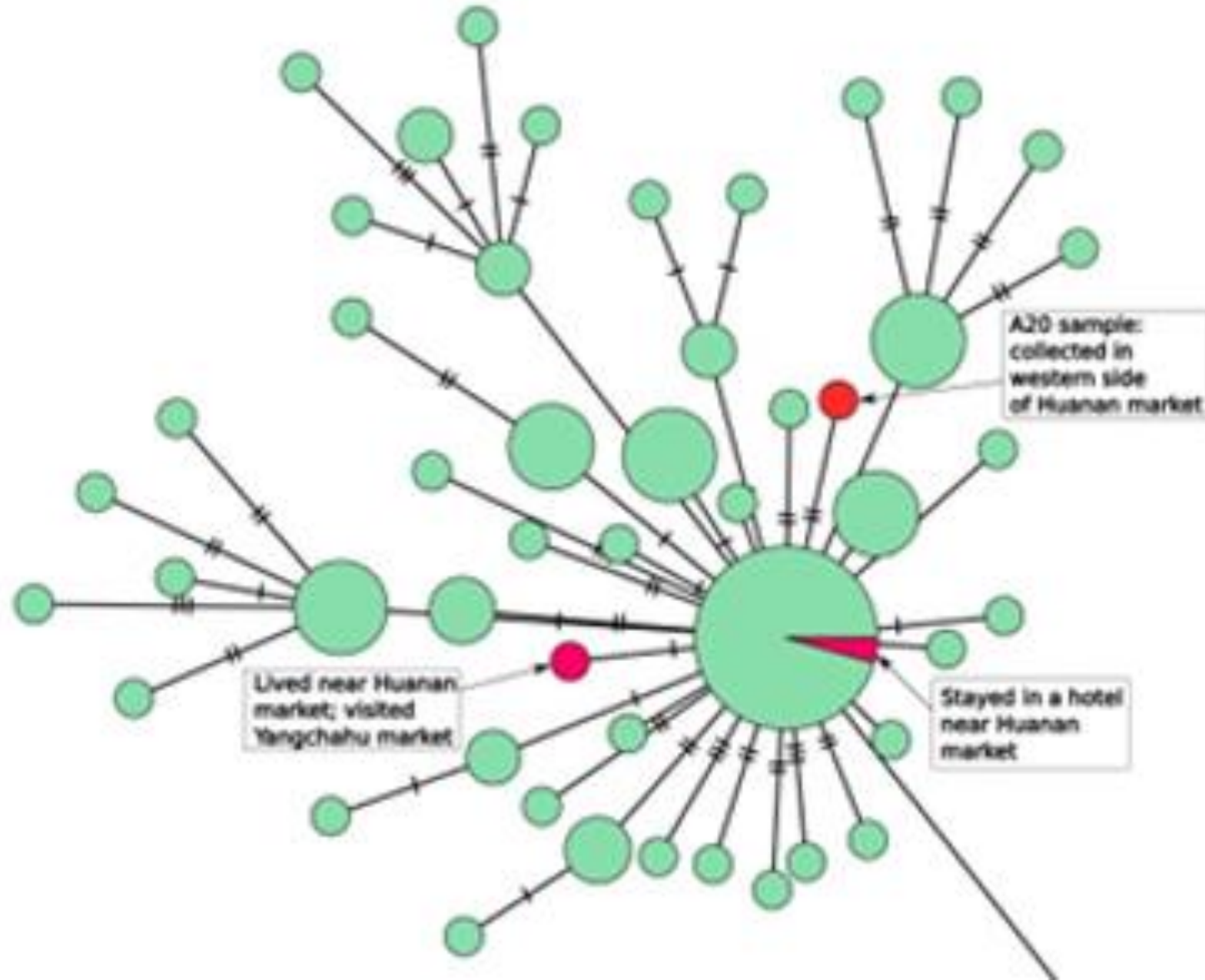
The earliest lineage A case lived closer to the market than expected for someone in Wuhan ($P = 0.034$).



Lineage A was found at the market, in 1 out of 4 samples sequenced
 (or maybe 2 out of 11, including those samples with very low read counts) .
 One of those samples was taken very close to the raccoon dog shop.



Those two early cases were at or near the root of the Lineage A genetic tree:



**The genetic clock says that Lineage B came first.
If lineage A came first, it should have evolved more than lineage B.**

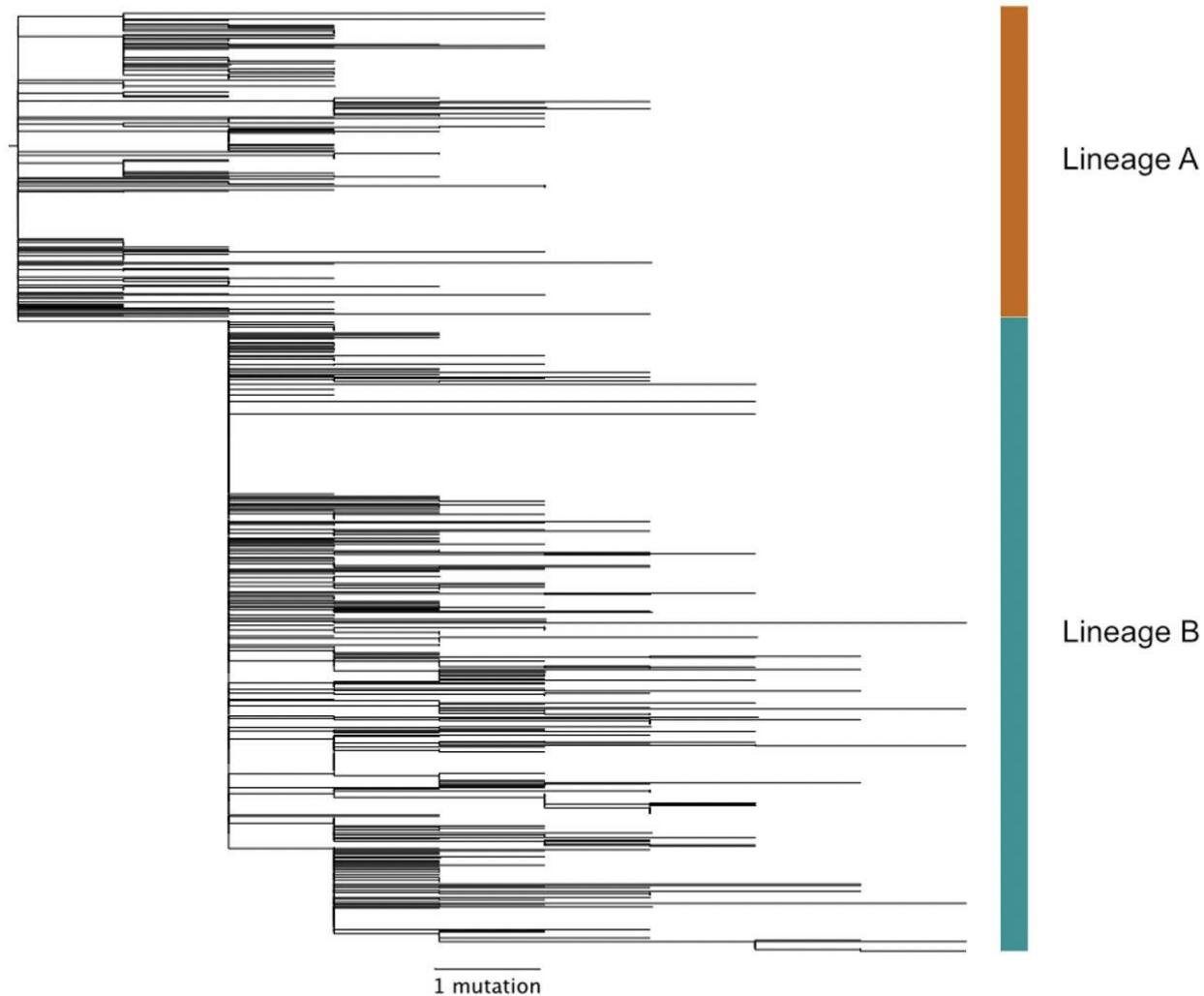
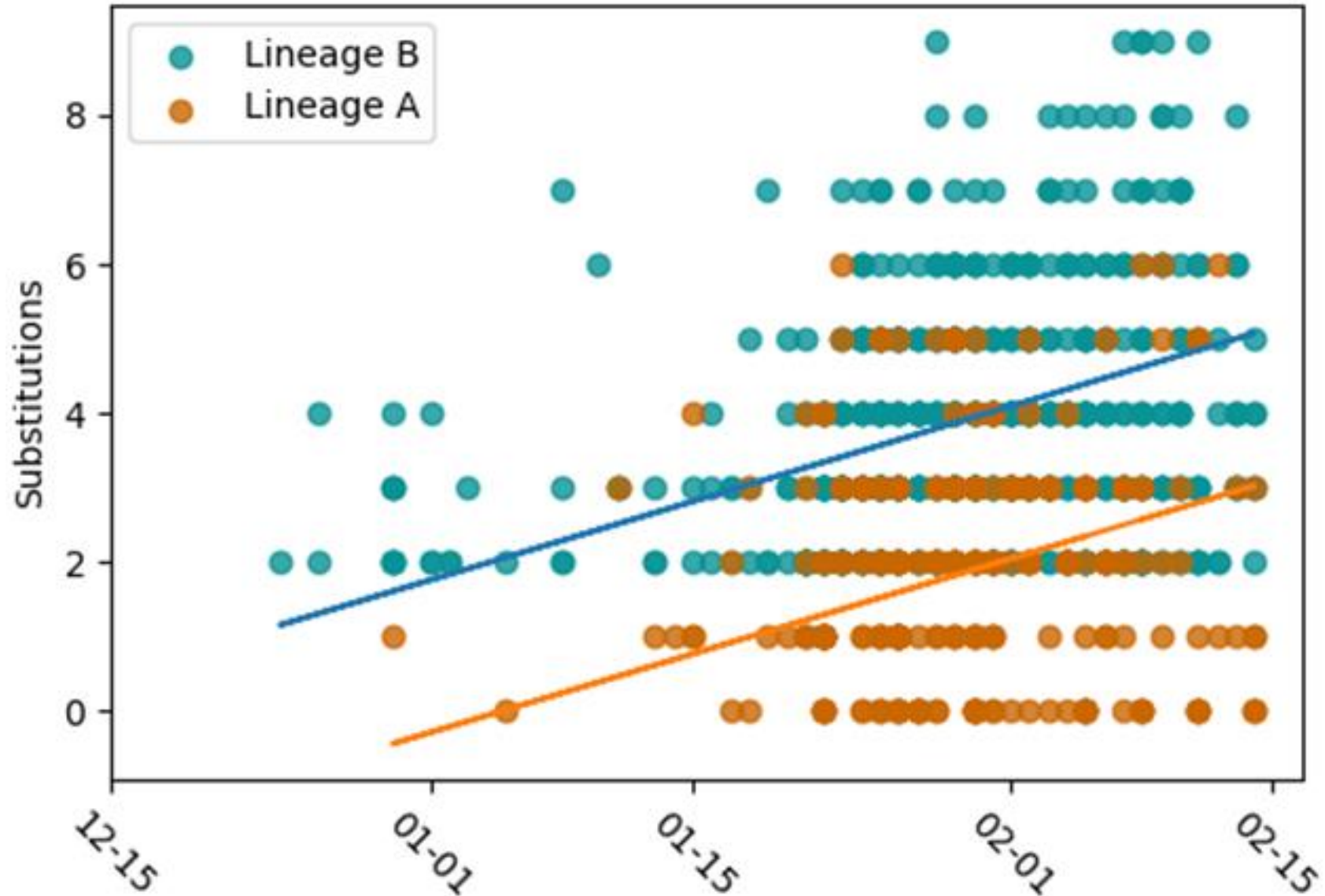


Figure S19. SARS-CoV-2 maximum likelihood tree rooted on lineage A (n=787 taxa, through 14 February 2020).

The genetic clock says Lineage B came before Lineage A

B was found earlier, has more diversity earlier, and continued to have more diversity than lineage A.

There are twice as many Lineage B cases than Lineage A cases everywhere in Wuhan, and also outside Wuhan.



For that, I'm claiming a bayes factor of 11, based on Pekar's model.

I'm being conservative here and excluding the market cases from the analysis, if you include those it's bayes factor 48. I'm excluding those because we've already decided that the market is the first major cluster, so I want to make sure I'm not double counting.

Pekar's numbers are consistent with other papers, like [Pipes et al 2021](#), which gives 96% odds Lineage B came first.

Haplotype	Mutations from Hu-1 reference	Representative genome	Phylogenetic analysis	
			Unconstrained (%)	No market (%)
B (C/T)	N/A	Hu-1	80.85 [†]	62.96 [†]
A (T/C)	C8782T+T28144C	WH04	1.68 ^{**}	5.73 ^{**}
C/C	T28144C	N/A	10.32 [*]	23.02

And lastly, I'm claiming a bayes factor of 4.2, because of the two lineages.

I think that's actually conservative and low, based on the many real world examples I found where 1 introduction = 1 polytomy.

So the combined odds for 2 lineages at the market are about 1 in 2,000

$(4.2 * 11 * 50 = 2,310)$

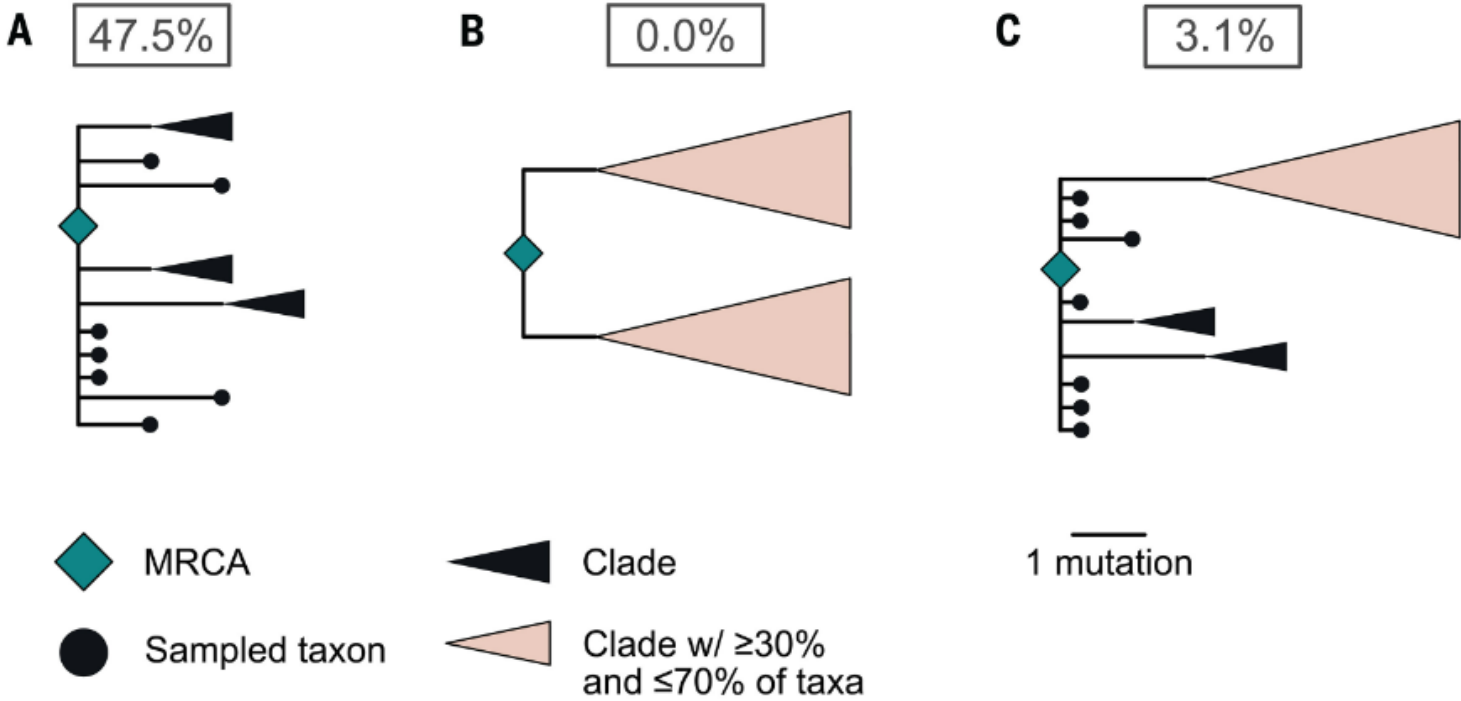


Fig. 2. Probability of phylogenetic structures arising from a single introduction of SARS-CoV-2 in epidemic simulations.

(A) A large polytomy of at least 100 descendent lineages, which is consistent with the base of both lineages A and B. (B) Topology matching a C/C ancestral haplotype: two clades, each one mutation from the ancestor, both with polytomies of at least 100 descendent lineages. (C) Topology matching either a lineage A or lineage B ancestral haplotype: a basal polytomy with at least 100 descendent lineages, including a large clade separated by two mutations, also possessing a polytomy of at least 100 descendent lineages. Basal taxa have short branch lengths for clarity. The probability of each phylogenetic structure after a single introduction is reported in the respective boxes.

No other theory for lineage A and B explains the facts as well:

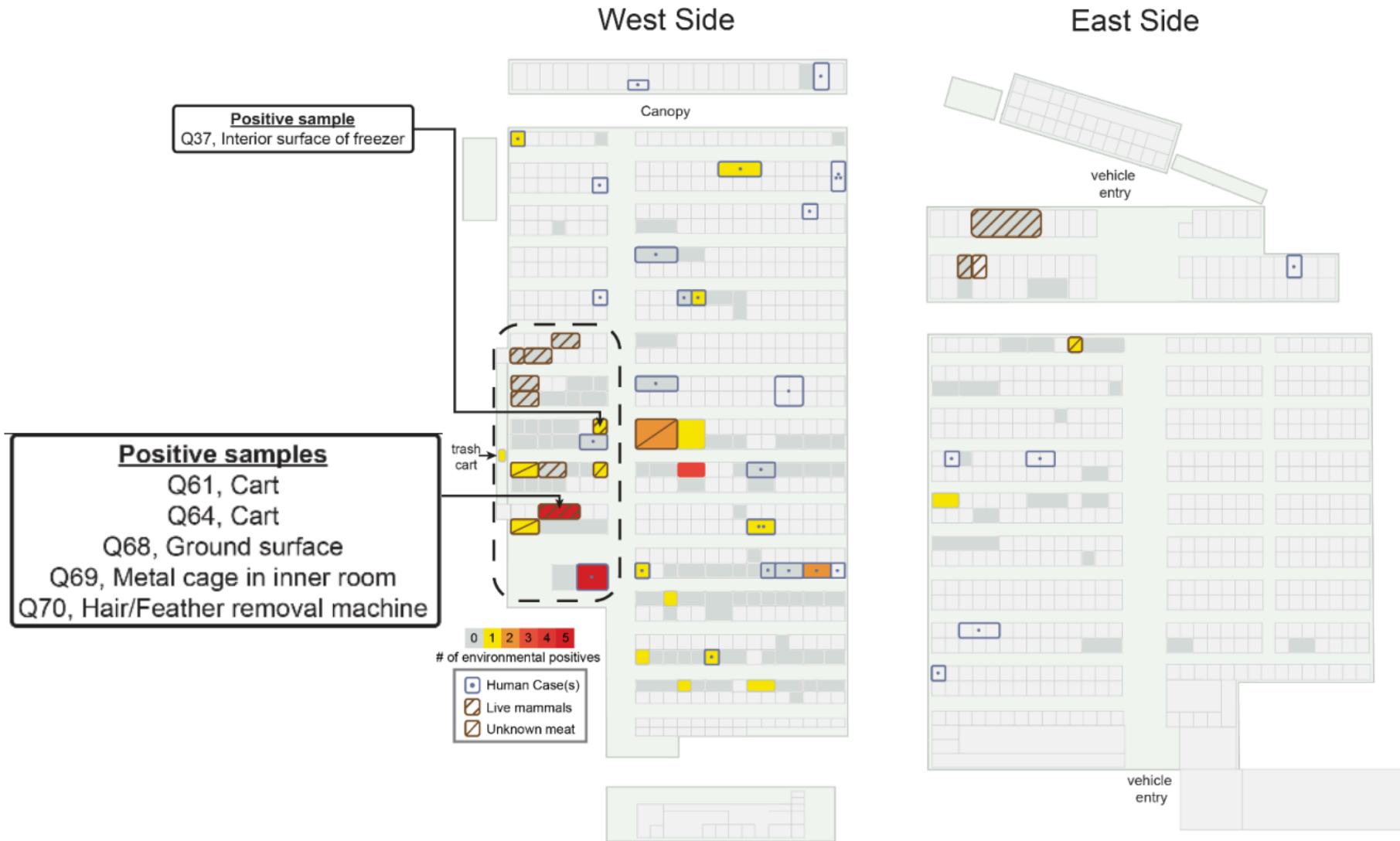
- If lineage A didn't start at the market, why are the earliest lineage A cases so close?
- If you think lineage A was widespread before the market, then the viral diversity of A should be higher than B.
- If you think lineage A was low prevalence before, then you have to explain why the market was the one and only superspreading event.

Inside the market, I'm also not sure how far our discussion got

The lab leak theory has gone through many stages, criticizing the market evidence:

- ~~There were no animals at the market (after WHO report)~~
- ~~There were raccoon dogs in 2014, but what about 2019? (after eddie holmes photo)~~
- ~~There were raccoon dogs in 2019, but what about December 2019? (after xiao xiao paper)~~
- ~~There was raccoon dog DNA, but the human DNA was removed (Alina Chan)~~
- ~~There was raccoon dog DNA, but those samples were negative (Steven Quay)~~
- ~~There was raccoon dog DNA in positive samples, but it's not correlated to covid RNA (Jesse Bloom)~~
- There was raccoon dog DNA in positive samples, but not enough reads ← we are here, maybe?
- Even if the raccoon dogs were sick, it's still a lab leak

To reiterate, two shops selling wildlife tested positive: shop 6/29 and 8/25.
 One of these shops had 5 positive samples, including one from a cage.
 Both shops continued to test positive into February.



Many of these samples are both PCR+ and NGS+, so if you're worried about the low number of reads found in the NGS test, you can also trust the positive PCR tests.

Table 1. Information of the positive environmental samples from Huanan Seafood

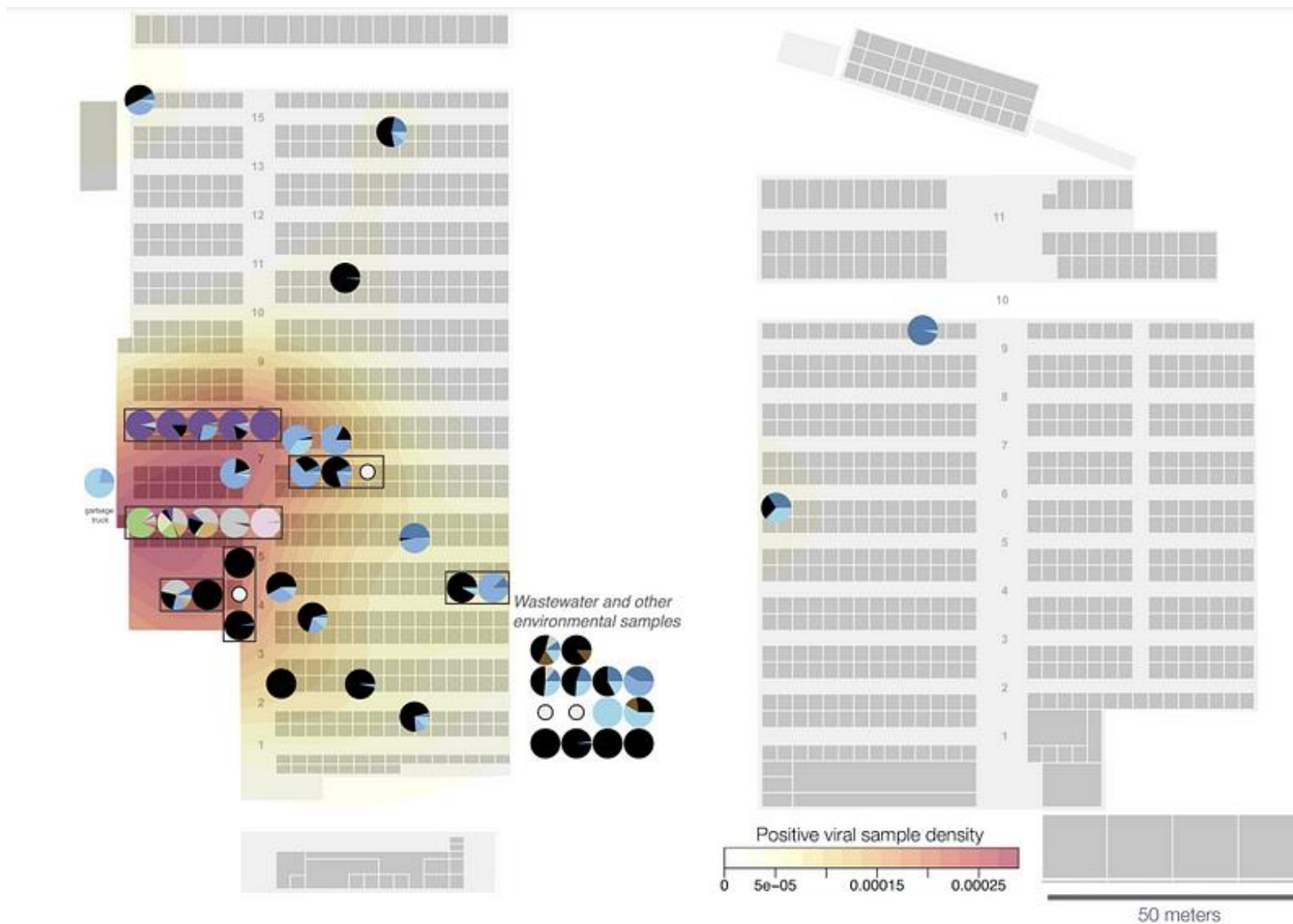
Market.

No.	Sample type	Sampling date	PCR	Ct	PCR target	NGS
E61	Ground	2020/1/1	+	36.04	ORF1ab/N	NA
A101	Surface of the door	2020/1/1	+	36.82	ORF1ab/N	NA
A14	Surface of packing bag for hairtail	2020/1/1	+	36.42	ORF1ab/N	NA
A15	Surface of the door	2020/1/1	+	35.51	ORF1ab/N	NA
A18	Shoe covers and soles	2020/1/1	+	33.79	ORF1ab/N	NA
A2	Ground	2020/1/1	+	35.07	ORF1ab/N	+
A20	Gloves	2020/1/1	+	32.48	ORF1ab/N	+
A33	Garbage truck	2020/1/1	+	34.46	ORF1ab/N	NA
A55	Ground	2020/1/1	+	34.84	ORF1ab/N	+
A61	Ground	2020/1/1	+	32.04	ORF1ab/N	NA
A63	Ground	2020/1/1	+	34.43	ORF1ab/N	+
A87	Surface of the door	2020/1/1	+	36.94	ORF1ab/N	NA
A88	Ground	2020/1/1	+	36.69	ORF1ab/N	NA
A90	Ground	2020/1/1	+	33.14	ORF1ab/N	NA
A96	Ground	2020/1/1	+	33.97	ORF1ab/N	NA
B17	Scale	2020/1/1	+	34.16	ORF1ab/N	NA
B5*	Ground	2020/1/1	+	29.32	ORF1ab/N	+
D32	Surface of a cart	2020/1/1	+	33.72	ORF1ab/N	NA
E48	Blood on the ground in front of the door	2020/1/1	+	35.93	ORF1ab/N	+
E7	Styrofoam desk in front of the door	2020/1/1	+	36.44	ORF1ab/N	+
F100	Ground	2020/1/1	+	34.72	ORF1ab/N	+
F13*	Surface of the wall	2020/1/1	+	23.85	ORF1ab/N	+
F33	Roller shutter	2020/1/1	+	34.13	ORF1ab/N	NA
F46	Ground	2020/1/1	+	31.8	ORF1ab/N	+
F54*	Ground	2020/1/1	+	25.8	ORF1ab/N	+
F98	Ground	2020/1/1	+	34	ORF1ab/N	+
G93	Sewage	2020/1/1	+	33.23	ORF1ab/N	NA
Q37	Inner surface of the freezer	2020/1/12	-	\	ORF1ab/N	+
Q61	Cart1	2020/1/12	-	\	ORF1ab/N	+
Q64	Cart2	2020/1/12	+	+	ORF1ab/N	+
Q68	Ground	2020/1/12	+	+	ORF1ab/N	+
Q69	Feather removal machine	2020/1/12	+	+	ORF1ab/N	+
Q70	Iron container in inner room	2020/1/12	-	\	ORF1ab/N	+

No.	Sample type	Sampling date	PCR	Ct	PCR target	NGS
1-27-33	Water drain	2020/1/27	+	36	ORF1ab	NA
1-27-37	Water drain	2020/1/27	+	35	ORF1ab	NA
1-29-4	Water drain	2020/1/29	+	36	ORF1ab	NA
1-29-8	Water drain	2020/1/29	+	37	ORF1ab	NA
8-25-D	Ground inside the stalls	2020/2/3	+	35.9	ORF1ab	+
8-25-M1	Ground inside the stalls	2020/2/3	+	36.5	ORF1ab	+
WS-1-1	West sewage well 1	2020/2/5	+	36.00	ORF1ab	NA
WS-1-2	West sewage well 2	2020/2/5	+	38.00	ORF1ab	NA
WS-1-3	West sewage well 3	2020/2/5	+	34.01	ORF1ab	NA
WS-1-4	West sewage well 4	2020/2/5	+	38.00	ORF1ab	NA
WS-2-1	West sewage well 1	2020/2/5	+	37.30	ORF1ab	NA
WS-2-2	West sewage well 2	2020/2/5	+	37.65	ORF1ab	NA
WS-2-3	West sewage well 2	2020/2/5	+	36.72	ORF1ab	NA
WS-3-2	West sewage well 3	2020/2/5	+	37.60	ORF1ab	NA
WS-4-2	West sewage well 4	2020/2/5	+	37.10	ORF1ab	NA
zong-1	Water drain	2020/2/9	+	34.94	ORF1ab	NA
w-6-29-33	Water drain	2020/2/15	+	37.97	ORF1ab	NA
w-zong-1	Water drain	2020/2/15	+	36.54	ORF1ab	NA
EWS-2#-2	East sewage well 2	2020/2/15	+	35.32	ORF1ab	NA
EWS-3#-2	West sewage well 3	2020/2/15	+	36.05	ORF1ab	NA
WWS-1#	West sewage well 1	2020/2/15	+	34.44	ORF1ab	NA
WWS-1#-2	West sewage well 1	2020/2/15	+	33.63	ORF1ab	NA
WWS-1#-3	West sewage well 1	2020/2/15	+	33.58	ORF1ab	NA
WWS-2#-2	West sewage well 2	2020/2/15	+	37.44	ORF1ab	NA
WWS-2#-3	West sewage well 2	2020/2/15	+	36.88	ORF1ab	NA
wws-1#-0	West sewage well 1	2020/2/15	+	33.75	ORF1ab	NA
W-8-25-D1	Ground inside the stalls	2020/2/15	+	36.77	ORF1ab	NA
W-8-25-D2	Ground inside the stalls	2020/2/15	+	33.91	ORF1ab	NA
W-8-25-L	Container	2020/2/15	+	34.58	ORF1ab	NA
W-8-25-L2	Container	2020/2/15	+	37.16	ORF1ab	NA
C8	Wall inside the stalls	2020/2/20	+	32.59-/37.34	RDRP/N/E	NA
SJ-D	Storehouse ground	2020/2/22	+	39.39/39.25/35.48	RDRP/N/E	NA
SJ-CS	Storehouse weight scale	2020/2/22	+	40.21/40.1/36.37	RDRP/N/E	NA
SJ-L3	Storehouse wire fence	2020/2/22	+	41.77/41.62/37.61	RDRP/N/E	NA
RLC-4	Storehouse bag surfaces	2020/2/22	+	41.71/41.53/37.31	RDRP/N/E	NA
RLC-3	Storehouse bag surfaces	2020/2/22	+	36.18/36.05/32.36	RDRP/N/E	NA

Note: Four positive samples were not included in this table. One sample was collected from other market in Wuhan and 3 collected from sewerage wells in surrounding areas.

Those infected wildlife shops had wildlife DNA, while many of the other positive market samples had human DNA.



Those two wildlife stalls both seem like plausible sources of the pandemic. Shop 6/29 looks better, to me. Shop 6/29 had the most positive samples, on the first (unbiased) days of testing the market.

The positive cage, carts, and especially the drains look like infected wildlife.

Shop 8/25 doesn't look quite as good because it tested negative on Jan 1st and mostly negative on Jan 12th. But then it did test positive repeatedly into February, so it's possible they figured out where to test in the shop, based on some knowledge they gained about which animals were important.

We don't know if either of those shop owners got sick. Neither of them were hospitalized for covid. That's not so weird – the odds of not being hospitalized for covid are perhaps 95%, on average (It depends on your age). Animal traders may even have some extra resistance on top of that.

It is still possible someone from the lab came straight to the market, infected the market, and the animals inside the raccoon dog shop got sick. Or perhaps someone sneezed on all the carts and cages and into the drains. But that's unlikely.

I'm claiming 1 in 68 odds that the internal market data looks like this. I say that because there are 10 wildlife shops in the market out of 680. Any shop in the market could have randomly had this pattern with lots of positive samples and heavily infected drains, but it just happened to be this one shop selling wildlife.

If Rootclaim wants to argue that this shop had infected animals because of a reverse zoonosis, and the odds of that are higher than 1 in 68, I'm willing to hear that argument and their estimate of those odds.

Here's relative hospitalization risk by age, from US [CDC data](#)

Mr Jin said his son in law is running shop 6/29. If the son is younger, in his 30's or 40's, then the odds that guy gets hospitalized by covid are lower than 5%.

Risk for COVID-19 Infection, Hospitalization, and Death By Age Group

Updated Nov. 22, 2021

Rate compared to 18-29 years old ¹	0-4 years old	5-17 years old	18-29 years old	30-39 years old	40-49 years old	50-64 years old	65-74 years old	75-84 years old	85+ years old
Cases ²	<1x	1x	Reference group	1x	1x	1x	1x	1x	1x
Hospitalization ³	<1x	<1x	Reference group	2x	2x	4x	5x	8x	10x
Death ⁴	<1x	<1x	Reference group	4x	10x	25x	65x	150x	370x

Both wildlife shops that tested positive were previously fined for selling illegal wildlife:

3 sellers at the Wuhan market were [fined in May, 2019](#) for selling illegal live hedgehogs (shops 6-29, 8-25, and 9-37)

The [fines were given on 5/29/2019](#) for Mr Jin, Mr Song, and Mr Zheng

Shop 6-29, with the raccoon dogs, was [owned by Mr Jin](#). (Jin Xiushan, 金秀山)

The shop name might be translated as Jiangnan Poultry Management Department.

Shop 8-25, owned by Mr Song Fabao, also had positive samples and positive warehouse samples.

Shop name is E'Dong Red Star Game Shop.

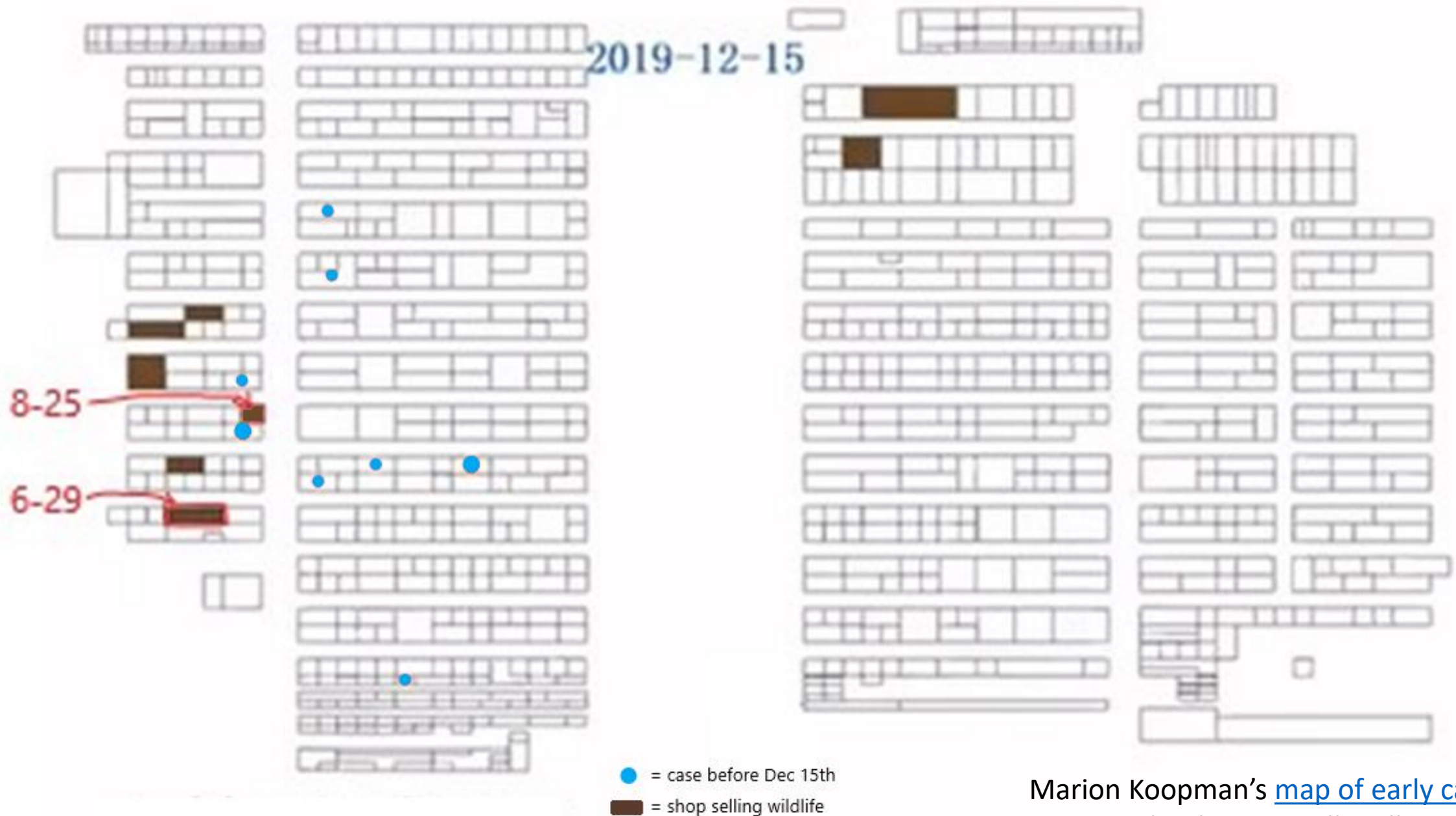
The fine notices [were taken down off the web](#) after Michael Standaert asked about them.

[No shops at the 3 other Wuhan wet markets](#) were fined in 2019 for illegal sales of wildlife.

I added a factor of 3/10 to my analysis to account for this.

This would be just another weird coincidence, if SARS-CoV-2 actually came from a lab and infected these shops, that it happened to infect the specific shops that were selling illegal wildlife.

Many of the earliest known covid cases at the market also worked close to these 2 suspicious shops:



Marion Koopman's [map of early cases](#), as compared to known stalls selling wild game.

These aren't just the earliest cases at the market, they're the earliest in town

Other than these, there's that one repeat visitor to the market that got sick on December 12th.

And there's one other non-market linked case who might be that early (they live very close to the market).

So it's really more like 12 of 13 of the first known cases are at the market, not just the first 5. And the 13th lived nearby.

The first case in Wuhan that's not at or near the market is Accountant Chen, on December 16th.

But Chen's case was sequenced and he's lineage B with no mutations, exactly the same strain seen at the market.

The lab leak theory we heard last week said that lineage A started at the lab and lineage B started at the market. Since Chen is lineage B, he must have been infected downstream of the market.

If Chen lived near the lab and had lineage A, or proCov2, that might be interesting lab leak evidence.

Instead, all it really shows is that the virus had started to spread around Wuhan by then.

And that shouldn't be too surprising, the market spillover was probably late November.

And covid can spread quickly – the first Beijing case was December 17th (and linked to the Huanan market).

China was obviously interested in shops 6/29 and 8/25.

They only tested warehouse samples for those 2 shops.

They only tested animals for a few shops, shop 6/29 was their highest priority.

Yes, the number of covid RNA reads in each wildlife sample is low.

But they're definitely positive, many of those samples are PCR+ as well, they're not false positives.

I think they're just weak positives because the market has been closed for 12 days, and the spill-over happened a month before that.

The market data is not perfect.

But it's pretty good.

And it's in comparison to the lab leak theory, which isn't even sure which lab covid leaked from or which month it leaked, and has no model of how many people were sick in December.

Just imagine if this were reversed, and the evidence for the lab was this good

- Imagine that the 12 earliest cases were at the lab.
- Imagine they were all fairly near Ben Hu's office.
- Imagine we didn't know for sure if Ben Hu was sick, but we did get 5 weak positive samples inside his office.
- Imagine that all the December cases radiated out away from the lab and were centered on it.
- Imagine those lab cases were also at the center of the pandemic, genetically.

And so on...

Imagine that was the evidence, and someone tried to argue the virus was actually from a market across town.

Suppose they did some analysis finding that many people at the lab got infected in the bathrooms or some break room where people socialized. Then they tried to argue that the toilets were the cause or the break room just has some property that grows covid infections faster than any other place in Wuhan, and that's the answer, not the research done at the lab.

Suppose they did some correlation analysis saying that the covid RNA was too spread around the lab, that the highest amount wasn't at Ben Hu's office, so that couldn't be the source.

That would sound absurd. But that's the mirror image of the lab leak theory.

I think this shows that people have a huge bias towards believing lab leak theories, for all the reasons I explained in the first debate. People have an innate bias to see a human motive behind every natural event that happens in the world.

How does the work done at the lab, or the DEFUSE grant, change these probabilities?

In theory, I could still support a lab leak at this point.

Suppose that what showed up at the market was the SARS virus from 2003, with zero mutations. That would be impossible in nature, but possible from a lab freezer.

That would be a bigger coincidence than any of the market evidence.

The same would be true if it was some chimeric virus with an exact WIV16 backbone. It would also be impossible to find the exact previous backbone in nature with no mutations, years after that virus was found.

But what we actually have is a novel virus that's only 80% similar to SARS, and it has many features that do not look engineered.

It's not at all clear that a lab made that virus, or whether they could make that virus.

We need to mark down the lab leak theory for the odds that they could create this virus.

The lab would first need a secret virus

Rootclaim says they had 180 secret viruses. It's clear that was actually 180 total viruses in the DEFUSE grant, because they said those were collected over 14 years, and the number matches the total number of viruses they had, as measured by other papers.

I think they have no secret viruses.

But if you make some assumption they secretly did all the work proposed in DEFUSE, maybe they have collected ~30 extra ones.

Odds of a secret starting virus?

DEFUSE grant says they'll collect 3,000 samples (page 31). That's going to yield ~30 sarbecoviruses.

The odds of finding one like BANAL-52 are ~1 in 100 viruses, if you look at exactly the right location in Laos.

But they're much lower elsewhere. The WIV plans to sample in their known bat cave. They've already sampled there, 100-200 times, without finding a SARS2 family virus. Looking for 30 more viruses there isn't likely to get one.

Full inventory of bat SARSr-CoV QS at our test cave sites, Yunnan, China. To provide data to train and validate our modeling, and as baseline for our immune modulation trial (TA2), DEFUSE fieldwork will target the high-risk cave site in Yunnan Province, SW China (Fig. 4, red triangle) where we will conduct our field trial, and where we have previously identified and isolated high-risk SARSr-CoVs^{2,11,33,34}. At three cave sites (one designated for our trial, two as controls), we will determine the baseline QS₀ risk of SARSr-CoV spillover. We will conduct longitudinal surveillance of bat populations to detect and isolate SARSr-CoVs, determine changes in viral prevalence over time, and measure bat population demographics and movement, definitively characterizing their SARSr-CoV host-viral dynamics. Field data will allow us to test the accuracy

They would need a virus with the correct Receptor binding domain.

Maybe 5 sarbecoviruses have that (3 BANAL viruses, 1 from Yunnan, 1 from vietnam) out of [1,500 total sarbecoviruses](#).
 $30 \text{ viruses} * (5/1,500) = 10\%$.

But 4 of those aren't close to SARS2, outside the RBD, you really need something like BANAL-52:
 $30 * (1/1,500) = 2\%$

2% is my high end guess for the odds they have a secret virus they could use to make SARS2.

I think it's lower than that. BANAL-52 is still only 97% similar to SARS2, it really needs to be 99+%

You need to adjust downward for that. I'm not sure by how much.

And then I think you should adjust downwards again for the fact that all these FOIA attempts and uncovered papers don't have any relevant viruses, and no evidence has shown up of secret sampling trips. Adjust downwards for that.

I went with 1 in 1,000 odds of a secret virus.

But it's hard to know the right factor to add for secret research. Maybe we should separate these into separate factors:

- At most 2% chance they have the right virus, if they did secret sampling trips.
- Multiply by some other factor for the odds they did those secret trips and hid them.

They would also need a new reverse genetics system, for that new backbone

The DEFUSE grants says that they will primarily work with 3 backbones: WIV1, WIV16, and SHC014

They also talk about making 3-5 novel backbones per year, but it's unclear how they will choose these.

They're definitely not going to make a new backbone for every virus. That takes too long and isn't necessary for most of what the lab wants to know.

In a discussion between Yuri Deigin and Stuart Neil, they agreed it might take about [3 months to make this](#).

In the last debate, Yuri said maybe you could do that in 3 weeks, if you were efficient.

Even then, 200 viruses * 3 weeks = 600 weeks. Or you need a large team to do the work in parallel.

Shi Zhengli does not have a large team, so there's a limit on how many they could work on at once.

They would need to find that SARS2 precursor interesting

The lab had RATG-13 for 6 years. That's 96% similar to Covid and they did not do any experiments with that.

DEFUSE talks about inserting viruses into 3 backbones highly similar to SARS (WIV1, WIV16, and SHC014). They were definitely interested in those viruses.

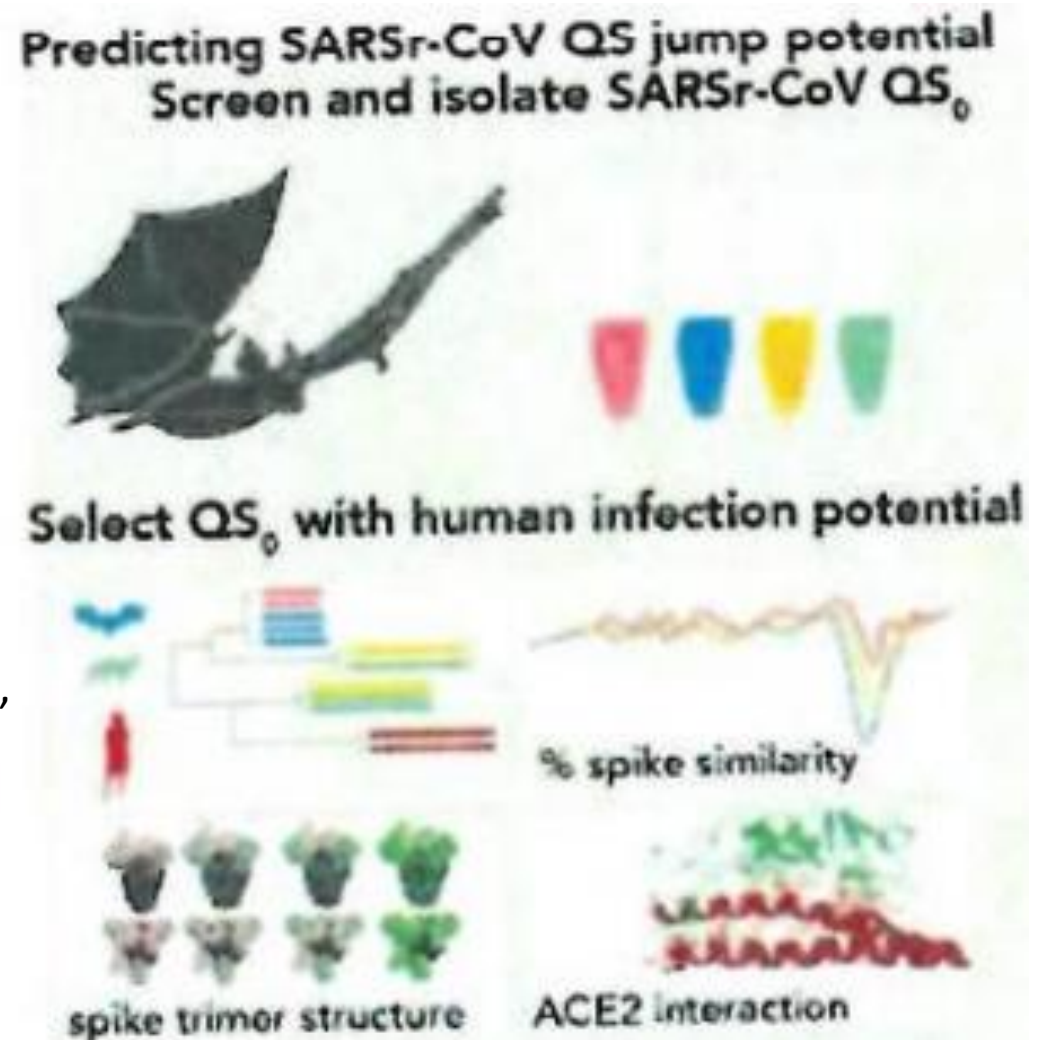
DEFUSE was interested in ACE2 interaction, but they were also interested in spike similarity (presumably to SARS)

The SARS2 spike is not uniquely well adapted to human cells.

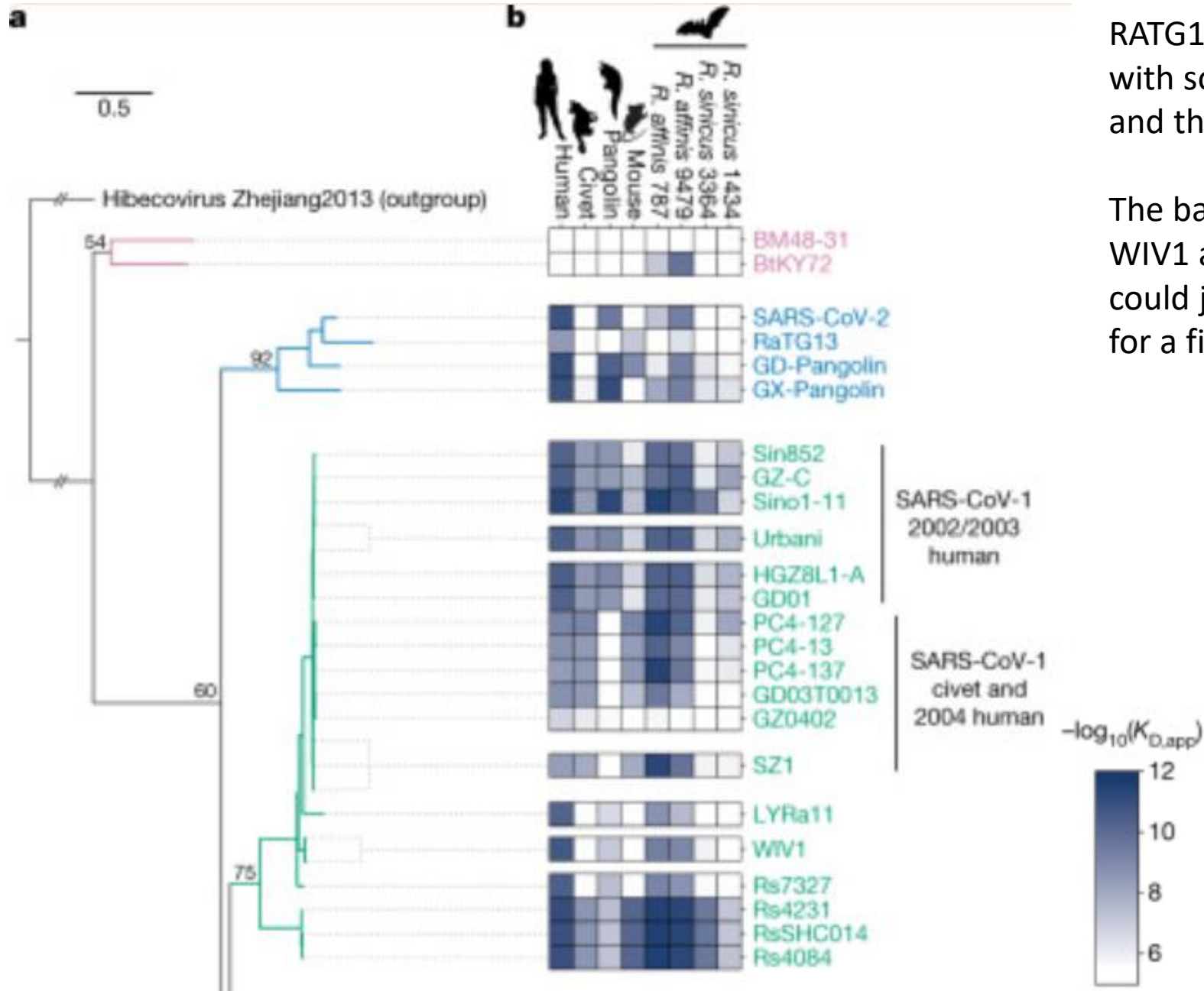
It works well, but [some pangolin viruses work even better](#).

And SARS2 has had mutations throughout the pandemic to improve cell entry.

I'm not sure it would stand out compared to other sarbecoviruses, prior to the pandemic.



SARS2 binds well to human ACE2, but so do many other bat viruses, most of them closer to SARS1



RATG13 doesn't bind as well, but it's on par with some of the less well adapted SARS1 strains, and the WIV1 seems to have ignored SARS1.

The backbones normally used at the lab, like WIV1 and SHC014, bind well to human ACE2. You could just throw a furin cleavage site into those for a first experiment and that might work.

Figure 1 from [Starr et al, 2022](#).
Darker colors indicate better ACE2 binding.

Even among the most closely related viruses, SARS-CoV-2 does not have the best human ACE2 binding.

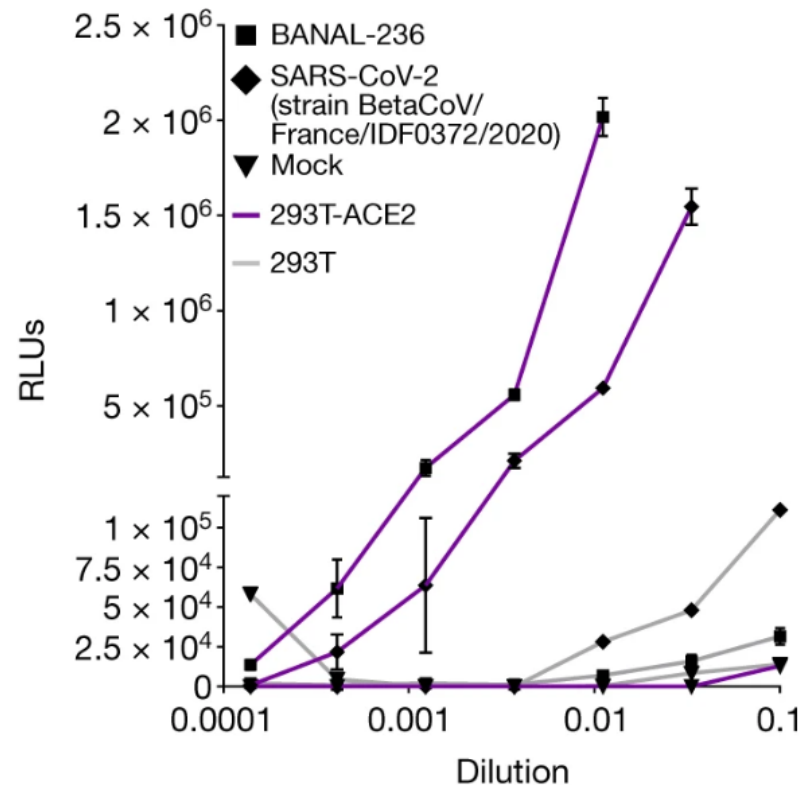
BANAL-236 binds better to human ACE2 than SARS-CoV-2 does.

If the lab had secretly gone to Laos and collected some of these same viruses, there's no guarantee they would have focused specifically on a SARS-CoV-2 precursor.

Fig. 4: BANAL-236 entry and propagation in human cells.

From: [Bat coronaviruses related to SARS-CoV-2 and infectious for human cells](#)

a



a, Results of spike-pseudotyped BANAL-236 (squares) and SARS-CoV-2 (strain BetaCoV/France/IDF0372/2020, GISAID accession number EPI_ISL_406596, diamonds) pseudovirus entry assay in HEK-293T cells expressing (purple lines) or not (grey lines) the hACE2 receptor, shown in relative luminescence units (RLUs) produced by the firefly luciferase present in the lentiviral backbone and the Bright-Glo luciferase substrate. A single experiment performed in triplicate representative of two experiments is shown. Centre values represent the average of the three replicates and error bars indicate s.d.

Figure from [Temmam et al 2022](#)

They would need to make a full-length backbone for that secret virus.

Let's assume that what they're doing is likely to leak, because you think the lab is highly unsafe.

So, first they're going to create 180 chimeras in a WIV1/WIV16/SHC014 backbone, to categorize the spikes of all the viruses they already have.

Then they're going to make up to 30 more, if they find 30 new viruses.

Then they're possibly going to make 3-5 full length viruses per year, but it's hard to understand which ones and what they would prioritize.

That's ~210 chances for a virus to leak, before you even get to making the full length viruses.

The first virus that leaks is going to only have (3-5)/210 odds that it's one of the full length recreated ones (1-2% range).

Existing spikes like WIV1 and SHC014 might work if you added to cause an outbreak, if you added the FCS to those.

Even if the SARS2 spike is some key thing that enables it to leak and spread, that spike might still work in one of the three existing backbones, and those chimeric viruses could leak first.

They would need to add a furin cleavage site

There's a [difference of opinion](#) as to whether DEFUSE was going to add cleavage sites where they don't exist, or look for partial sequences and complete them, or what exactly was happening .

You'd probably want to add some probability here, because it's unclear.

I didn't, I just called it 100% to steelman the lab leak theory.

I did add a 50% to my analysis, because it's not clear if they're talking about the S1/S2 site or S2'.

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 proteases⁶⁸⁻⁷¹ cleave the SARS-CoV S glycoprotein causing massive changes in S structure⁷² 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-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. We will also review deep sequence data for low abundant high risk SARSr-CoV that encode functional proteolytic cleavage sites, and if so, introduce these changes into the appropriate high abundant, low risk parental strain. N-linked glycosylation: Some glycosylation events regulate SARS-CoV particle.

They would need to add a very weird cleavage site

Scientists normally choose RRKR. The cleavage site found is PRRAR, and it's out of frame.

Out of 5 previous experiments I found adding cleavage sites, neither used Proline or Alanine.

I'm saying it's a factor of 10 for the Alanine. That's suboptimal but not seriously detrimental.

I added a factor of 20 for the proline because you'd expect that to be detrimental.

And I added a factor of 6 for "out of frame" – with 6 possible codons for serine, maybe the secret virus had the right one.

Suppose you also think, like Yuri did, that they put an optimal FCS in first (RRKR), then tried different ones, or maybe they tried lots of different combinations.

In that case, the RRKR one is still likely to leak first, before they get to something weird like PRRAR.

They would need to do this with live viruses, not pseudoviruses.

Many experiments can be conducted with pseudoviruses, with less effort and higher safety.

The prior furin cleavage site insertion experiments I found used pseudoviruses.

But the DEFUSE grant has references to both live and pseudoviruses, so it's hard to say what they were going to do here. Maybe it could be either.

I'm only adding a factor of 2 here, to steelman the lab leak theory.

They would need to successfully culture the virus, with no mutations.

Vero cells don't work. They lose the furin cleavage site.

Calu-3 cells don't work, various mutations are seen.

HAE cells or Vero cells+trypsin might work, but only if the virus escapes the culture quickly.

Humanized mice probably don't work, because of the N501Y mutation, which arises in [multiple experiments](#). But they could work if the mouse bites a person among the first few passages from mouse to mouse.

I called it 1 in 10 that the lab would figure all this stuff out and culture the virus properly (i.e. they didn't even use HAE cells at the WIV in previous experiments).

But this one is a hard number to say precisely.

For instance, I didn't calculate the odds that the first passaged mouse would bite a lab worker as compared to the third passaged mouse.

The virus would have to leak from the lab.

Lab infections are normally about 1 in 500, per lab, per year.

Rootclaim says this should be higher, because the lab was unsafe (some experiments at BSL-2) and they created a highly pathogenic virus like SARS-CoV-2.

I conservatively upgraded my analysis to say 1 in 50 (= 2%) chance of a leak, because of this.

Rootclaim says it's a 15% chance of a leak.

That sounds too high, to me, but I suppose this is a hard number to know.

One resource I found to estimate this is [a database of lab accidents](#). We can search this to see how many times SARS-CoV-2 has leaked, since the pandemic has started. The database lists 3 incidents but they're all from Taiwan and it's unclear, I think they might be referring to the same incident.

That accident appear to be from working with Covid infected mice, and it was not done in a biosafety cabinet.

When we say BSL-2, they are still working with negative air flow cabinets, that should be the primary line of defense, not the mask the person is wearing:

Biosafety Cabinets



Class I



Class II



Class III

Also, it's unclear how much virus you'd have to grow in cell culture before a person could easily get infected.

Since SARS-CoV-2 is not leaking from many labs today, despite lots of work being done on the virus, maybe we can conclude that the rate of initial lab leaks would also be low, and Rootclaim's 15% estimate is too high.

But it's also possible that the WIV had much worse safety than all labs working with the virus today. And it's also possible that SARS-CoV-2 has leaked from other labs since 2020 and it just wasn't reported.

Putting this all together:

Odds DEFUSE grant happened secretly 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** – my guess based on Latinne FOIA, 2018 paper, sampling rates, etc.)
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?** could be lower/higher, hard to guess 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 a lab leak = 1 in 5×10^{25}

Putting this all together:

Odds DEFUSE grant happened secretly 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 – my guess based on Latinne FOIA, 2018 paper, sampling rates, etc.)

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? could be lower/higher, hard to guess 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 a lab leak = 1 in 5×10^{25}

Uses real data

less reliable but can be estimated from other data

Hard number to guess

This isn't a full Bayesian analysis with priors.

I have suggested starting at 99 to 1 for zoonosis vs lab, based on flu data, and the conservative (but unproven) assumption that 1977 was caused by some lab.

I would mark down zoonosis for the Furin Cleavage Site found in Wuhan.

CGG might lean weakly towards the lab, but it's not a high bayes factor and I'm not sure it counts at all, since it seems to have a function that was selected for.

I would give no weight to N-glycans or the Faul site.

I would not count ACE2 binding, since we've now found that feature in nature.

Suppose that you start at 50/50 after including both Wuhan location and the FCS, to match Kristian Andersen's priors. That's equal to starting at 99 to 1 for zoonosis and then giving Wuhan+FCS a weight of 100. Start there, for simplicity.



Kristian Andersen 11:47

February 2nd, 2020 ▾

Reading through Ron's comments again I agree on pretty much everything he's saying - I come to the same conclusions. Where we differ is that he's looking for very specific evidence proving that this is unnatural (which is understandable), but except for the most simple scenario where somebody plugged a gene into a preexisting backbone, that would simply be impossible to prove.

Natural selection and accidental release are both plausible scenarios explaining the data - and *a priori* should be equally weighed as possible explanations. The presence of furin *a posteriori* moves me slightly more towards accidental release, but it's well above my paygrade to call the shots on a final conclusion.

After adding all these new things we've learned since February 2020, zoonosis wins by a factor of $5 \cdot 10^{25}$

That leaves a wide margin for error, in terms of interpreting the market evidence and exactly what the DEFUSE grant means.

There is still a fair amount of uncertainty in this analysis.

We can debate the details more.

Some values are based on real data.

Some are estimates based on the number of prior experiments.

Some are guesses that should be treated with a wider range of uncertainty.

I may be double counting in a few places, that's a common risk in Bayesian analysis.

I said it's 1 in 10 they'd find the SARS2 spike interesting and 1 in 100 they'd make a SARS2 backbone.

Those might be dependent on each other -- the odds of a SARS2 precursor becoming a backbone might go up if they find the spike interesting and thus be double counted. But maybe not, because WIV1+SARS2 spike might be interesting enough, and that could also leak.

I said they didn't publish their work but also counted them having no secret virus. That might be double counting.

I also counted the fact that governments/sleuths haven't solved this. That overlaps a bit with "no secret virus".

But it doesn't overlap entirely, because there are so many different methods that intelligence agencies and sleuths could use to solve this mystery.

If I include a factor for the lab leak cover-up being so good that no one has proved it, do I need to include a similar factor, on zoonosis side?

Maybe. Neither one has been definitively proven, otherwise we wouldn't be debating this. But there's a lot more evidence pointing to the market than the lab.

I could consider removing this factor from my analysis, but my opponent relies heavily on the fact that no animal tested positive, so I think he would likewise have to remove that.

I also think the dispositive evidence for a natural origin could only be found by surveying Chinese wildlife, and no one in China has done that or published it.

None of the live wildlife was tested at the market and only 600 animals were tested from supplying farms, most of which were not even relevant species:

Table 5.1. Survey of animals from Huanan market suppliers in Hubei

Nucleic Acid Testing (NAT)	
Hubei	
Number of species	10
Specific types of animals	Bamboo Rat, Porcupine, Duck, Snake, Rabbit/Hare, Chicken, Ostrich/Turkey, Wild Boar
Total sample size	616
Test results	Negative

The dispositive evidence for a lab leak could be found in a variety of ways, maybe even outside China

There could be a whistleblower from the lab, or any foreign scientist who worked with the lab.

There could be a saved copy of the Wuhan database. Or a FOIA'ed copy of a paper. Or some e-mails.

There could be evidence from the supplies that the Wuhan lab purchased.

There could be witnesses if the Wuhan lab had secret sampling trips in Yunnan or Laos.

Years of searching have not turned up anything.

The best evidence they've found is the DEFUSE grant, which is ambiguous and probably didn't even happen (even my opponent gives it less than 50% odds).

The human evidence we do have points in the other direction.

Australian scientist Danielle Andersen worked at the WIV in November. She didn't get sick, and didn't know anyone that got sick. She tested negative for covid antibodies.

The Wuhan lab says they tested their staff for covid antibodies and all were negative (obviously, this is the one thing that would be easiest to lie about).

We do have something of a whistleblower from the Trump state department, [Christopher Ford](#). He says that he pushed his colleagues to investigate the science more carefully before accusing China of a lab leak. He says there was an internal review at the state department of Steven Quay's Bayesian analysis, and the scientific experts gathered said the analysis did not hold up. Ford says that his colleagues, like David Asher, went far beyond what the science could prove.

We've also [uncovered e-mails from the people supporting Brexit](#), where they plotted to get Yuri Deigin onto Bret Weinstein's show and [Weinstein onto Joe Rogan](#), to popularize the lab leak theory.

And we may even have some evidence that Robert Redfield might have been lying to congress, when he said he tried to alert Fauci and Collins about a lab leak, before February 1st. [Multiple FOIA attempts](#) have failed to turn up any such e-mails.

There are many ways to answer a question, probabilistic arguments are only one.

One way to resolve an argument is to lay out all the evidence.

The facts point to a natural origin at the Huanan market, the lab leak theory is entirely speculation:

Fact:

Earliest known cases were at the Huanan market.

Later cases center on the market.

There were Covid susceptible animals at the market.

There is no known virus which could be used to create Covid.

RATG-13 is not close enough to create Covid.

The DEFUSE grant was not funded.

That project was supposed to happen in America.

The DEFUSE grant was interested in SARS1 viruses and chimeras.

Scientists usually insert RRKR or RRSRR as a furin cleavage site.

Scientists have no clear codon preference for arginine.

The Wuhan database went down at some point in 2020.

Kristian Andersen initially thought covid might be a lab leak

Speculation:

Maybe China missed earlier cases or hid them.

Maybe China used a biased search.

Maybe a person from the lab came and infected those animals.

Maybe the WIV secretly had other viruses.

Maybe RATG-13 was modified, or it's fake, or it's a diversion.

Maybe the project secretly happened anyways.

Maybe work secretly happened in Wuhan.

Maybe Shi Zhengli was interested in SARS2 viruses.

Maybe Shi Zhengli had some reason to choose PRRAR.

Maybe Shi Zhengli always uses CGG.

Maybe the database went down in September 2019, to hide a lab leak.

Maybe Fauci forced Andersen to change his mind

Another approach is to look at who's spreading the information.

The lab leak theory was started by Steve Bannon, one month after the pandemic started.

The 3 sick WIV employees story was promoted by Michael Gordon, who also spread the stories about Iraq's weapons of mass destruction.

There are big overlaps with other misinformation movements.

Alex Washburne works with the Brownstone institute, an organization that also promotes anti-vax authors.

Matt Ridley supports the lab leak theory and also doubts climate change.

Matt Ridley supports lab leak and also doubts climate change

In one article, he writes:

Matt Ridley: Cooling Down the Fears of Climate Change

Evidence points to a further rise of just 1°C by 2100. The net effect on the planet may actually be beneficial.

Most forecasts point to substantially more global warming than that.

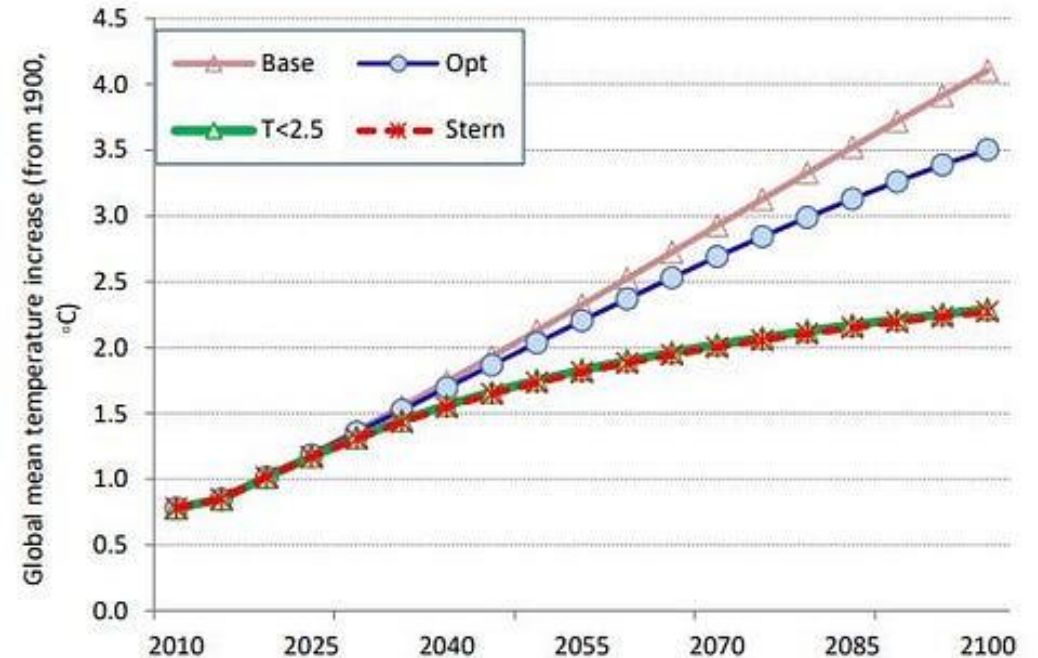


Figure 4. Temperature change in different scenarios
The most ambitious scenarios cannot limit temperature to 2 ½ °C, and the cost-benefit optimum with standard parameters has sharply rising temperatures.ⁱⁱⁱ

[one forecast](#) from Nobel economist William Nordhaus

Matt Ridley uses the same tactics for lab leak and climate change denial

In 2010, he [praised bloggers for questioning climate change](#) and said that climate researchers were corrupted by research grants. That sounds just like praising DRASTIC and accusing Kristian Anderson and other Covid researchers of being corrupted by funding from Fauci.

Ridley framed climate change as a “free speech” issue:

THE MANY ATTEMPTS TO STIFLE FREE SPEECH ON CLIMATE CHANGE - MATT RIDLEY

[Home](#) > [Blog](#) > The many attempts to stifle free speech on climate change

Published on: May 2, 2016

There's room for disagreement within the “consensus”
My [Times column](#) on free speech and climate change:

That sounds just like framing lab leak as “important to talk about because it is censored”.

Lab leak was censored by Facebook in 2020, but mostly just stuff like “Covid is a bioweapon”.

Ridley might question climate change because he [owns and profits from coal mines on the property he inherited](#).

Ridley is also known for helping crash the British economy during the global financial crisis. He was the chairman of Northern Rock bank, when it had to be bailed out in 2007.

Yuri's co-author, Daoyu Zhang, is "Flavinkins" on Gab. Here are some of the things he posts:

Flavinkins reposted

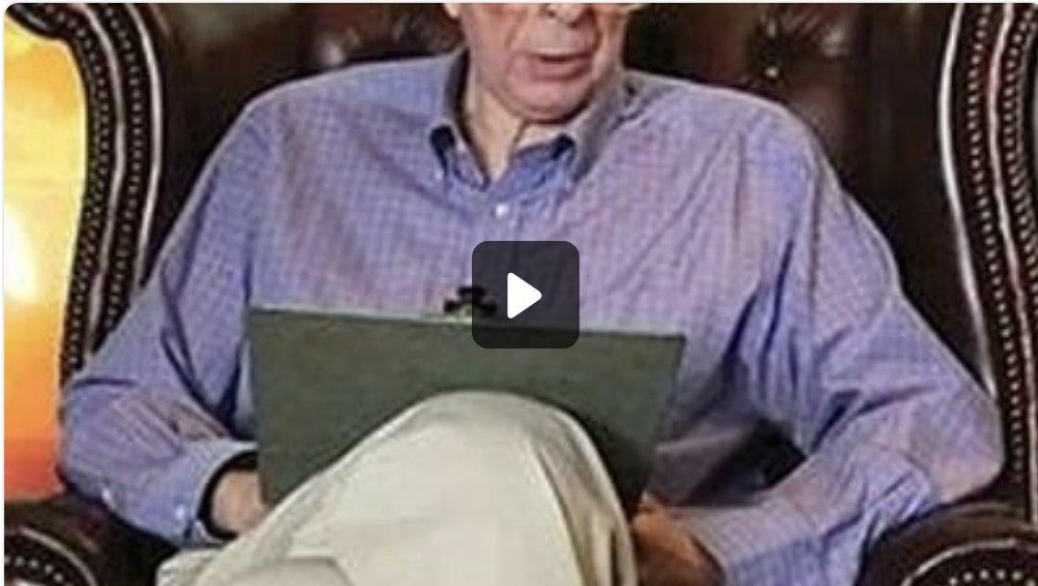


BardsFM @BardsFM

Mar 14 · 🌐

Weapons of Mass Destruction #VaxWMD 🚫💉

soundcloud.com/user-222860353/covid-19-vaccin...



Covid-19 Vaccines Are Weapons of Mass Destruction - and Could Wipe out the Human Race

Covid-19 Vaccines Are Weapons of Mass Destruction - and Could Wipe out the Human Race

Dr Coleman says this is the most crucial video he's ever made about covid-19....

soundcloud.com

View Link Feed

Flavinkins reposted



FA355 Pureblood @FA355

Apr 3 · 🌐

Quotable quote: "Fabricating data is easy. Scientists must start to treat it with skepticism."

Flavinkins @Flavinkins

Apr 3 · 🌐

dogsbreakfast.substack.com/p/sars-evidence-of...



SARS-1: Evidence of an Artificial Origin

As the world debates the origin of SARS-COV-2, most assume the SARS outbreak of 2003 was a natural event. But revisiting the evidence I found parallels, direct linkages...

dogsbreakfast.substack.com

Link Feed

Flavinkins reposted a comment on a post



Flavinkins @Flavinkins

May 1 · 🌐

@CapitolSheila @jhas5 @Nih @nypost georgewebb.substack.com/p/was-corona-911-real...



University of California Pre...
The Dark Winter's Sun: A...



Wikipedia
R. James Woolsey Jr. - W

Was Corona 9/11 Really Anthrax 9/11 Live?

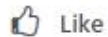
Was Operation Dark Winter Moved From A Table Top Exercise To The Real World For Corona 9/11?

georgewebb.substack.com

Link Feed

👍 2

1 repost



Like



Comment



Repost



Quote



Share

Flavinkins reposted a comment on a post



Nehming Names @ShemNehm

May 29 · 📱 · Edited

@Flavinkins The French started rescheduling HCQ in Oct 2019, and it was made official in Jan 2020.

The Sabotage of Hydroxychloroquine

gab.com/ShemNehm/posts/106851904025280183



Flavinkins @Flavinkins

Feb 22 · 🌐

hsgac.senate.gov/media/majority-media/johnson...

Likely that ballots for trump is simply stopped midway and destroyed, and counterfeits were created.

thegatewaypundit.com/2020/12/china-whistleblo...

projectveritas.com/news/breaking-new-pennsylv...



Majority Media |

Homeland Security & Governmental Affairs Committee |

Homeland Security & Governmental Affairs Committee

🔗 w...

View Link Feed

6 likes 2 reposts



Like



Comment



Repost



Quote



Share



What's on your mind?

There are a few reasonable people creating lab leak theories.

There are also some reasonable scientists that aren't involved in creating specific lab leak theories, but who just remain on the fence about the origins issue, or who just think that lab safety is an important issue that should be talked about.

But there are also a large number of bad actors creating misinformation.

For Rootclaim's Bayesian analysis to be valid, it needs to be very careful to exclude that misinformation, otherwise the analysis will be flawed ("garbage in = garbage out")

Another approach is to look for patterns common in conspiracy theories.

Goalpost shifting is a typical one.

Conspiracy theories can't be killed by contradictory facts, they just change into new versions.

Think about hydroxychloroquine. It wasn't enough to do a trial showing that hydroxychloroquine [doesn't work against covid](#). People said it only works if you treat the patients earlier. When [that failed](#), they said you have to add azithromycin. When [that failed](#), they said you have to add zinc. When [that failed](#), they still didn't give up, they just said that the trials were all rigged.



Lab leak follows a similar pattern

At first, covid binds well to human ACE2, so it must be engineered.

Then we find a pangolin virus which also binds well. So the theory becomes that covid is a lab chimera between bat and pangolin viruses.

Then we find a bat virus which binds even better. So the theory becomes that the lab secretly had that virus.

The scientists proposing a natural origin make theories that fit evidence, and their theories often predict new developments:

- they predicted lineage A would be found at the market, and it was

- they predicted raccoon dogs were at the market, and they were

You can't make predictions with the lab leak theory.

The only prediction is: no matter what evidence is found, they'll say it's good news for the lab leak theory.

Goalpost shifting means that this debate will never be settled:



Alex Washburne

@WashburneAlex



Replying to [@MJnanostretch](#)

I'm concerned that there could be an effort by those likely responsible for creating SARS-CoV-2 to plant a false ancestor in bat caves, then emerge and say "See? Natural origin!"

It's horrific that EHA received million\$ to continue this research given such an existential COI.

11:27 AM · Dec 13, 2022

What evidence would be good enough?

Washburne says that finding a bat ancestor to SARS-CoV-2 wouldn't be good enough, he thinks it would be fake.

Suppose you could prove that the raccoon dogs were infected on December 31st, through some more careful analysis of the market DNA samples, or maybe some uncovered frozen sample.

The theory would be that someone from the lab infected those animals back in November. Rootclaim already proposed that.

Or they would say that the evidence was fake.

Suppose you found a natural raccoon dog virus that was a likely precursor to covid.

The theory could be that the lab secretly had that virus, just like it secretly had the right bat virus from Laos.
Or they could say that virus was fake.

Suppose the Wuhan lab made 100% of their records and databases publicly accessible and provided clearer alibis showing that key scientists did not get sick in November.

The theory would be that the Wuhan lab is lying and hiding additional viruses and databases.

Another approach is to look at behavior

If you think there was a lab leak, did people respond like it was a lab leak?

For instance, Shi Zhengli went travelling to a conference in Singapore, in December, right after a November lab leak would have occurred. Her lab kept publishing, throughout the fall.

Her team went out to dinner without concern in January.

She published RATG-13, a virus that was so similar to Covid that it created a year of conspiracy theories.

She didn't even have to publish that virus, if they'd used it or a similar one for engineering.

Shi Zhengli also gave an interview where [she said she was worried the virus could have come from her lab](#). And she said she had previously thought of Wuhan as an unlikely place for a pandemic to start.

Why would she say things like that, if she had created the virus? Wouldn't that just implicate her? Wouldn't she say the opposite to cover it up?

Shi, a virologist who is often called China's "bat woman" by her colleagues because of her virus-hunting expeditions in bat caves over the past 16 years, walked out of the conference she was attending in Shanghai and hopped on the next train back to Wuhan. "I wondered if [the municipal health authority] got it wrong," she says. "I had never expected this kind of thing to happen in Wuhan, in central China." Her studies had shown that the southern, subtropical provinces of Guangdong, Guangxi and Yunnan have the greatest risk of coronaviruses jumping to humans from animals—particularly bats, a known reservoir. If coronaviruses were the culprit, she remembers thinking, "Could they have come from our lab?"

Some lab leak conspiracies also include:

The Chinese government

Peter Daszak and Ecohealth Alliance

The WHO team that worked with Peter Daszak

Ralph Baric

Kristian Anderson

Eddie Holmes

Tony Fauci

Michael Worobey and all his co-authors

Maybe I'll make the list, after this debate?

Which of these people acted like they were in on the plot?

Peter Daszak

Dec 9: TWIV interview, talks about swapping out spike proteins

Jan 11: He's still tweeting about infecting humanized mice with SARS related viruses.

Jan 25: He's still talking about "50 SARS-CoV viruses" when asked if the virus is natural.

Feb 10: sends out [an e-mail](#) asking other Ecohealth scientists to sign a letter condemning conspiracy theories.

Feb 19: The Lancet puts out [that letter](#), signed by 27 scientists

Tony Fauci

February 1: when informed of a possible lab leak, his first response is to suggest someone call the FBI:

From: Fauci, Anthony (NIH/NIAID) [E]
Sent: Sat, 1 Feb 2020 00:38:35 +0000
To: Jeremy Farrar
Cc: Kristian G. Andersen
Bcc: Conrad, Patricia (NIH/NIAID) [E];Mascola, John (NIH/VRC) [E];Conrad, Patricia (NIH/NIAID) [E]
Subject: RE: Phone call

Jeremy:

I just got off the phone with Kristian Anderson and he related to me his concern about the Furine site mutation in the spike protein of the currently circulating 2019-nCoV. I told him that as soon as possible he and Eddie Holmes should get a group of evolutionary biologists together to examine carefully the data to determine if his concerns are validated. He should do this very quickly and if everyone agrees with this concern, they should report it to the appropriate authorities. I would imagine that in the USA this would be the FBI and in the UK it would be MI5. It would be important to quickly get confirmation of the cause of his concern by experts in the field of coronaviruses and evolutionary biology. In the meantime, I will alert my US. Government official colleagues of my conversation with you and Kristian and determine what further investigation they recommend. Let us stay in touch.

Best regards,

Tony

Anthony S. Fauci, MD
Director
National Institute of Allergy and Infectious Diseases

Wuhan and Beijing governments

Dec 10: first known case

Dec 27: Doctor Zhang Jixian discovers human to human transmission.

Dec 29: Market link is known across several hospitals.

Dec 31: China closes and cleans the market.

Jan 15: WIV scientists are still out having dinner in public, not wearing masks.

Jan 18: Wuhan [potluck dinner for 40,000 families](#), designed to break some record.

Jan 20: China [admits human to human transmission](#).

Jan 21: Chinese new year celebration

Jan 22: order to wear masks in Wuhan

Jan 22: Chinese CDC: “it is highly suspected that the current epidemic is related to the trade of wild animals”

Jan 23: China quarantines Wuhan and 3 other cities

Jan 23: China [shuts down Hubei wildlife farms](#)

Jan 27: Wuhan mayor steps down, publicly shamed

Feb 24: Beijing [restricts wildlife farming nationwide](#).

The Wuhan government doesn't seem to know about a lab leak. They didn't panic or respond aggressively.

Killing tens of millions of animals on farms could be a way to hide a lab leak, but that sounds quite aggressive.

Kristian Andersen

Early on, he says he's 50/50 for natural/lab origin.


He gets less certain as they find the pangolin viruses, notices the furin cleavage site doesn't look engineered, finds the PAA insert in RmYN02. And presumably he leans even more towards a natural origin when he sees clearer data showing the virus came from the Huanan market.

By the end of the slack messages, he declares his frustration with all the misinformation, and says he's ready to fight back against it. He says he will "die on this hill":



Kristian Andersen 14:32

So much bullshit again. I have decided that I *am* going to die on this hill, so I'll talk to a few reporters and try to beat some sense into them. NYT had an article earlier today (I talked to them a couple of weeks back): <https://www.nytimes.com/2020/04/30/us/politics/trump-administration-intelligence-coronavirus-china.html>

 **The New York Times** | By Mark Mazzetti, Julian E. Barnes, Edward Wong and Adam Goldman
Trump Officials Are Said to Press Spies to Link Virus and Wuhan Labs
Some analysts are worried that the pressure from senior officials could distort assessments about the coronavirus and be used as a weapon in an escalating battle with China.



Kristian Andersen 16:25

Indeed. In fact, I blew up the call with the White House panel I'm on earlier this morning by suggesting that maybe we as a country should stop blaming others for our own failures and instead focus on making science-based decisions to get in front of this disaster - and that maybe we could write a letter to the president about that. I doubt I'll be invited back.



Robert Garry 16:43

Kinda shocking to see the "WIV or China CDC released this thing on the world" coming from both the left and the right. Trump has a few advisors that know exactly how to create a distraction. (edited)



Andrew Rambaut 18:12

It really doesn't help that the Chinese are trying to suggest that it didn't start in Wuhan (or Hubei, or even China).



Kristian Andersen 18:23

No. The Chinese blaming the Americans is about as unhelpful as the Americans blaming the Chinese.

From my perspective:

Shi Zhengli is not acting like it's a lab leak.

Kristian Andersen honestly changed his mind as he learned more.

The Chinese government is often evasive, but that doesn't mean anything because they're also denying an outbreak at the market. China has also tried to cover up other zoonotic disease outbreaks like a [pneumonic plague outbreak](#).

The most suspicious people of all are the ones selling the lab leak theory. Their behavior is the one thing we can look at to most easily see who's lying. To be clear, some of them probably believe everything they're saying – I'm mostly accusing specific known liars.

The covid origins scientists like Worobey and Pekar have done a good and honest job at showing that Covid spilled over at the Huanan market.

Is Bayesian analysis the best approach?
Some general thoughts on
Probabilistic arguments

Bayesian/probabilistic arguments can be more precise than other methods, but there are also many ways they can be wrong.

These analyses are not inherently superior to human reasoning. A human still chooses which evidence to include in the analysis, and how to weight that evidence.

The analysis can be skewed in one direction or the other by including or omitting evidence.

Rootclaim and I have both done probabilistic analyses that point in entirely opposite directions.

Only one of these can be correct. Perhaps one or both of us left out important factors.

A probabilistic analysis can fail if the underlying evidence is wrong (“garbage in = garbage out”)

My analysis could be wrong if the market data is wrong.

Rootclaim’s analysis could be wrong because they ignore the market data.

We need to establish the facts before we can set the probabilities.

Sometimes the odds are simply hard to estimate:

What is the effect on the odds of the virus having a furin cleavage site?

Alex Washburne says this is very rare: a [1 in 1,000 coincidence](#).

Rootclaim's [2020 analysis](#) says this only increases the odds of a lab origin by a **factor of 8**.

Probabilistic Estimates:

Adding a [furin cleavage site](#) is a known procedure in [gain-of-function](#) research involving coronaviruses, while almost no other SARS-related coronaviruses contain furin cleavage sites (out of [hundreds of related viruses](#)). Since this virus resulted in a pandemic, it is much more likely to have such a feature, and we thus [conservatively](#) estimate it is equally likely for a zoonotic virus and a [gain-of-function](#) virus.

However, the placement in the genetic code looks like it was the result of an insertion and not a mutation, which is estimated to be at least 5x less likely to occur [zoonotically](#). Additionally, it is estimated that the appearance of two successive CGG codons are at least 3x less likely if SARS-CoV-2 developed [zoonotically](#). Therefore, the likelihood that COVID-19 developed [zoonotically](#) is reduced by 15x, [conservatively](#) adjusted to 8x.

The first version of Rootclaim's 2023 arguments (in week 1) said FCS only increases the odds of a lab origin by a **factor of 2**.

Then in the week 2 version, they changed it to give it a **factor of 20-100**, because of the "clean insert".

I agree that a sarbecovirus with a furin cleavage site is interesting and rare, but Rootclaim doesn't seem to know how rare that is, they're just guessing as to which numbers to use.

There's nothing wrong with updating an analysis, I've also been trying to refine my analysis as this debate has progressed. I use data where possible, but some things are just guesses. The furin cleavage site odds are just a guess.

Let's look at some other Bayesian analyses

Some Bayesian analyses for lab leak simply ignore all the market evidence

This is one example, [from Michael Weissman](#)

He starts with 99% odds of a natural origin vs a lab leak, much higher than Rootclaim does.

He assigns a large weight to things like CGGCGG or “pre-adaptation”.

He throws in other “suspicious lab leak leaning things” and doesn’t give them any weight, to pretend like he’s being conservative.

He ignores all the evidence of a market origin.

And he ignores any questions of whether the lab was actually doing relevant work, whether they had a secret virus, why they would insert PRRAR out of frame, etc.

Observation	Comments on methods	Logit \pm S.E.
Priors	Previous lab vs wild pandemics, prior estimates of relevant lab leak rates, counts of emerging pathogens, expert estimates from Proximal Origin authors (wide disagreements found, often w.o. explanation)	-4.6 \pm 2.3
Found in Wuhan	Fraction of China’s population, fraction of China’s wet markets vs. fraction of locations where coronavirus modifications including FCS were planned at BSL2 or BSL3 (pretty much consensus view)	+4.6 \pm 0.7
Negative correlation between SC2 RNA and potential host DNA in HSM	Surprising for ZW at HSM, conservatively don’t use since compatible with ZW elsewhere
Worobey et al. early cases near or associated with HSM	Not used due to internal HSM negative correlation, late timing of cases, distance of sequences from MRCA, case ascertainment bias, unexplained missing 35% of cases, contradictory Weibo data, opposite conclusions of WHO-CDC study of same data, rejection of Worobey conclusion by China CDC head
No wildlife host found	Comparison with other emergent coronaviruses, 80000 animals sampled (conservatively ignores that intense testing)	+1.4 \pm 0.6
Sangon SC2	As expected for LL but consistent with ZW
Presence of FCS	Expected for LL under DEFUSE, unclear how surprising for ZW due to post-selection (conservatively not used)
FCS at S1/S2 junction	Expected for LL under DEFUSE, unclear how surprising for ZW due to post-selection (conservatively not used)
No amino acid changes near FCS	Expected for LL, surprising for ZW (conservatively not used)
CGGCGG for ArgArg in FCS	Pretty likely for LL, highly unlikely in related coronavirus sequences, no post-selection explanation	5.1 \pm 0.8
Pre-adaptation to humans/ lab cultures	Expected for LL, strong contrast with SC1 behavior and with SC2 in minks (conservative subjective factor)	1.1 \pm 0.5
post-bat adaptation but no hosts found	For ZW but not LL these features are in tension (hard to model, conservatively not used)
Inconsistencies and omissions in Wuhan data	Not used, should be analyzed by social scientists
Rumors of suspect Wuhan events	Unverified and potentially politically corrupted, not used
Proximal authors’ communications	Used only for vivid language supporting independently estimated factors

Table 1. A summary of the main observations considered in calculating the odds.

Here's a similar analysis, [from Louis Nemzer](#).

He starts with natural origin at 99.9%, a much higher value than Rootclaim does.

He increases lab leak by 100X for "near a BSL-4 lab", by 200X for the furin cleavage site, and by 55X for the double CGG.

Those are high odds for each of those things.

He also includes zero mention of the wildlife market, no odds for secret viruses, and no mention of any of the questionable engineering choices.

$$\text{Posterior Odds}(\text{Natural Origin}) = \frac{\text{Prior Odds}(\text{Natural Origin}) \times \text{BayesFactor}(\text{Has BSL4 Lab}) \times \text{BayesFactor}(\text{Has FCS}) \times \text{BayesFactor}(\text{CGGCGG})}{1}$$

Prior odds: I start with a natural occurrence being 1,000 times more likely than a nonnatural origin. This seems like a conservative estimate, but I think that before considering any evidence, it is a reasonable place to start. Although the possibility of a lab leak has raised for many years, and there have been lab leaks in the past (including with SARS), all known major pandemics so far have had a natural origin.

Outbreak starts near BSL-4 Lab: There is no reason to expect a natural origin should occur specifically in a city with a BSL-4 lab studying bat coronaviruses, as opposed to any other population center. That is why I assign it a likelihood of 1 out of 100. However, if a lab leak occurred, it makes sense that the pandemic would be first noticed in such a location (9 out of 10).

Furin Cleavage Site: The grant proposal EHA submitted to DARPA specifically mentions inserting a furin cleavage site into a consensus sequence of collected coronaviruses. It was known to the researchers that this kind of insert can significantly increase the pathogenicity of viruses (1 out of 5). Since known furin cleavage site exist only in more distantly related viruses, it was unlikely to come from recombination (1 out of 1000).

CGG CGG: Within the furin cleavage site, two adjacent arginine amino acids use the codon CGG in their RNA coding sequence. This motif does not occur anywhere else in sabecovirus genomes, (the family of coronaviruses SARS and COVID belong to). Finding a C followed by a G is disfavored in viruses, since it activates the host immune system. The baseline rate for an arginine amino acid to use a CGG codon is 3% in SARS-CoV-2, so the probability of naturally finding CGG CGG, assuming it has to be two arginines in a row to maintain the furin cleavage site, is $(0.03)^2 = 0.0009$. Scientists may have wanted to purposely activate the immune system for an attenuated virus, using CG as an adjuvant (1 out of 20). So, the complete calculation is:

$$\frac{1000}{1} \times \frac{1}{0.9} \times \frac{1}{100} \times \frac{1}{1000} \times \frac{(0.03)^2}{1/20} = \frac{1}{1000}$$

This implies that the estimated probability of a natural origin for SARS2 is $\frac{1/1000}{1+1/1000} \approx 0.001$, which is **0.1%**.

This is why, while a natural origin is obviously still possible, I don't think it should be considered the most likely explanation when compared with an anthropogenic cause.

Steven Quay's [Bayesian Analysis](#), 2021:

Zoonosis starts at 98.8%, higher than Rootclaim.

The heavy lifting is done by 6 things:

“location of first cases near WIV”

“Lack of evidence of seroconversion in Wuhan”

“Lack of posterior diversity”

WIV research and capacity to edit viruses

“Lack of furin cleavage site in other sarbecoviruses”

“Rare usage of CGG codons”

The first 3 items are actually evidence in favor of a market origin:

The first cases were at the market, not the lab.

The lack of diversity is because it started at the market.

The lack of seroconversion is because it just started.

Furin cleavage site does lean towards a lab origin.

The “opportunity” section is unclear.

And CGG codons are ambiguous or weak bayes factor.

This is yet another misleading and one sided analysis --

It includes lots of lab leak leaning evidence, but ignores all market leaning evidence.

Evidence	Zoonotic Origin	Laboratory Origin
Initial State	98.8%	1.2%
International committees to determine CoV-2 origin may not be impartial	98.8%	1.2%
Three key zoonotic papers: pros and cons	98.8%	1.2%
SARS-like infections among employees of the Wuhan Institute of Virology in the fall of 2019 reported by US Government	98.8%	1.2%
Location of first cases near Wuhan Institute of Virology	95.1%	4.9%
Lack of evidence of seroconversion in Wuhan and Shanghai	80.9%	19.1%
Lack of posterior diversity	30.8%	69.2%
Opportunity: The Wuhan Institute of Virology has publicly disclosed that by 2017 it had developed the techniques to collect novel coronaviruses, systematically modify the receptor binding domain to improve binding or alter zoonotic tropism and transmission, insert a furin site to permit human cell infection, make chimera and synthetic viruses, perform experiments in humanized mice, and optimize the ORF8 gene to increase human cell death.	30.8%	69.2%
Lack of furin cleavage sites in any other sarbecovirus	4.7%	95.3%
Rare usage of -CGG- single codons & no CGG-CGG pairs	0.5%	99.5%
Routine use of CGG in laboratory codon optimization, including Daszak & Shi	0.2%	99.8%
Spike Protein receptor binding region (200 amino acids) optimized for humans	0.2%	99.8%
Whole genome analysis shows pre-adaption of CoV-2	0.2%	99.8%
The finding of CoV-2 in Barcelona wastewater in early 2019 was an artifact	0.2%	99.8%
Shi and the WHO comment early on that CoV-2 seemed to begin with a single patient	0.2%	99.8%
Mammalian biodiversity between Yunnan and Hubei is significantly different, limiting a potential common intermediate host	0.2%	99.8%
The ancestor of CoV-2 can only obtain a furin site from other subgenera viruses but recombination is limited/non-existent between subgenera	0.2%	99.8%
Canvas of 410 animals shows humans and primates are the best, bats are the worst, for ACE2-Spike Protein interaction	0.2%	99.8%
A government requested review of samples collected from a mineshaft may have caused the COVID-19 pandemic	0.2%	99.8%
The Hunan Seafood Market and farmed animals in Hubei province are not the source of CoV-2	0.2%	99.8%
Line 2 of the Wuhan Metro System is the likely conduit of the pandemic and is the closest subway line to the WIV	0.2%	99.8%
Feral and domestic cats are not the intermediate host	0.2%	99.8%
Extraordinary pre-adaption for the use of human tRNA is observed	0.2%	99.8%
Evidence of lax operations and disregard of laboratory safety protocols and regulations in China	0.2%	99.8%
Previous SARS-CoV-1 laboratory accidents	0.2%	99.8%
Shi and Daszak use Wuhan residents as negative control for zoonotic coronavirus exposure	0.2%	99.8%
RaTG13 could be CoV-2 precursor using the synthetic biology 'No See 'Em' technique	0.2%	99.8%
Location, location, location: Based on the distance between known SARS-CoV-1 laboratory-acquired infections and the hospital of admission of the infected personnel, the WIV is within the expected hospital catchment for a CoV-2 LAI	0.2%	99.8%

Here's the conclusion of Rootclaim's 2020 analysis on covid's origins:

Hypotheses Considered ⓘ

Calculated Results ⓘ

1	Lab escape: The virus was developed during <u>gain-of-function</u> research and was released by accident.	89%
2	Bioweapon: The virus was genetically engineered as a bioweapon and was deliberately released.	4.5%
3	Zoonotic collection: The virus evolved in nature, and was contracted by virus researchers.	3.2%
4	Zoonotic: The virus evolved in nature and was transmitted to humans <u>zoonotically</u> .	3.2%





They worked through the probabilities, step by step, here are the first few steps:

Starting Point ⓘ

There have been many more viruses introduced to humanity zoonotically than through lab failures. Specifically, there were several major pandemics involving novel coronaviruses from natural origin in recent years. Although there have been no known outbreaks involving any novel viruses (coronavirus or otherwise) that came from research, there have been cases of lab leaks that were caught before causing widespread infections, including one lab leak (of a previously known virus) that led to secondary infections. There are also no known cases of a virus being released deliberately in modern history.

Before examining the specific evidence, the initial estimate of the probabilities of Zoonotic : Zoonotic collection : Bioweapon : Lab escape (based on their respective likelihood of incidents per year) is 78% : 6% : 16% : 0.6%.

Initial Probabilities

Name	Initial Likelihoods ⓘ	
Zoonotic		78%
Lab escape		0.7%
Bioweapon		16%
Zoonotic collection		6%





Show more ▾

Evidence ⓘ

Contagion and mortality

COVID-19 is more contagious than the typical flu, but not as fatal as recent viruses like MERS or SARS. Overall, it is not particularly well-suited as a traditional bioweapon, and COVID-19 broke out during a relatively peaceful time. This indicates that, if it was used as a bioweapon, it would probably not be released as a method of killing people but for a different purpose such as disrupting the world economy.

Effect ⓘ





Name	Effect ⓘ	Updated Likelihoods ⓘ	
Zoonotic	-		91%
Lab escape	-		0.8%
Bioweapon	+10		1.8%
Zoonotic collection	-		7%

Show more ▾

Outbreak location: Wuhan

The COVID-19 outbreak was first recorded in Wuhan, one of the larger cities in China. Large cities are often the initial breakout sites of zoonotic pandemics, but in that sense Wuhan is no more likely than any other city. It also isn't a particularly desirable target for releasing a bioweapon.

However, Wuhan stands out for housing the Wuhan Institute of Virology, one of only a few labs engaged in gain-of-function research.

Name	Effect ⓘ	Updated Likelihoods ⓘ	
Zoonotic	+20		51%
Lab escape	-		9%
Bioweapon	+15		1.4%
Zoonotic collection	+2		38%

Show more ▾

Mistakes made in Rootclaim's 2020 analysis:

They start zoonosis at 78% and bioweapon at 16%, very different priors from other Bayesian analyses.

They claim any mild virus showing up in Wuhan would have a 50/50 chance of being a lab leak, before even considering any features of the virus, like the furin cleavage site.

They assume that Covid is a chimera of a pangolin and bat virus. Since finding BANAL-52, that's not necessary.

The key mistake is that it doesn't recognize that the first known cases were at the Huanan market. The analysis states that if that were true:

Wet market: If the source was either a lab escape or a bioweapon, then the wet market is no more likely than the many other possible breakout locations in Wuhan. If it was certain that the wet market was the breakout location this would have a very large effect based on the large number of such locations, perhaps by 50x. But since it is not certain, as there were many documented cases before the wet market cluster, this potential breakout location conservatively reduces the likelihood of a lab escape, zoonotic collection, or bioweapon by 10x.

As we've seen, the odds the first cluster would be at the Huanan market are around 1 in 10,000.

They used only 1 in 10. And in their latest analysis they don't count the market at all.

Correcting for that one mistake alone, Rootclaim's 2020 analysis would show a 99+% probability that covid is natural.

Any honest Bayesian analysis would include factors from both sides

It should include the Furin Cleavage Site and the presence of a lab in the same city as the virus. Those should make us suspicious of that lab, and it was reasonable to be suspicious of the lab in early 2020.

But it also needs to include the factors that don't look engineered, like the out of frame PRRAR furin site.

It needs to include the odds they were secretly doing the DEFUSE grant, and the even lower odds they had a secret virus to start the engineering from.

It needs to include the failure of 4 years worth of lab leak theories to come up with any one theory.

And it needs to include the strong evidence pointing to the Huanan market as the point of origin.

Each of these lab leak analyses are skewed entirely towards the lab.

Very few non-lab leak people have even tried making a balanced Bayesian analysis. Here's [one analysis](#) I found that said [75% market, 25% lab](#), after including factors from both sides. They still don't apply much weight to the market data, but even including it at all makes zoonosis win.

And most scientists support the zoonotic origin, but they prefer not to just make up numbers for everything.

In other words:

What Rootclaim thinks
bayesian analysis is:



Superior to all human inference

What it actually is:



Motivated reasoning with
some numbers attached

Other problems to avoid, in a Bayesian analysis

Bayesian/probabilistic arguments can fail because of double counting.

Consider this argument:

“There’s a 1 in 100 chance the virus showed up in Wuhan, the same city that has the Wuhan institute of Virology.”

“And, since that lab is the best coronavirus lab in China, let’s adjust by a factor of 10 and say the odds are 1 in 1,000 that the virus is natural.”

As we’ve already seen, most big cities in China have labs.

The only reason we’ve specified Wuhan as interesting is because it has the best lab.

That’s a form of double counting.

Now consider this argument:

“The odds of a virus showing up in Wuhan are 1 in 100
They were researching furin cleavage sites, and the virus has a furin cleavage site.
The odds you’d find a furin cleavage site naturally are 1 in 100.
Now the odds are one in 10,000 that it’s a natural virus.”

Is that double counting?

Again, you only chose Wuhan as important because the lab was doing that work.

Almost every Chinese city has a lab, Wuhan is only important because they had “the furin cleavage site lab”. (which had never actually inserted any furin cleavage sites, mind you, but they were co-authors on the DEFUSE proposal, so this is a maybe)

Does specifying the cleavage site change the odds or not?

I think this may still be double counting. Maybe you get the odds of the cleavage site or the odds of the city, but not both.

Maybe you need to adjust the Wuhan odds because you could also make an FCS theory for Beijing and a few other cities.

I’m not sure there is a simple way to define double counting here. One definition might be: would you still make the lab leak theory if the virus did not have a furin cleavage site? If not, then you’re double counting.

Your Bayesian argument can be wrong if you specify things that are rare but unimportant:

Example: someone is murdered in one building in Wuhan. There are 10 million buildings in Wuhan. The odds are 1 in 10 million that it would happen in that building. Therefore, the odds are extremely low it would happen there, this murder should be seen as highly suspicious.

But the 1 in 10 million is not relevant, in that case. It had to happen in some building.

When I mentioned the importance of the Huanan market, I said it was a 1 in 10,000 coincidence that the virus would start there. The market was important because it had previously been identified as a high risk place for a viral spillover, 4 places in Wuhan sold such wild animals, and it sold more wild animals than any such place in Wuhan.

Those odds were based on analysis of traffic to the market, or the number of employees working at the market.

I did not say it was a 1 in 10 million coincidence, because there are 10 million buildings in Wuhan.

Had the virus first been found at some other random place that did not sell susceptible animals, that location might have no significance at all. If the first cases were in a movie theater, that wouldn't adjust the odds at all, because there's no clear source for a novel virus in a movie theater.

It might be the same with patterns like the Faul site or the double CGG.

Faul is only one of many restriction sites. The odds are ~50% you'd find a similar length restriction site in the furin cleavage site, or 99.5% you'd find a restriction site of any length.

CGG is not the preferred codon of labs.

It is rare in nature, but if you don't understand why people would choose it, then maybe it doesn't matter.

It's not that much rarer as a natural insertion (bayes factor between 1 and 5.5, depending on whether it was copied from host or viral RNA).

Also, it seems to have an evolutionary purpose: it improves protein folding accuracy. So maybe it was selected for and relative odds in nature and engineering don't matter.

Motivated reasoning can find suspicious patterns anywhere

You can find suspicious patterns anywhere

To demonstrate how motivated reasoning works, let's make a mock conspiracy theory about another virus called HKU-1.

There have been 3 coronaviruses coming out of China in the last 20 years.

After SARS and before covid, another human coronavirus was discovered. It's called HKU-1.

The first case was discovered in a man in Hong Kong, but he'd just travelled to Shenzhen, and we think he got sick there.

The virus has a furin cleavage site, RRKRR. That's an optimal cleavage site, exactly the kind of thing scientists would insert into a virus in an experiment.

Alpha	HCoV-NL63	735	-	GICADGSLI----	PVRRPNSS	-	751
Alpha	HCoV-229E	554	-	GVCADGSII----	AVQPRNVS	-	570
Beta 2a	HCoV-OC43	753	-	GYCVDYSK-----	NRRSRGAI	-	768
Beta 2a	HCoV-HKU1	742	-	GFCVDYNSPSSSSSS	RRKRRSI	-	762
Beta 2b	SARS-CoV	655	-	GICASYHTVS-L----	LRSTS	-	670
Beta 2b	SARS-CoV-2	669	-	GICASYQTQT-NSP	RRARSVA	-	688
Beta 2c	MERS-CoV	734	-	SLCALPDTPSTLTP	RSVRSVP	-	754

What are the odds a virus has a furin cleavage site?

Estimates range anywhere from 1 in 2 up to Alex Washburne's [1 in 1,000](#).
Let's go with 1 in 100.

Let's look at how those [amino acids are encoded](#) for RRKRR:

CGT CGT AAA CGT AGA

There's a triple CGT!

In covid, 2 out of 3 of the arginines are spelled CGG.

HKU-1 is even more suspicious, with 3 out of 4 spelled CGT.

Each one is 10%, so the odds are 1 in 1,000 of that happening by chance.

So we're at 1 in 100 odds for the FCS times 1 in 1,000 for the triple CGT =
1 in 100,000 odds this virus is natural.

codon frequency for HKU-1

CGT	9.6%
CGC	3.3%
CGA	8.5%
CGG	3.0%
AGA	50.2%
AGG	25.4%

The furin cleavage site isn't the only thing that jumps out at me here.

Right before that we see SSSSS, five serines in a row!

Alpha	HCoV-NL63	735	-	GICADGSLI----	PVRRPNSS	-	751
Alpha	HCoV-229E	554	-	GVCADGSII----	AVQPRNVS	-	570
Beta 2a	HCoV-OC43	753	-	GYCVDYSK-----	NRRSRGAI	-	768
Beta 2a	HCoV-HKU1	742	-	GFCVDYNSP	SSSSSRRKRRSI	-	762
Beta 2b	SARS-CoV	655	-	GICASYHTVS-L----	LRSTS	-	670
Beta 2b	SARS-CoV-2	669	-	GICASYQTQT-NSP	RRARSVA	-	688
Beta 2c	MERS-CoV	734	-	SLCALPDTPSTLTP	RSVRSVP	-	754

I searched through the rest of the genome. S is never repeated more than 3 times in a row.

I searched for the other 19 amino acids. None of them are ever repeated more than 3 times in a row.

Yet, here, right at this critical junction, the furin cleavage site, we find five serines in a row!

What are the odds of seeing S, five times in a row? There are 20 amino acids, so it's:

$(1/20)^5$ or 1 in 3.2 million!

Multiply 3.2 million by 100,000 and we now have **1 in 320 billion!**

This is clear evidence of human design.

HKU-1 was found in Shenzhen, the same city that has the “[Shenzhen institute of synthetic biology](#)”.

Coincidence? Or, did **they** use synthetic biology to create it?

The odds are 1 in 100 that it would show up in Shenzhen, compared to some other city in China.

Put that all together: the odds are only **1 in 32 trillion** that HKU-1 is a natural virus.

It’s overwhelmingly clear that HKU-1 was made by scientists!

If I wanted to increase that number, I could pick out other things about it.

[Going over wikipedia:](#)

“HKU-1’s S protein has been noted for [its large size](#).”

Did scientists enlarge the spike protein?

“The HE protein differs from conventional ones (such as the one in OC43) by having a much smaller vestigial lectin domain. This domain is shielded from recognition by the immune system via size changes and glycosylation.”

Did scientists change the glycans to create a virus that evades the immune system?

I looked for restriction sites in HKU-1's furin cleavage site and I found 10 of them!

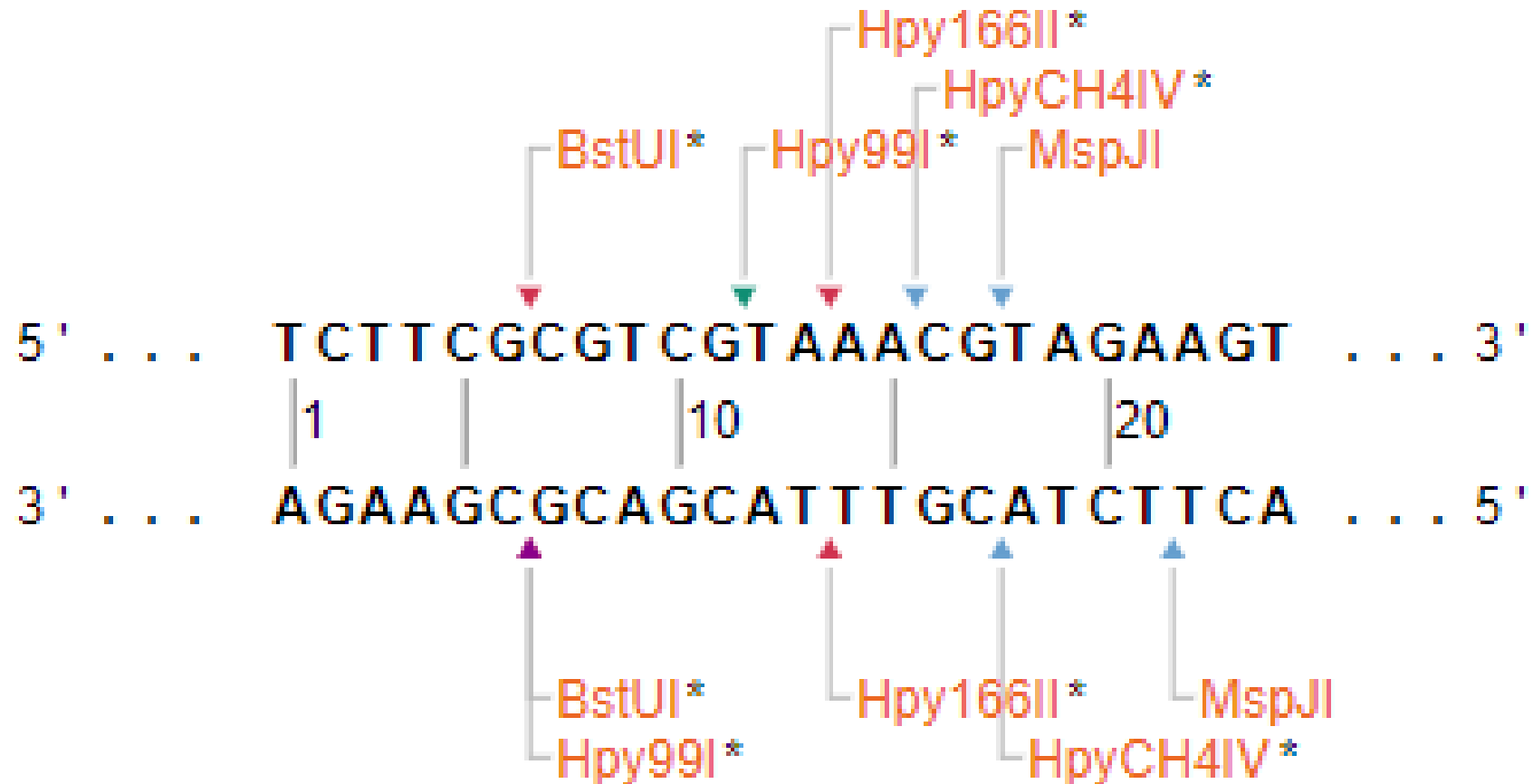


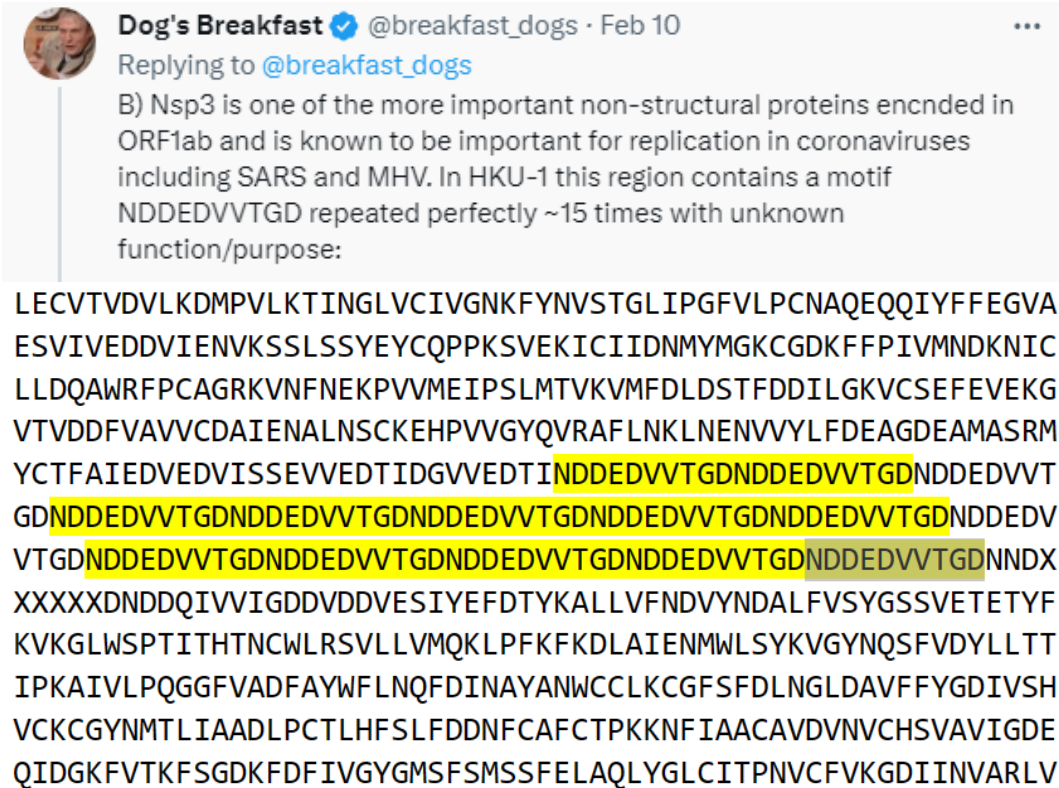
Image made [with NEBcutter](#)

After making this theory, I noticed that some guy on Twitter [made an actual conspiracy theory for HKU-1](#).

And now I feel like an amateur. He found all the same stuff I did, like the quintuple serine. But he found even more suspicious things elsewhere that could be seen as “smoking guns” for lab creation.

This one was my favorite weird and spooky thing that he found – 15 repeats of the same 10 amino acid sequence. That looks completely unnatural.

The odds of that happening randomly would be $(1/20)^{150}$. Let’s just call that zero.



Dog's Breakfast @breakfast_dogs · Feb 10
Replying to @breakfast_dogs

B) Nsp3 is one of the more important non-structural proteins encded in ORF1ab and is known to be important for replication in coronaviruses including SARS and MHV. In HKU-1 this region contains a motif NDDDEDVVTGD repeated perfectly ~15 times with unknown function/purpose:

LECVTVDLKMPVLKTINGLVCIVGNKFYNVSTGLIPGFVLPNAQQIYFFEGVA
ESVIVEDDVIENVKSSLSSYEYCQPPKSVEKICIIDNMYMGKCGDKFFPIVMNDKNIC
LLDQAWRFPCAGRKVNFNKPVVMEIPSLMTVKVMFDLDSTFDDILGKVCSEFEVEKG
VTVDDFVAVVCDAIENALNSCKEHPVVGQVRAFLNKLNENVVYLFDEAGDEAMASRM
YCTFAIEDVEDVISSEVEDTIDGVVEDTINDDDEDVVTGDNDDEDVVTGDNDDEDVVT
GDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDV
VTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDX
XXXXDNDQIVVIGDDVDDVESIYEFDTYKALLVFNDVYNDALFVSYGSSVETETYF
KVKGLWSPTITHTNCWLRVLLVMQKLPFKFDLAIENMWLSYKVGYNQSFVDYLLTT
IPKAIVLPQGGFVADFAYWFLNQFDINAYANWCCLKCGFSFDLNGLDAVFFYGDIVSH
VCKCGYNMTLIAADLPCTLHFSLFDDNFCAFCTPKKNFIAACAVDENVCHSVAVIGDE
QIDGKFVTKFSGDKFDFIVGYGMSFSMSSFELAQLYGLCITPNVCFVKGDIINVARLV

Here's HKU-1's furin cleavage site, [as compared to some related virus](#):

```

AYR18599.1  734  GSGFCVDY - - - - STAHRGRRALSTGYR
HKU1       740  GSGFCVDYNSPSSSSRRKRRSISASYR
    
```

The FCS doesn't look inserted, that was already there but it mutated to be a better one (Or, did the lab choose a perfect cleavage site?)

But there is a 15 nucleotide insertion right before that.

If this 15 NT insert is natural, it gets a lot harder to say that Covid's 12 NT insert is clearly engineered.

HKU-1 also has some 15-18 nucleotide "clean insertions" relative to its ancestors [in the CTD region of S1](#):

AYR18599.1	366	SCNNMDASKVYGMCFSSILIDKFAIPNSRRVDLQLGNSGFLQSFNYKIDTTATSCQLYYSLPQNNVTVNNRNPS	439
HKU1	369	SCNNFDESKIYGSCFKSIIIDKFAIPNSRRSDLQLGSSGFLQSSNYKIDTTSSSCQLYYSLPAIINVTINNYNPS	442
AYR18599.1	440	SWNKRFGFSGFNLDSYS DHRYS VAYAQQCFQVPTDYCPADPNVVNTCVQNKPLSAACPVGTNQRACRFTA--	511
HKU1	443	SWNRRYGFNNFNLSHS -----VVYSRYCFSVNNTFCPCANPSFASSCKSHKPPSASCPIGTNYRSCESTVLD	511
AYR18599.1	512	----- CTCSCQPDPINTIAGDSCLOKKVLVGVGEHCPGIGINEDKCGNPSG ----- CFCAADAFVGSHESSCLSN	576
HKU1	512	RTDWCRCSCLPDPITAYDPRSCSQKKSLLVGVGEHCAGFGVDEEKCGVLDG SYNAS CLCSTDAFLGWSYDTCVSN	585
AYR18599.1	577	GRCQIFANLILNGVNSGTTCSSTDNQKANTEVLIDVCVDYDLYGITGRGIFKEVSADYYNSWQNLLYDVNGNLYG	650
HKU1	586	NRCNIFSNFILNGINSGTTCSNDLLQPNTEVFTDVCVDYDLYGITGQGIFKEVSAVYYNSWQNFLYDSNGNIIG	659

You can make a conspiracy theory about anything.

If HKU1 had been a bad pandemic, perhaps we'd all be talking about the triple CGT, the quintuple serine, and the Shenzhen institute of synthetic biology.

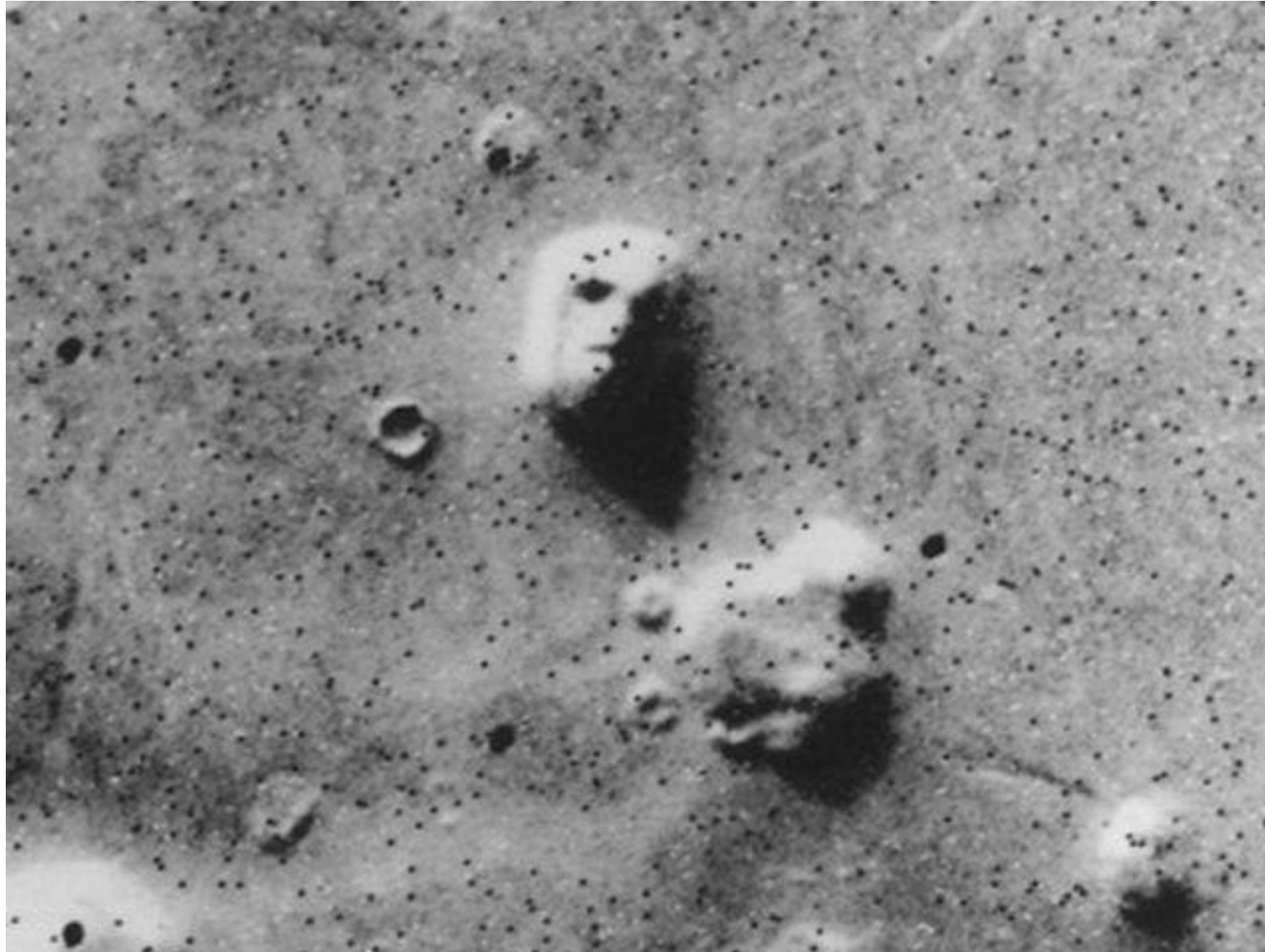
Maybe we'd talk about that mysterious sequence of 10 amino acids that was repeated 15 times in a row.

Or maybe we'd talk about the 15 nucleotides inserted before the furin cleavage site, or the ones inserted into the CTD.

Lab leak theorists have had 4 years to hunt for similar anomalies with Covid, and they've found lots of patterns that look suspicious.

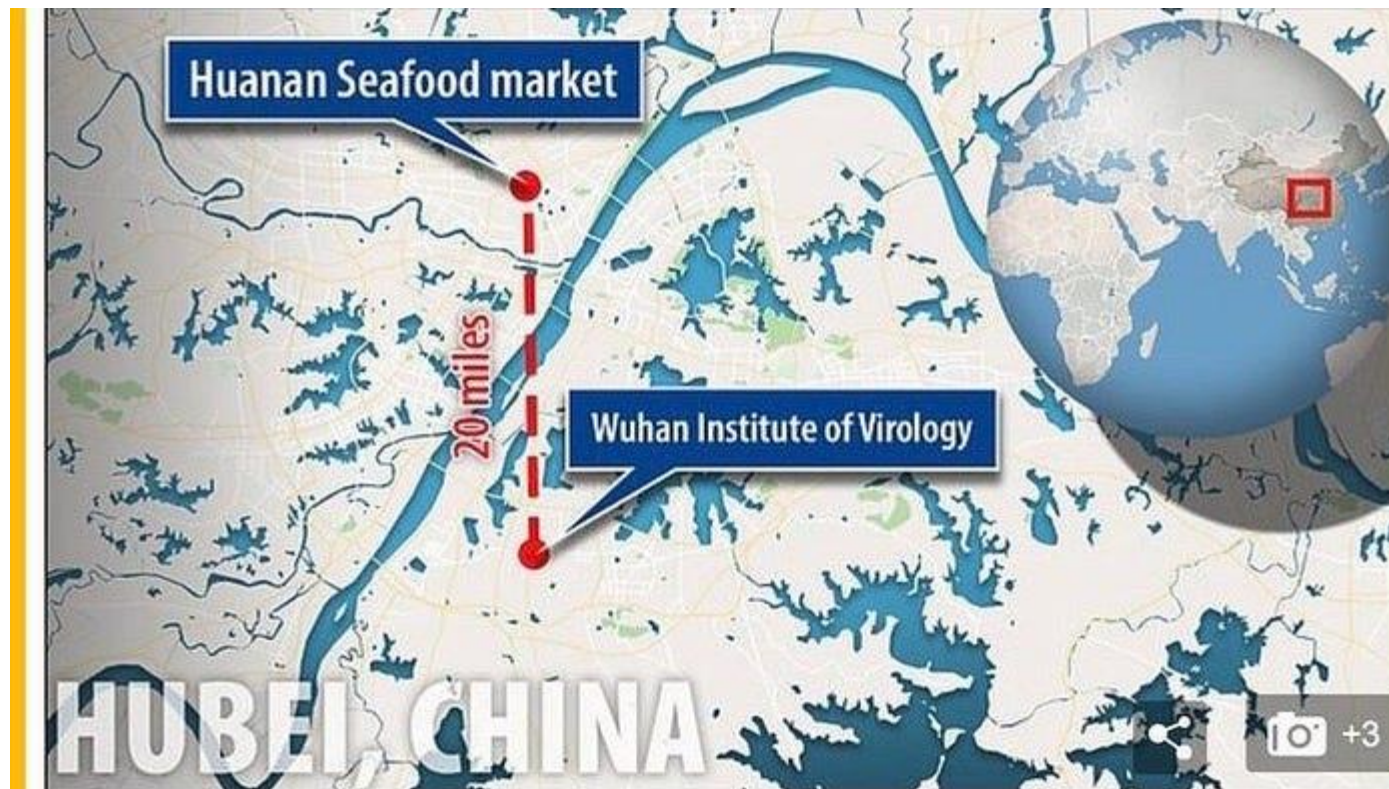
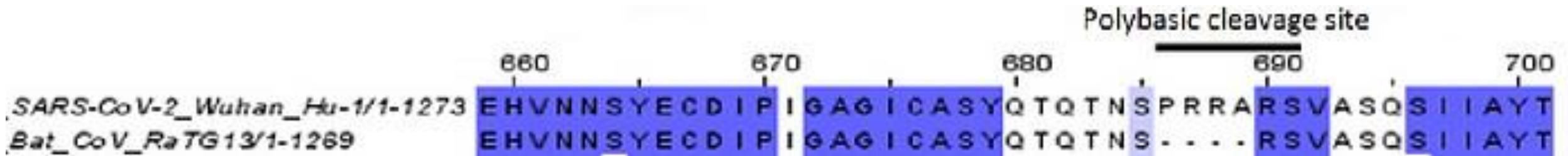
We see patterns in everything, and see signs of human intent

Sometimes those patterns are spooky. Everyone sees them:

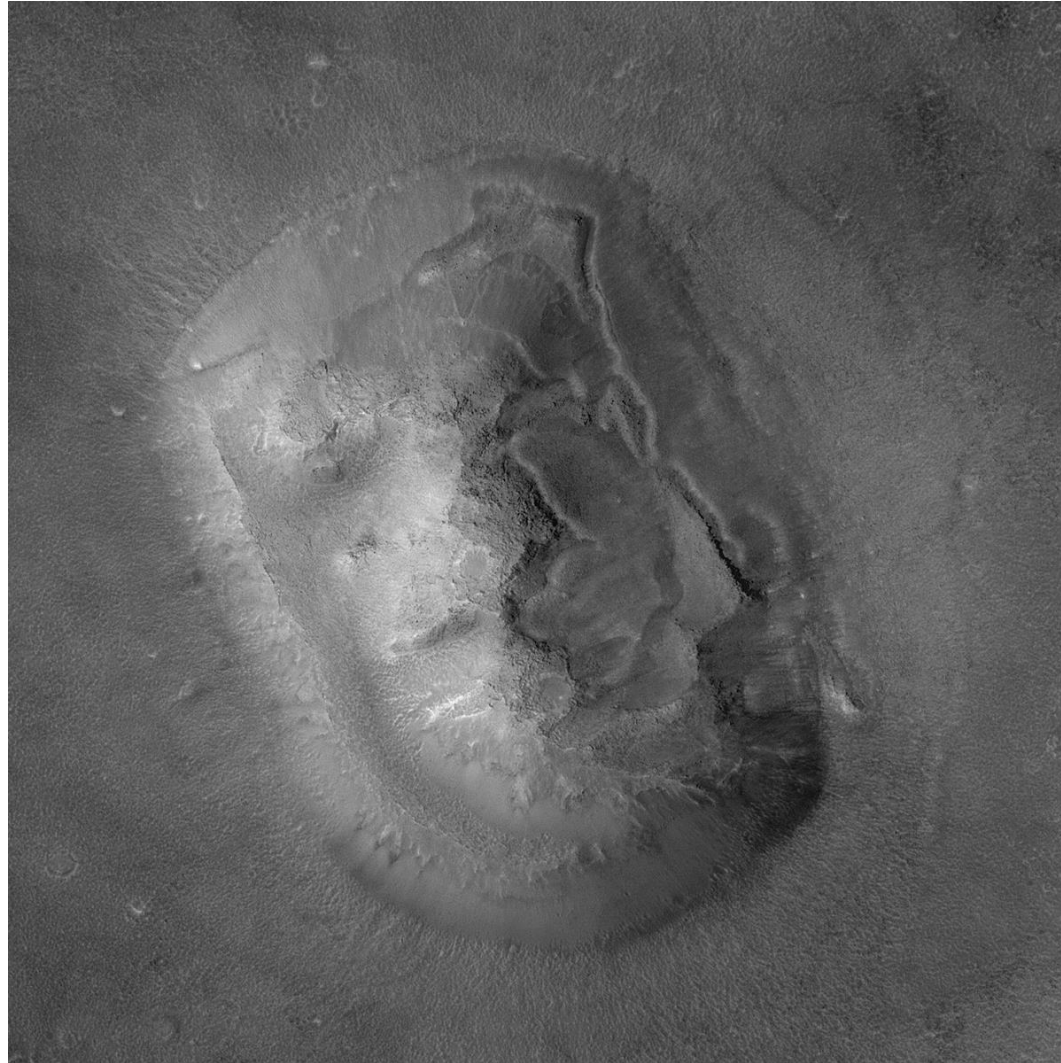


Face on Mars, Viking 1976

In this case, the virus has a furin cleavage site and it started near a lab.



Those patterns go away with more time and attention:



Mars global surveyor, 2001

As we continue to investigate the origins of Covid, I predict all suspicious patterns will continue to fall away, one by one.

It can take time to find the answers.

It took 25 years to launch another spacecraft to orbit Mars.

It took decades to find the source of Ebola (also bats). We still don't know how Ebola gets from bats to humans.

It took 5 months to find the host for SARS, even with the markets still open.

It took 15 years to find a cave that was more or less convincing of the bat origin of SARS. It still wasn't exact — the closest bat virus is 96% similar to SARS. We still don't know how SARS got from bats into civets. We still don't know where the infected civets came from. It could have been a farm in Hubei or Yunnan or somewhere else. We may never know.

With covid, we've already found a 96.8% similar bat virus.

We already understand where the human ACE2 binding comes from.

We've found the virus originated at a market that was already flagged as dangerous.

We've found the most likely shop and some likely animals.

And we'll probably learn more over the next 5 to 10 years.

I'm suggesting the judges vote for a natural origin, making a prediction that we'll resolve any other mysteries in time.

It's important that we understand that Covid is natural, so that we can focus on the changes that will prevent the next pandemic.