



Origin of Covid-19: **Lab Leak**

Session 3: Additional Evidence + Conclusion

A case by Rootclaim, presented by Saar Wilf

Loose Ends - Rootclaim & Fact-checking

In case it wasn't obvious, the description that Rootclaim does not do fact-checking does not mean we don't fact-check our evidence, just that we don't publish fact-check reports.



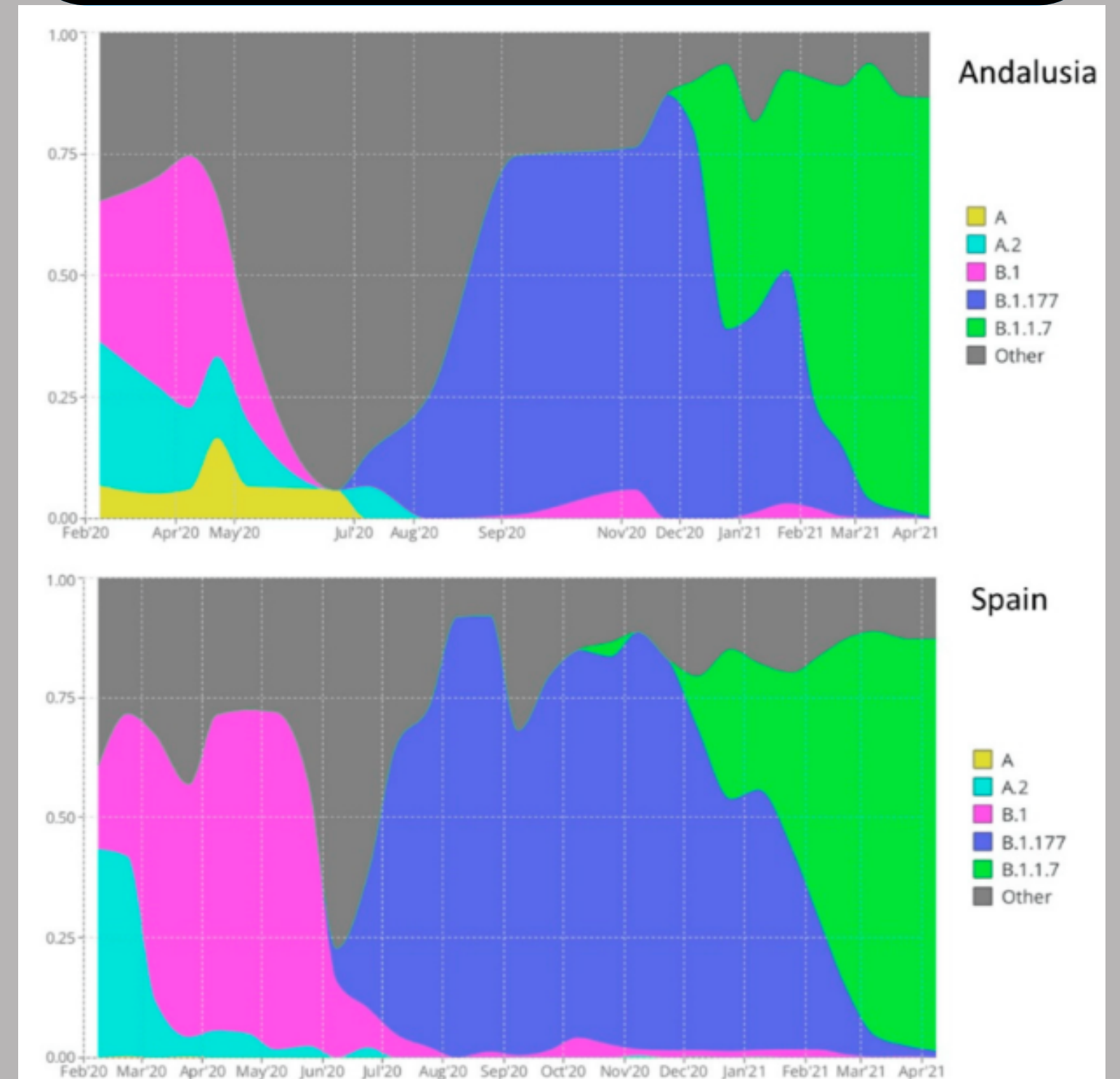
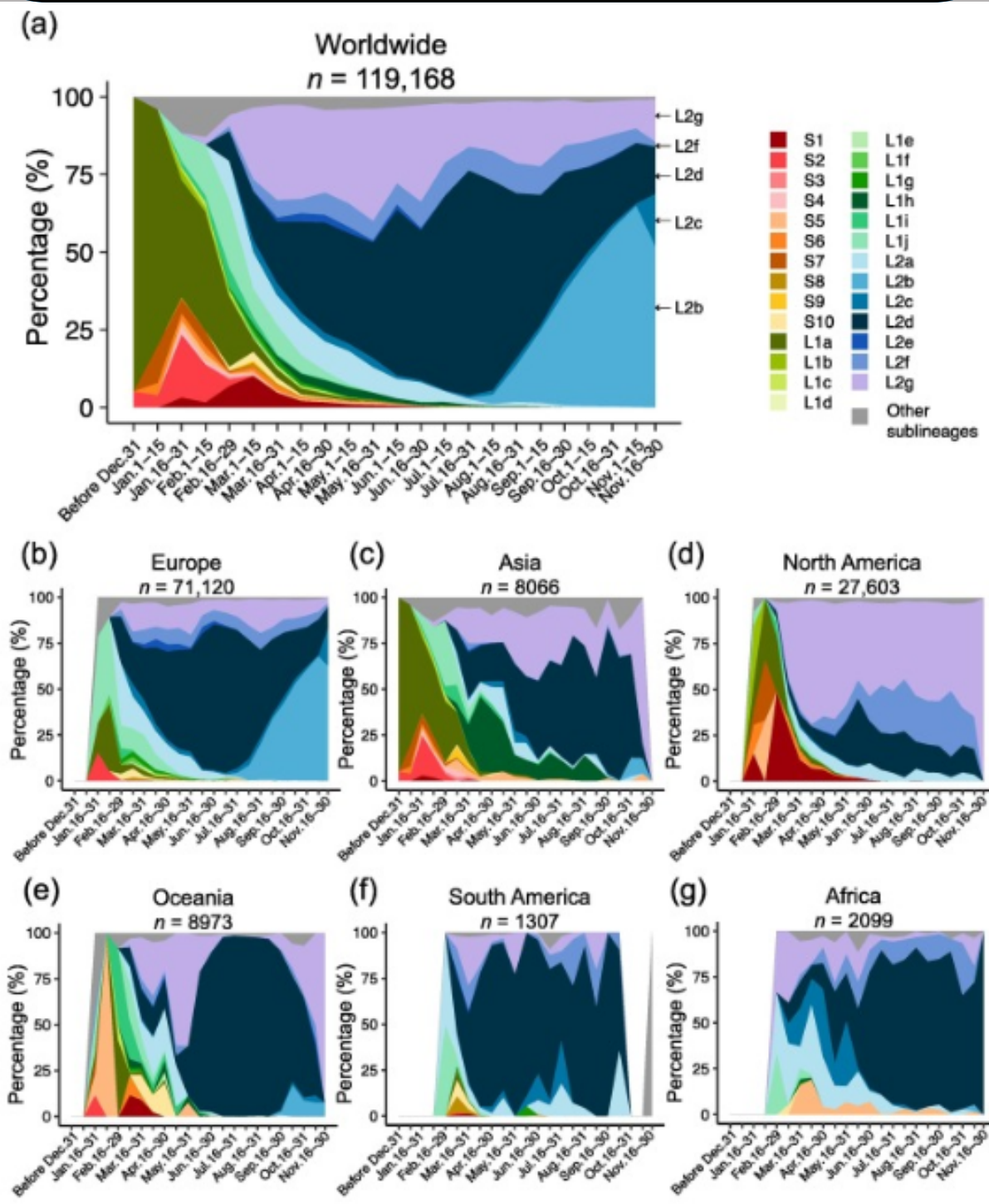
We occasionally get a fact wrong in our analysis, but the methodology knows to account for that possibility, so it never significantly changes a conclusion.

Loose Ends - B Advantage Over A

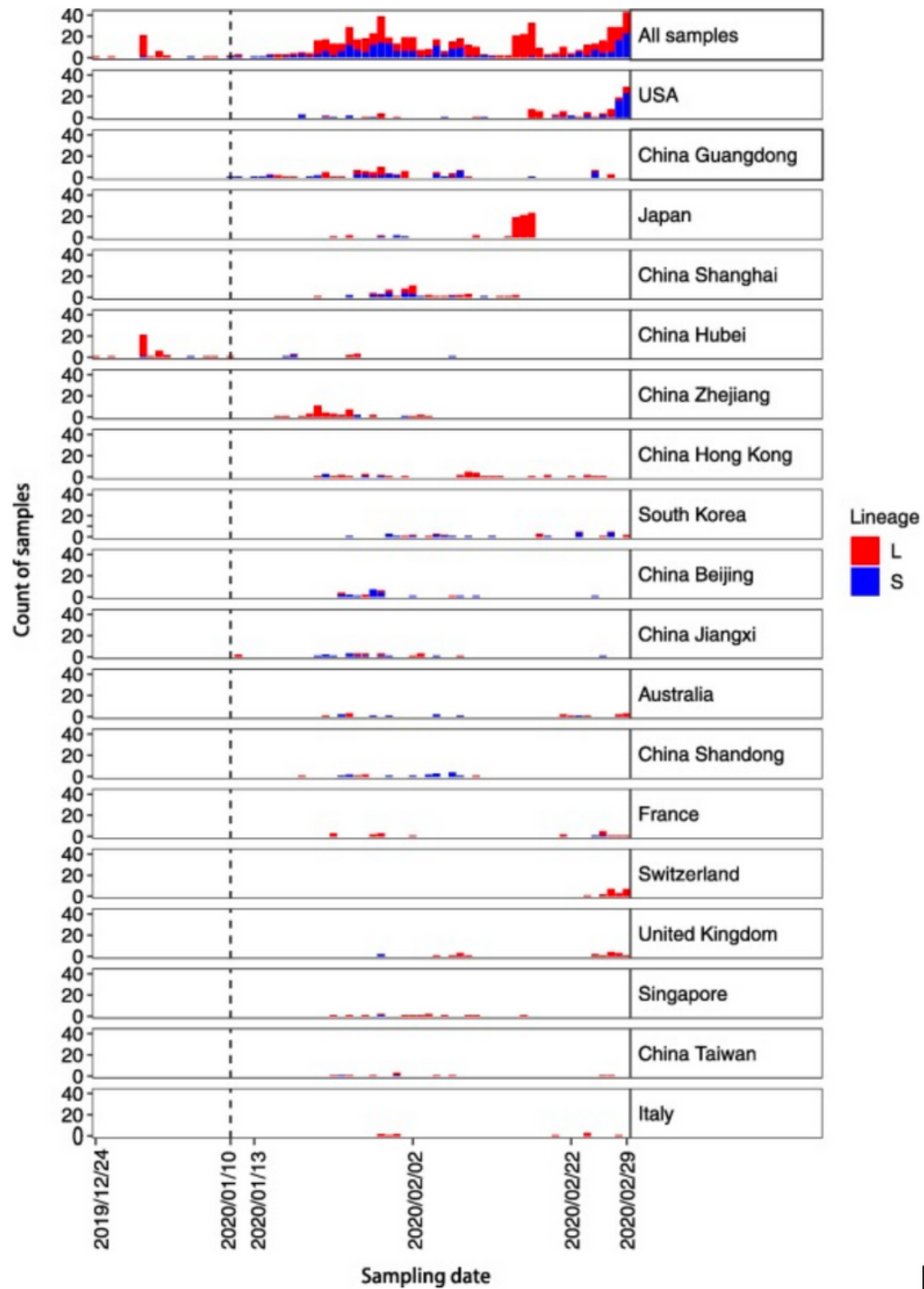
Peter has suggested that B does not have a fitness advantage over A and the replacement was due to a founder effect.

This paper charts the progress of A(S), and B(L) in various countries:

A similar picture seen in Spain



Loose Ends - B Advantage Over A



Similar pattern in other countries and within China

The studies show A losing to B in multiple geographies, indicating a founder effect is unlikely.

Loose Ends - No Animal Vendors Tested

In previous sessions, our opponent claimed that we should disregard the lack of infected animal vendors, which we would expect in a spillover and saw in SARS1, because no animal vendors were tested, implying their infections might have simply not been recorded.

- However, this does not explain why no animal vendor was hospitalized or discovered in the immediate and retrospective investigations.**
- We also have 53 HSM infections with direct exposure to HSM, most of whom either work there or are repeat buyers, and another dozen or more interviews to the media of people from the market. Though both in the investigations and the interviews people were asked if they knew of others who were sick in the market, many people were mentioned but none were animal vendors.**
- In summary: It's not just testing. No one knows of sick animal vendors and no animal vendor needed any medical attention.**

- Our opponent further tried to explain this by stating that maybe animal vendors developed immunity to SARS over the years of selling animals.**
- This explanation is extremely unlikely. People lost immunity from delta to omicron which is 30 mutations. they won't have SARS2 immunity from some other sarbecovirus with 1000 mutations.**

Loose Ends - Negative Mahjong Sample (1/2)

- In Session 1 our opponent mentioned there was a negative sample taken from a Mahjong table.
- We do not actually know this table sample is from the mahjong room. Mahjong is extremely popular in China and the table could be from the back of one the shops.
- Our opponent presented the following screenshot for the negative Mahjong sample, without linking to the source.

Sample ID	Lab code	Sampling date	Sampling location	Street No. ^a	Vendor No. ^b	Sample type	Sample information	Animal Spe	SARS-CoV-2 qPCR r
Env_0843	C13	20/02/2020	West Wine of HSM	5	NA	Environmental swab	Mahjong table	\	Negative
Env_0844	C14	20/02/2020	West Wine of HSM	5	NA	Environmental swab	Mahjong table	\	Negative

However, he cut out the columns on the right (after “negative”):

Aquatic (f	Seafood (Poultry (t	Livestock	Wildlife (Vegetable	Cold chain (type of vendor
no	no	yes	yes	no	no	yes

- According to the paper, “Yes” in these columns “means that the vendor where the sample was collected sells this type of product [..]. These pieces of information together indicate the scope of business of the vendor where the sample was collected.”
- We can not say with certainty, as the table does not provide enough details, but this might mean that the location of the Mahjong table sampled is a place that has a vendor that sells poultry, livestock, and cold chain products, which doesn't fit a mahjong room but a mahjong table in a vendor’s shop.

Loose Ends - Negative Mahjong Sample (2/2)

- Even if we assume this sample is from the Mahjong room, this sample is from Feb 20, and while it could have degraded on any surface by then, it is almost certain to have done so on a mahjong table.
 - Mahjong tables are covered with cloth made from cotton or wool. According to studies, COVID-19 can not survive on such fabrics for more than 24 hours and so, detection 2 months after exposure is extremely unlikely.
- In any case, we have very reliable testimonies that dozens of players from the mahjong hall were infected (See slide 82 in our first presentation). One negative sample is significantly weaker evidence.
- Our opponent further stated that “Drains & shop still tested positive” on the same late date of Feb 20.
 - However, drains have moisture combined with organic material, and multiple studies have shown these conditions significantly extend COVID’s survivability.
 - And while our opponent stated the other positive sample was from a “shop”, the samples were from the ground and a “container” inside this shop.
 - A container might be sealed or partially closed, which would trap moisture and avoid sanitization and UV light.
 - The ground is also likely asphalt or old concrete, which would trap moisture and protein (from dead animals and blood on the market’s floor).

Loose Ends - Unequal Conditional Probabilities

- When discussing evidence of engineering in the genome, we say that certain features (the 12nt insertion) is strong evidence for engineering, but it's difficult to find strong evidence against engineering (e.g. the leading Proline), since it's hard to speculate what an engineer would or would not find useful.

- Peter viewed this as unfair, comparing it to moving the goalposts.
 - That is just a fact about probability.
 - Imagine trying to determine whether markings on a paper are random or designed. A designer can always decide to make random-looking markings, but randomness cannot produce a very ordered marking.
 - However, depending on how frequently we expect to see the evidence, the lack of it could be strong evidence for the other side.
 - More mathematically.
 - $p(\text{an ordered marking}|\text{natural})=0.001$
 - $p(\text{an ordered marking}|\text{man made})=0.5$
 - If priors 1:1
 - Posterior with a clean marking: 99.9%
 - Posterior without a clean marking: 66.7%
 - But if $p(\text{an ordered marking}|\text{man made})=0.99$, then seeing no ordered marking would be strong evidence for randomness.
 - It is “fair” - because if the evidence is common, then its absence is strong. And if it's not common, then it rarely has an effect, and the total effect is small. We just happened to be “lucky” to have it in SARS2.
 - Note that sometimes it works for the other side: An animal infected with an ancestral virus would be very strong evidence for zoonosis. But an animal with no infection is very weak evidence.
 - However, testing many animals with no results is interesting, because at some point we'd expect to find at least one.

Loose Ends - 9 out of 23 animal vendors infected in SARS1

On Slide 81 of Day 1, Peter states that we say 9 out of 23 vendors infected in SARS1. He says *“It’s actually 1 of 7 for earliest SARS Cases”*

This was a mistake confusing two studies.

The first study was in a Shenzhen market selling wildlife during SARS, showing slightly different numbers: not 9 out of 23, but 8 out of 20:

“Although 8 out of 20 (40 percent) of the wild-animal traders and 3 of 15 (20 percent) of those who slaughter these animals had evidence of antibody, only 1 (5 percent) of 20 vegetable traders was seropositive.”

The second study that mentioned “9 out of 23” was not about vendors in a market but the total number of early cases in the SARS outbreak who were food handlers:

“A high proportion (9/23, 39%) of early cases were food handlers (this category includes persons who handle, kill, and sell food animals, as well as those who prepare and serve food), but none were farmers handling livestock or poultry. Of the nine early cases in food handlers, seven were restaurant chefs working in township restaurants (where a variety of animals were slaughtered on the premises), one was a market produce buyer for a restaurant, and one was a snake seller in a produce market (where a variety of live animals were offered for sale).”

- *This study further illustrates the striking difference between SARS1 and SARS2: 9 out of the first 23 cases in SARS1, which was a zoonotic spillover, were people directly working with the infected animals.*
 - *Additionally, we see that restaurants, and not wet markets, are the common spillover location. This was repeated a year later with an outbreak starting in a restaurant handling live civets.*
 - *<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3367621/>*
- *Contrary to SARS1, in HSM no one working with the hypothesized infected animals tested positive, was treated or hospitalized, and no such case was reported by the roughly 80 people either working in the market who were questioned in investigations, or who gave testimonies to the media.*

Loose Ends - Xinfadi Numbers (1/2)

Originally we compared "linked cases" in HSM vs Xinfadi as 33% vs 78%.

- In Xinfadi, 78% of cases had direct exposure to the market, and the rest (22%) were indirect infections. No other infections were found.

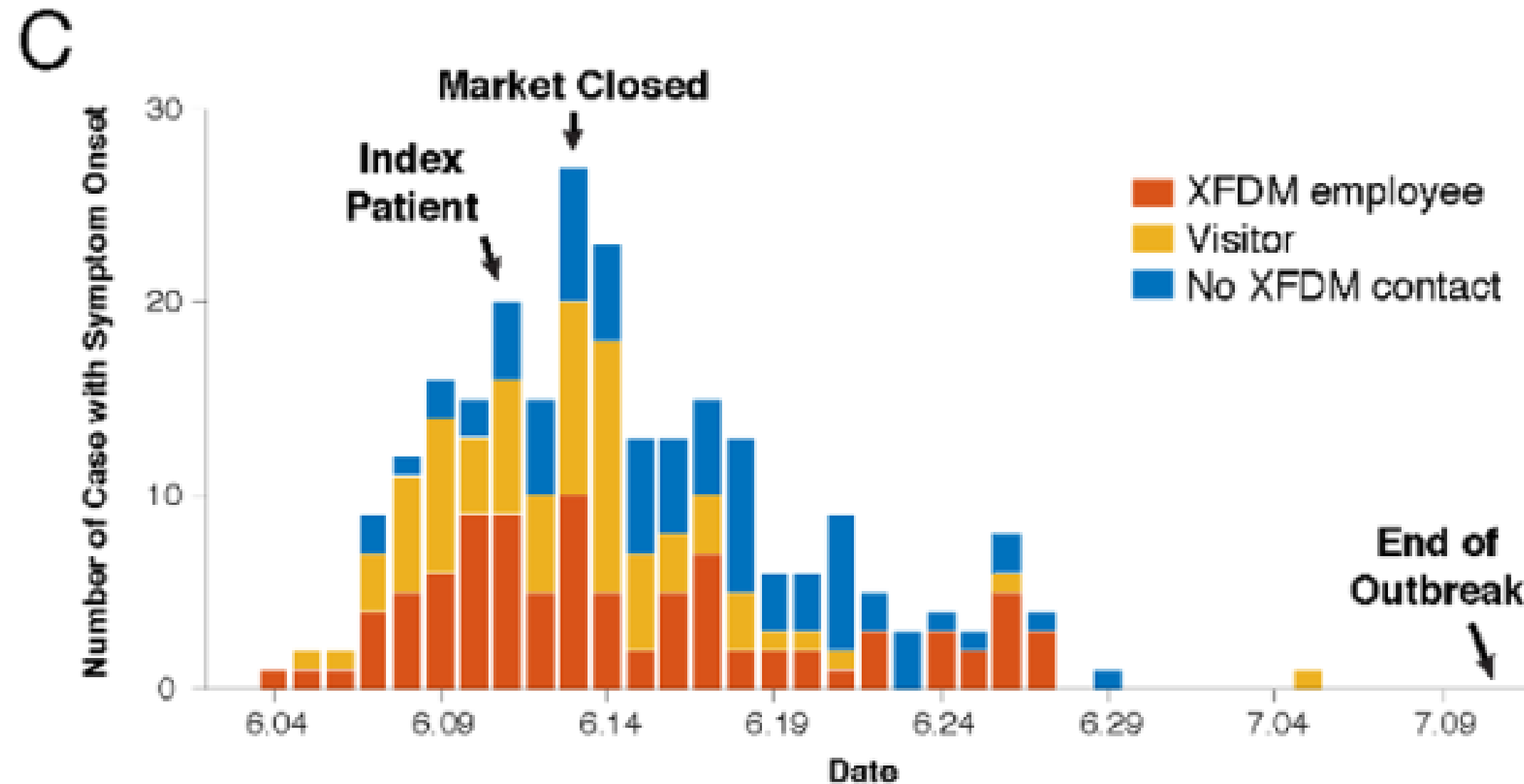
HSM direct exposure cases accounted for 33% (55/168) of the total cases but this included 2 indirect exposures (which we incorrectly listed as direct).

Types of exposure		
Direct exposure	Vendor ^a	30
	Purchaser ^b	12
	Passer-by ^c	5
	Buyer ^d	3
	Deliveryman ^e	2
	Visitors ^f	1
Indirect exposure	Contact of the Huanan	
	Market exposed population	2

Thus, we see only 33% of cases were HSM-related, whereas 100% were Xinfadi-related.

Loose Ends - Xinfadi Numbers (2/2)

This is partially explained by the fact HSM had roughly a month between the first infection and closure of the market, While Xinfadi had about two weeks (adding 2-4 days from infection to symptom onset).



However, this is still insufficient to account for the extreme difference we see between the markets with regards to the proportion of cases that have direct or indirect exposure to them. Specifically, Chen and Connor Reed (discussed later) are 2 and 5 weeks before closing HSM respectively, and seem to be completely unrelated to the market.

Loose Ends - How did the WIV P0 get infected?

- In Session 1, Eric said he understands how COVID in the air would easily infect a lab worker working in a BSL-2 lab but how did it get there, to begin with?
- Some options include sharing air with infected animals, inhalation of aerosolized virus from culture supernatant, an accidental needle prick, or an animal bite.
- We would like to add that the CDC safety Guidelines for bio-labs state that in 50-80% of lab leaks, the exposure method is unknown but most of these unknown cases are assumed to be inhalation-based infections, as these are the only ones a lab worker can be unaware of.

1.4. Routes of Laboratory Infection - The five most predominant routes of LAIs are:

- parenteral inoculations with syringe needles or other contaminated sharps;
- spills and splashes onto skin and mucous membranes;
- ingestion or exposure through mouth pipetting or touching mouth or eyes with fingers or contaminated objects;
- animal bites and scratches (research laboratories or activities); and
- inhalation of infectious aerosols.

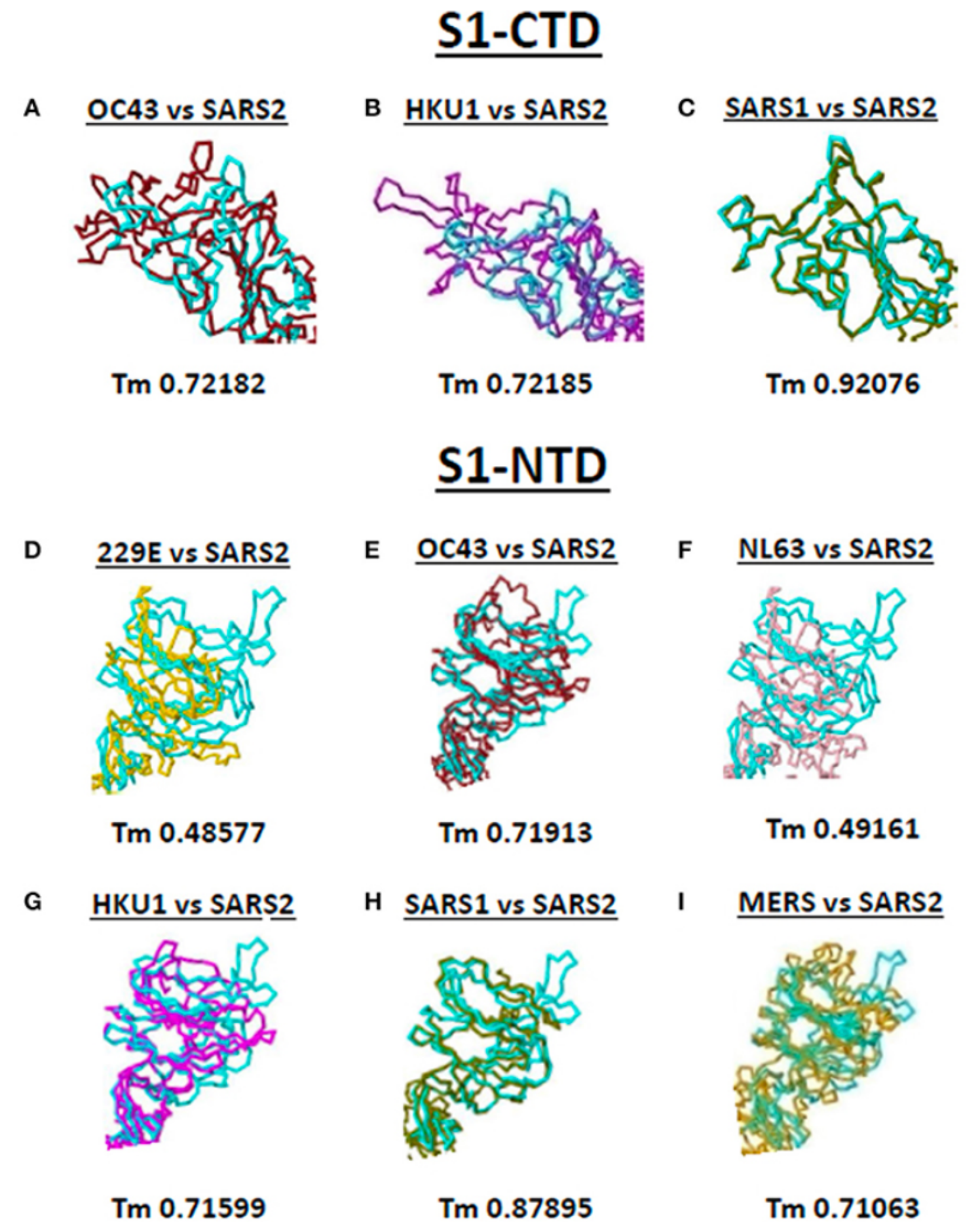
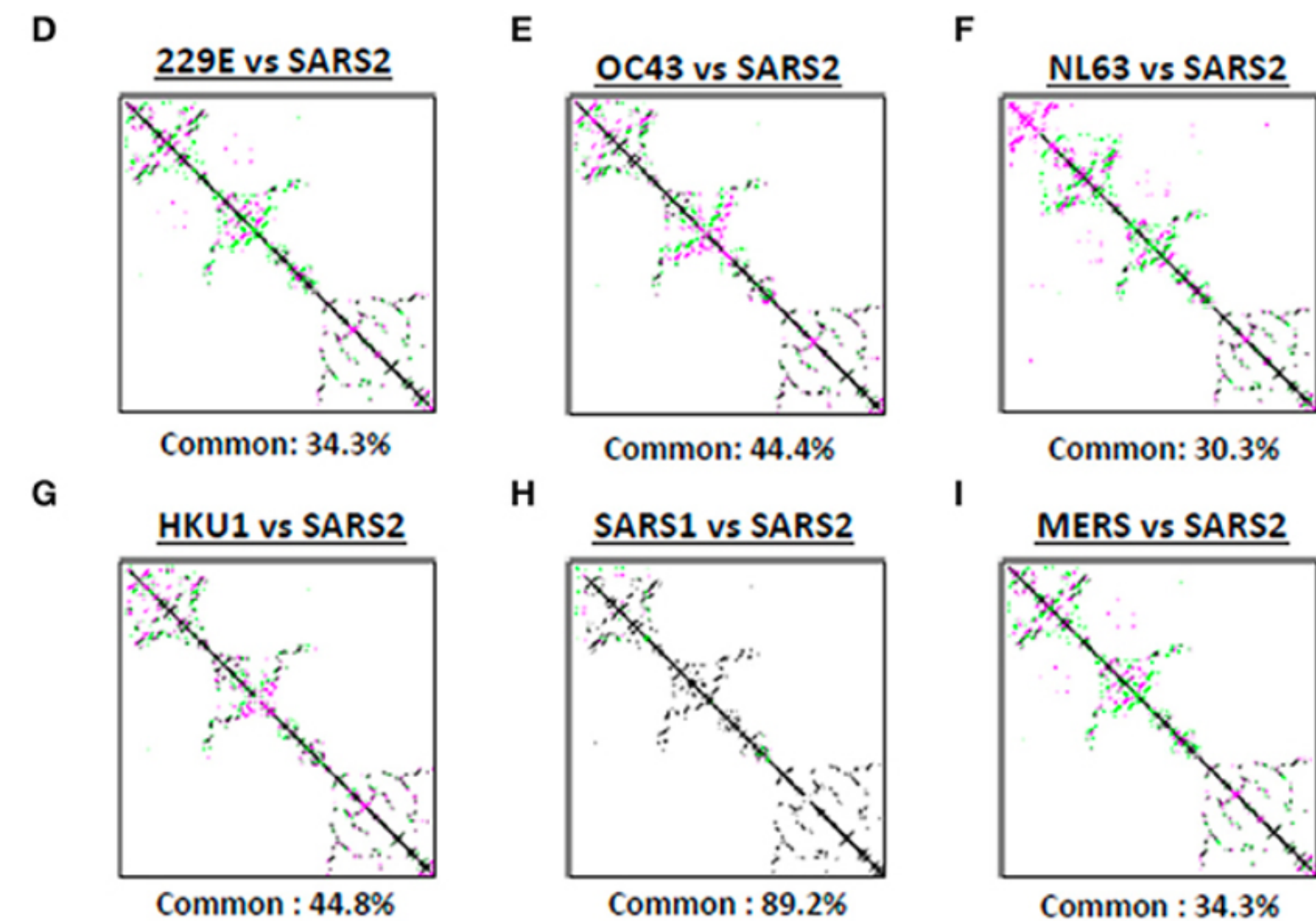
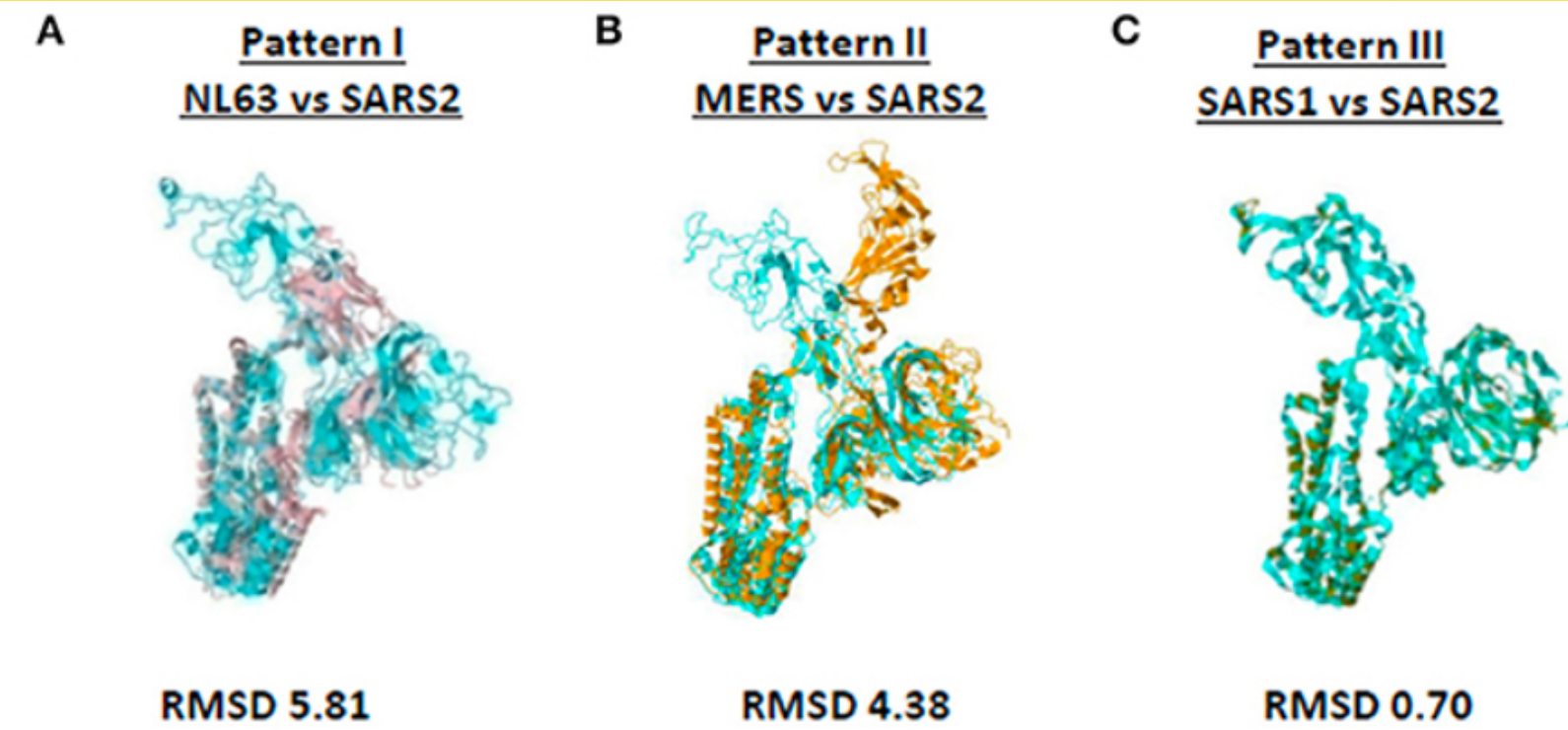
The first four routes are relatively easy to detect, but they account for <20% of all reported LAIs (23,24). No distinguishable exposure events were identified in approximately 80% of LAIs reported before 1978 (24–26). In many cases, the only association was that the infected person worked with a microbiological agent or was in the vicinity of a person handling a microbiological agent. The inability to identify a specific event was also reported in a more recent study (27), which found that the probable sources of LAIs were apparent in only 50% of cases. These data suggest that unsuspected infectious aerosols can play a large role in LAIs (1,23,24,28).”

Loose Ends - SARS and SARS2 Spike 3D Modeling

SARS1 and SARS2 Spike Models Are Structurally Similar

Among the seven human-infecting CoVs, only SARS2 resulted to a pandemic (18, 19) which may suggest that the overall SARS2 spike protein differs from the other human-infecting CoVs. To structurally differentiate SARS2 and other human-infecting CoV spike models, model superimposition was performed. As a follow-up from our earlier results (Figure 1), we utilized representative spike models (NL63, MERS, SARS1) for superimposition against SARS2 since these models putatively share different S1-CTD orientation (based on visual observation) and have both RMSD <1.0 and Tm score > 0.95 among spike models within the same phylogenetic cluster and lineage. For purposes of this study, we classified distinct S1-CTD orientations as patterns and, likewise, established which among the spike protein models share the same S1-CTD orientation, whereby, spike protein models with the same S1-CTD orientation would be classified into one pattern. In this regard, we observed three distinct S1-CTD orientations which we classified into three patterns among the superimposed spike models: (1) Pattern I (NL63 and SARS2 superimposition; Figure 3A); (2) Pattern II (MERS and SARS2 superimposition; Figure 3B), and (3) Pattern III (SARS1 and SARS2 superimposition; Figure 3C). This is consistent with our earlier observations (Figures 2A–G) which would further suggest that spike models within the same phylogenetic cluster and lineage share the same spike S1-CTD model orientation. In this regard, based on Figures 2A–G, we think that 229E follows a Pattern I orientation while both OC43 and HKU1 follows a Pattern III orientation.

Loose Ends - SARS and SARS2 Spike 3D Modeling



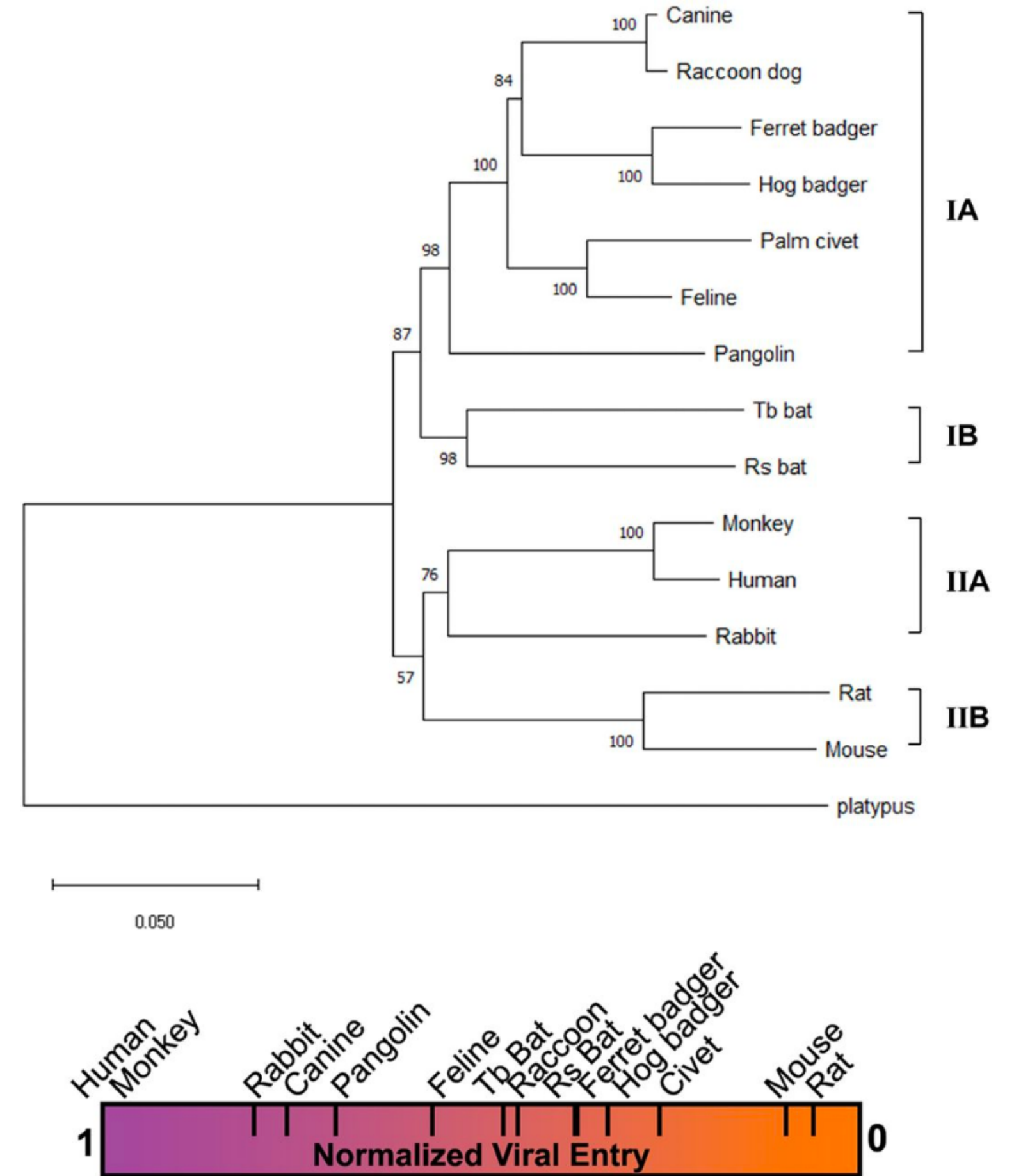
Loose Ends - in vitro testing of SARS2 binding to animal ACE2 receptors

Peter claimed that human ACE2 affinity is not exceptional:

Peter actually showed data from [this paper](#) showing that human ACE2 has among the best binding affinities to SARS2:

FIG 4 Phylogenetic clustering of ACE2s correlates with their receptor activities. At top is a phylogenetic tree of 14 ACE2s. The tree was constructed based on nucleotide sequences using the neighbor-joining method implemented in the program MEGA X.

The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The tree was rooted by the ACE2 of platypus (*Ornithorhynchus anatinus*). The taxonomic orders into which these animals are classified are shown on the right-hand side of the tree. A heat bar summarizing the relative levels of pseudotyped virus entry supported by different animal ACE2s is shown below the tree.



Loose Ends - in vitro testing of SARS2 binding to animal ACE2 receptors

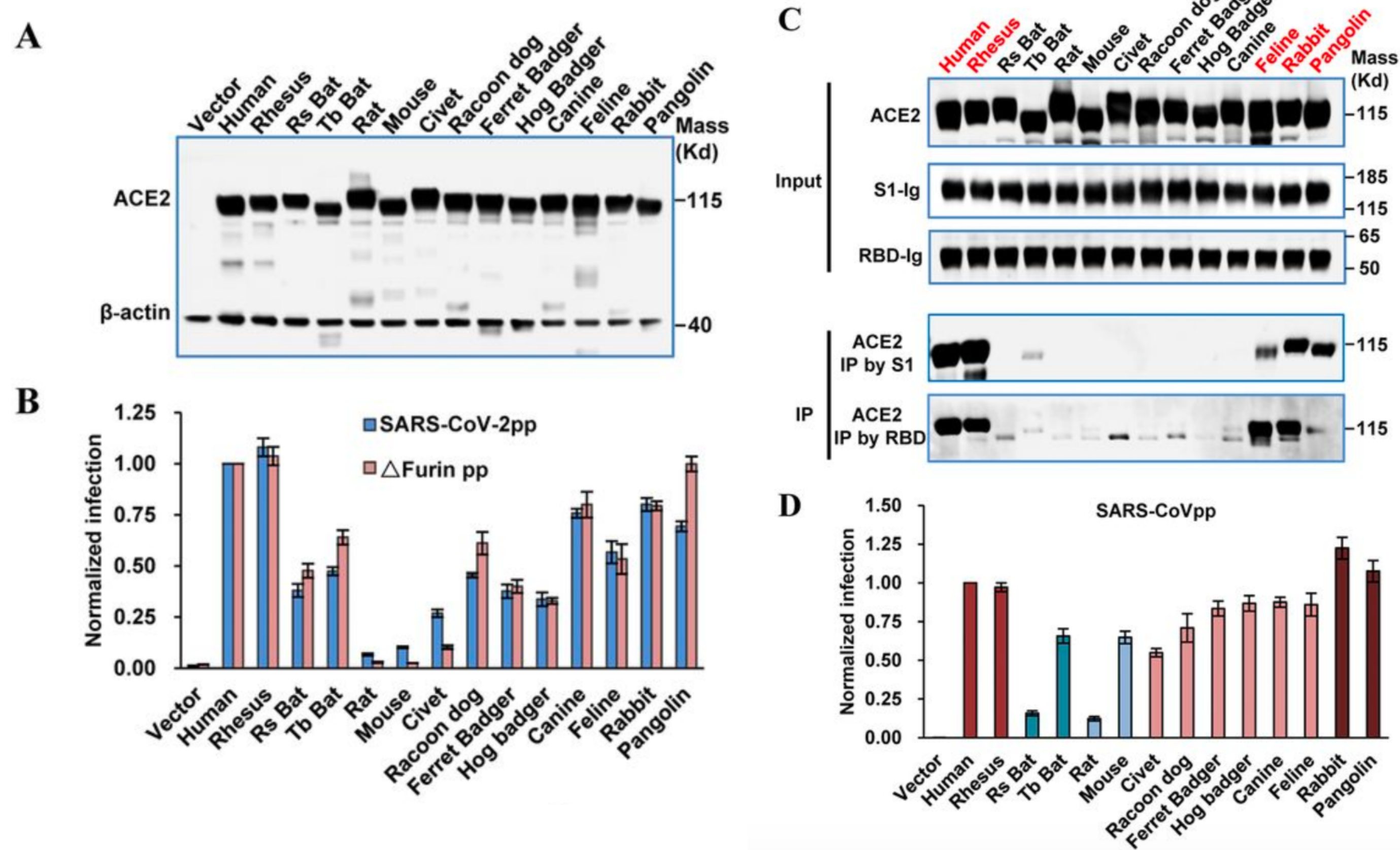


FIG 3 Multiple ACE2 orthologs served as receptors for SARS-CoV-2. (A) Transient expression of ACE2 orthologs in 293T cells. The cell lysates were detected by Western blot assay using an anti-C9 monoclonal antibody. (B) HIV-Luc-based pseudotyped virus entry. 293T cells were transfected with ACE2 orthologs. At 48 h posttransfection, the cells were infected by pseudotyped virus of wild-type SARS-CoV-2 or a mutant lacking furin (Δ Furin). At 48 h postinfection, luciferase activity was measured and normalized to that of human ACE2. Error bars represent the standard deviations of the means from four biological repeats. (C) IP assay. The upper panel shows the input of ACE2 protein with a C9 tag and S1 and RBD with an IgG Fc tag (S1-Ig or RBD-Ig). The lower panel shows the ACE2 pulled down by an S1-Ig or RBD-Ig fusion protein. (D) SARS-CoV spike-mediated entry. 293T cells were transfected with ACE2 orthologs. At 48 h posttransfection, the cells were infected by the pseudotyped virus of SARS-CoV. At 48 h postinfection, luciferase activity was measured and normalized to that of human ACE2. Error bars represent the standard deviations of the means from four biological repeats.

Loose Ends - Early A Cases

Originally the first lineage A patient was thought to have onset on Dec 1. It is the 62M who was first published in Huang 2020 as the famous Dec 1 case in the onset curve graph. Then in Zhou 2020, Shi Zhengli paper had 62M as a Dec 12 onset with a note that he got sick twice and recovered with antibiotics.

Then, in the WHO report, they moved him to Dec 26 which is the day he was hospitalized, which is clearly too late to be the true onset. So, we have no official onset date for 62M. However, if he got sick twice, was it Dec 1 or 12? Common sense has his onset probably +/- 3 days from his wife who is firm for Dec 15 onset and very likely had the same lineage. See more details [here](#)

Loose Ends - Early A Cases

From: [A pneumonia outbreak associated with a new coronavirus of probable bat origin](#)

Patient No.	Gender	Age	Date of Onset	Date of Admission	Symptoms When Admitted	Current Status (2020.01.13)	Diagnosis history
ICU-01*	Male	62	2019.12.12	2019.12.27	fever	recover, discharged	negative
ICU-04	Male	32	2019.12.19	2019.12.29	fever, cough, dyspnea	fever, intermittent cough	negative
ICU-05	Male	40	2019.12.17	2019.12.27	fever (38 °C), expectoration, malaise, dyspnea	fever, malaise, intermittent cough	AdV (IgM)
ICU-06	Female	49	2019.12.23	2019.12.27	fever (37.9 °C), palpitation	fever, malaise, cough	Coronavirus (nt)
ICU-08	Female	52	2019.12.22	2019.12.29	fever (38.5 °C), expectoration, malaise, dyspnea	recover, discharged	Streptococcus pneumoniae (nt)
ICU-09	Male	40	2019.12.22	2019.12.28	fever (38.5 °C), expectoration	fever (38.5 °C), malaise, expectoration, dizziness	negative
ICU-10	Male	56	2019.12.20	2019.12.20	fever, dyspnea, chest tightness	fever, malaise, cough, dyspnea	negative

Note, some records are missing. All patients are sellers or deliverymen at the seafood market except ICU-01, whose contact history is unclear. All patients were admitted to intensive care unit (ICU) during the first investigation and were now in stable condition. Blood IgM tests have been performed for the following respiratory pathogens for all patients: *Legionella pneumophila*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, respiratory syncytial virus, adenovirus, *Rickettsia*, influenza A virus, influenza B virus and parainfluenza virus.

*This patient reported fever on 12 December 2019 and then recovered without medical treatment. He came back to the hospital on 27 December 2019 with a fever. His wife was also ill and admitted to the hospital. Both individuals recovered.

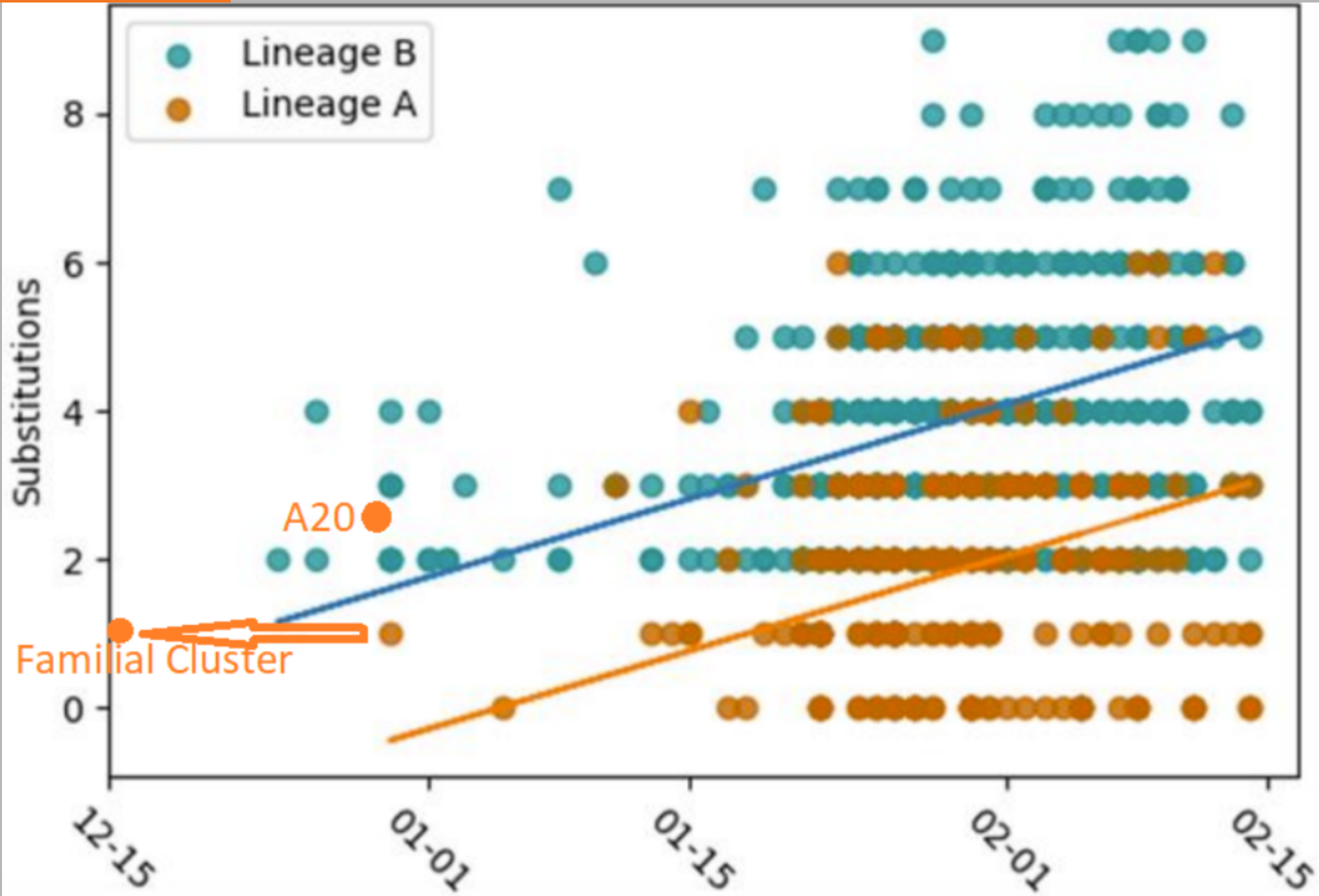
Loose Ends - Early A Cases

Additionally, the sequenced A20 sample from HSM had at least 2 extra mutations (G26262T, C6145T) and possibly 3 (T24979C), according to the Crits-Christoph preprint (Table S1).

REGION	POS	REF	ALT	REF_DP	REF_RV	REF_QUAL	ALT_DP	ALT_RV	ALT_QUAL	ALT_FREQ	TOTAL_DP	PVAL
NC_045512.2	28144	T	C	54	6	18	49136	8558	35	0.998192	49225	0
NC_045512.2	26262	G	T	1	1	37	15103	13131	35	0.99578	15167	0
NC_045512.2	8782	C	T	155	150	22	38766	35780	30	0.977631	39653	0
NC_045512.2	6145	C	T	3893	3061	33	13049	10386	35	0.768221	16986	0
NC_045512.2	16066	G	A	7	5	31	10	0	32	0.588235	17	0.00021011
NC_045512.2	24979	T	C	8254	517	35	10326	713	34	0.55552	18588	0
NC_045512.2	24143	T	C	2288	2129	34	703	652	35	0.231098	3042	2.2481E-228
For field descriptions, see the iVar documentation: https://andersen-lab.github.io/ivar/html/manualpage.html												
Cutoffs: >15x coverage, >20% alternative allele frequency												

Loose Ends - Early A Cases

Shown on the early A and B case chart



Loose Ends - Responding to claim that SARS1 didn't come from Yunnan

During the first debate, Peter states: “Rootclaim says [SARS] came from Yunnan. [...] The reality is that no SARS-infected people or civets were ever found in Yunnan. Where did Rootclaim get that idea? I have no idea, because Rootclaim didn’t provide a source.”

The source is a quote from an interview with Shi Zhengli published in an online Chinese newspaper

According to Zheng-li Shi, in [comments made at the time to a Chinese online newspaper](#), this mystery can be resolved:

"The Paper: Is the civet being wronged?"

Shi Zhengli: Not wronged. It is a fact that it spreads the SARS virus, it is the intermediate host, and bats are the source.

We went to a township under Kunming, Yunnan. I checked the information at that time. In 2003, there was a civet breeding farm in Kunming, but there is no more now. At that time, the country's civet cats were sold in Guangdong, mainly for food." [Google translate]

<https://www.linkedin.com/pulse/phylogeographic-mapping-newly-discovered-pinpoints-from-latham-phd/>

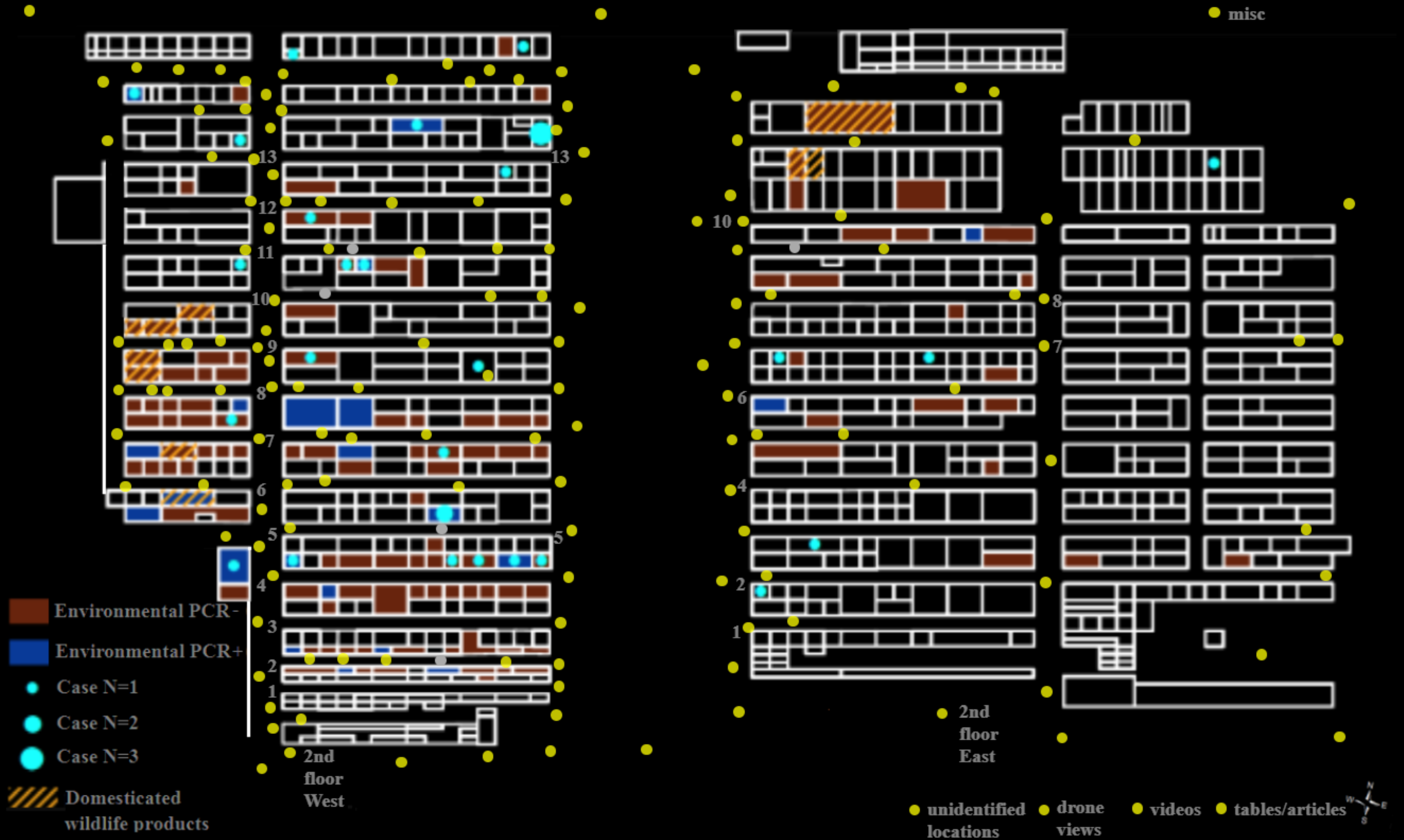
Original Chinese newspaper interview: https://www.thepaper.cn/newsDetail_forward_1897724

Higher positivity rates in environmental samples near animal stalls

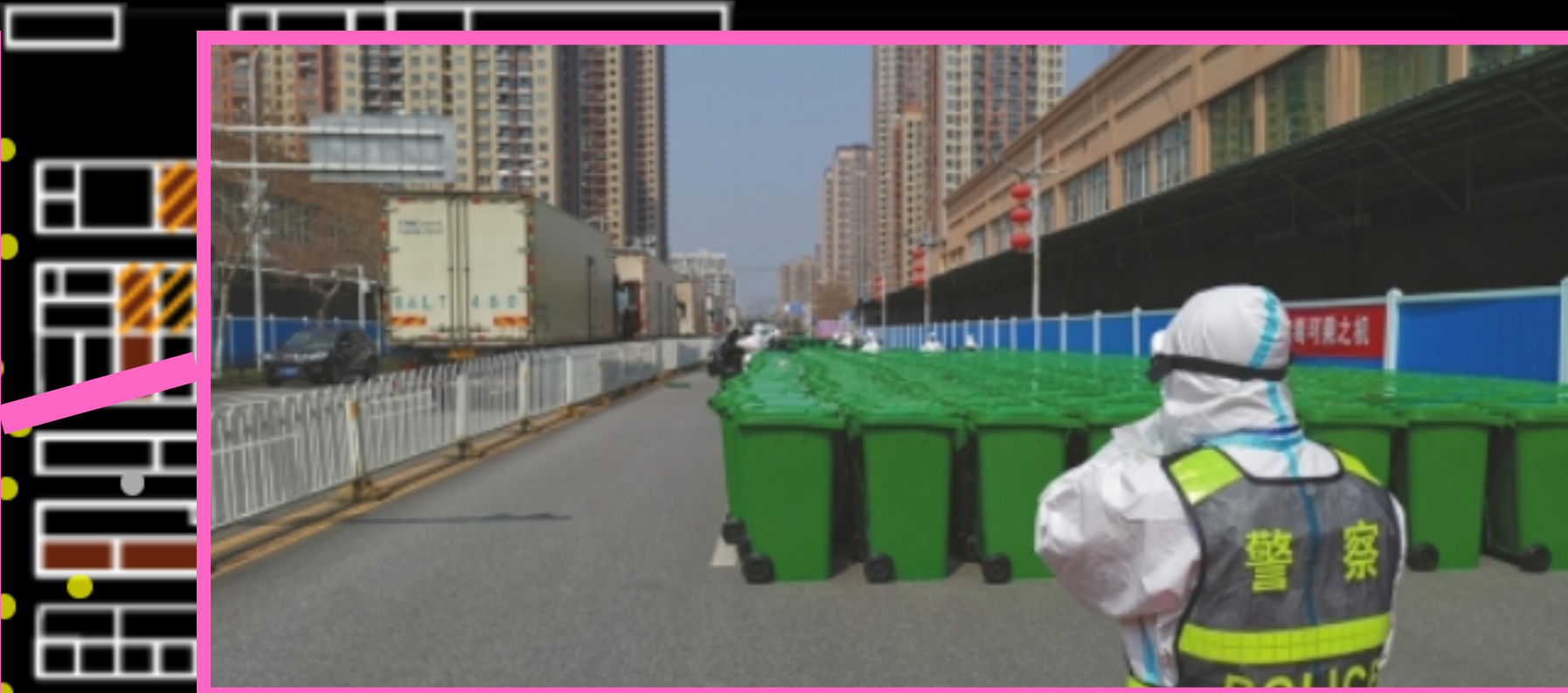
Early biases can explain the (slightly) higher positivity rate detected near animal stalls. As there was an initial expectation that the virus arrived in the market zoonotically, a focus on the live animal stalls becomes expected.

This is in addition to the stalls happening to be near the toilets and mahjong room, as well as in the least ventilated area, all of which are not necessarily a coincidence as all of these could be features of being in the “cheap” area of the market.

- Possible factors biasing positivity rate in wildlife stalls:
 - Order of sampling was biased (rightfully so but still biased) on an assumption of zoonosis. Animals first, Mahjong last.
 - Sampling more meticulously
 - Sending more experienced staff to these stalls
 - Using better sampling and detection methods
 - Contamination by multiple repeat visits of testers
 - Retesting negative samples “to make sure”

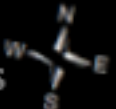


● misc



● 2nd floor East

● unidentified locations ● drone views ● videos ● tables/articles



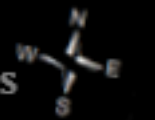


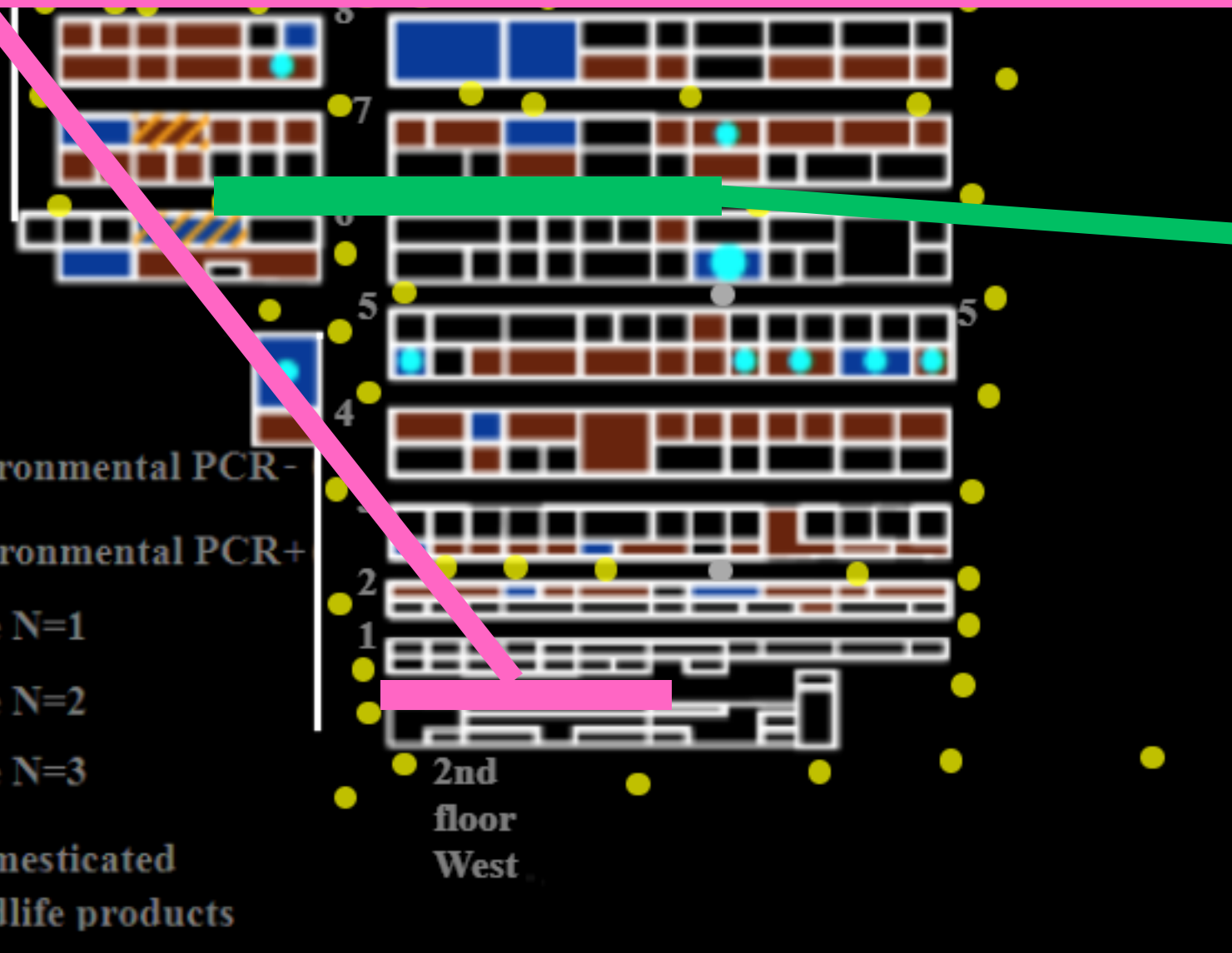
- Envir
- Envir
- Case
- Case
- Case
- Dom
- wildl

- 2nd floor East
- unidentified locations
- drone views
- videos
- tables/articles



● misc



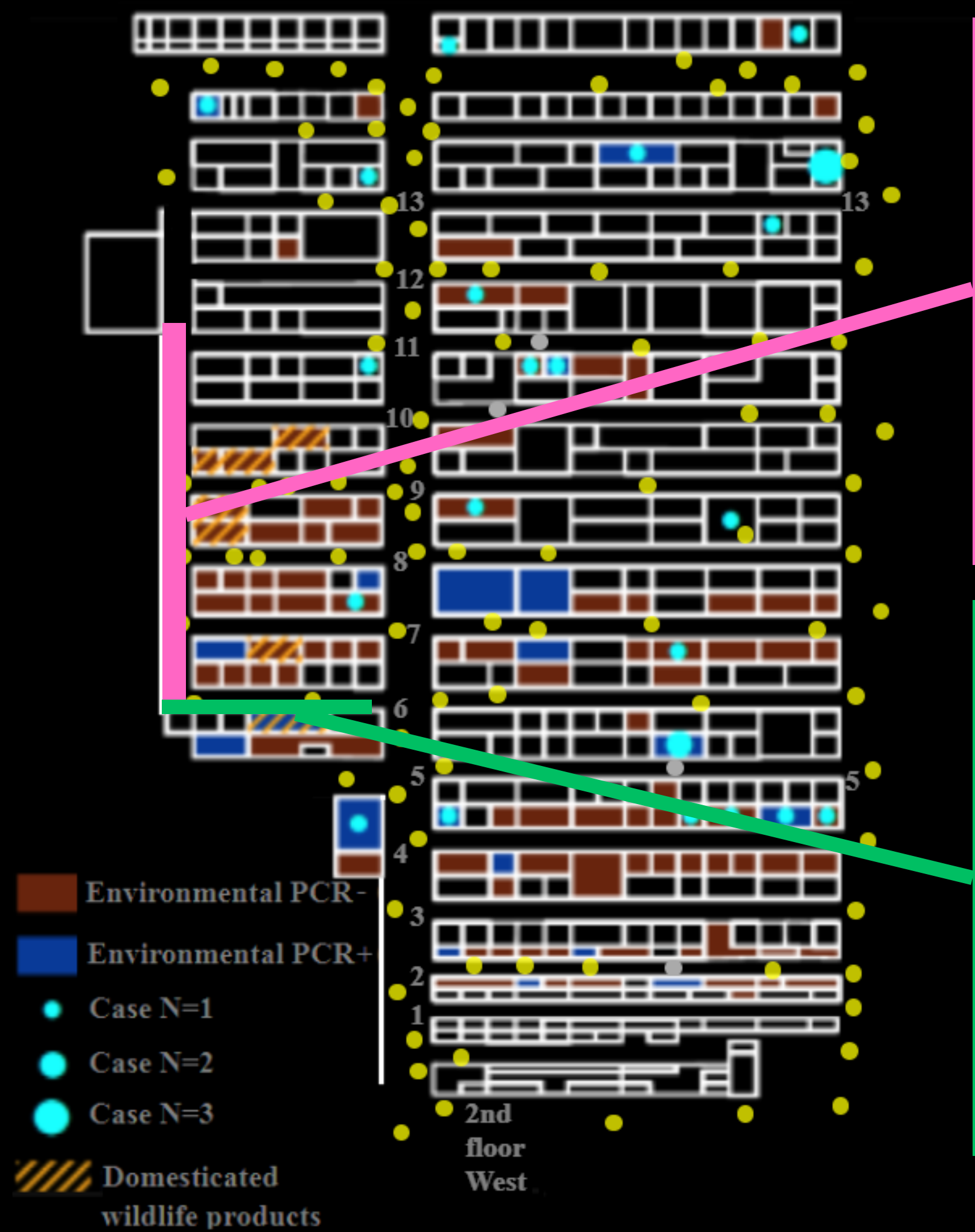


- Environmental PCR-
- Environmental PCR+
- Case N=1
- Case N=2
- Case N=3
- Domesticated wildlife products

2nd floor West

unidentified locations

● misc



Loose Ends - Smallpox leaks

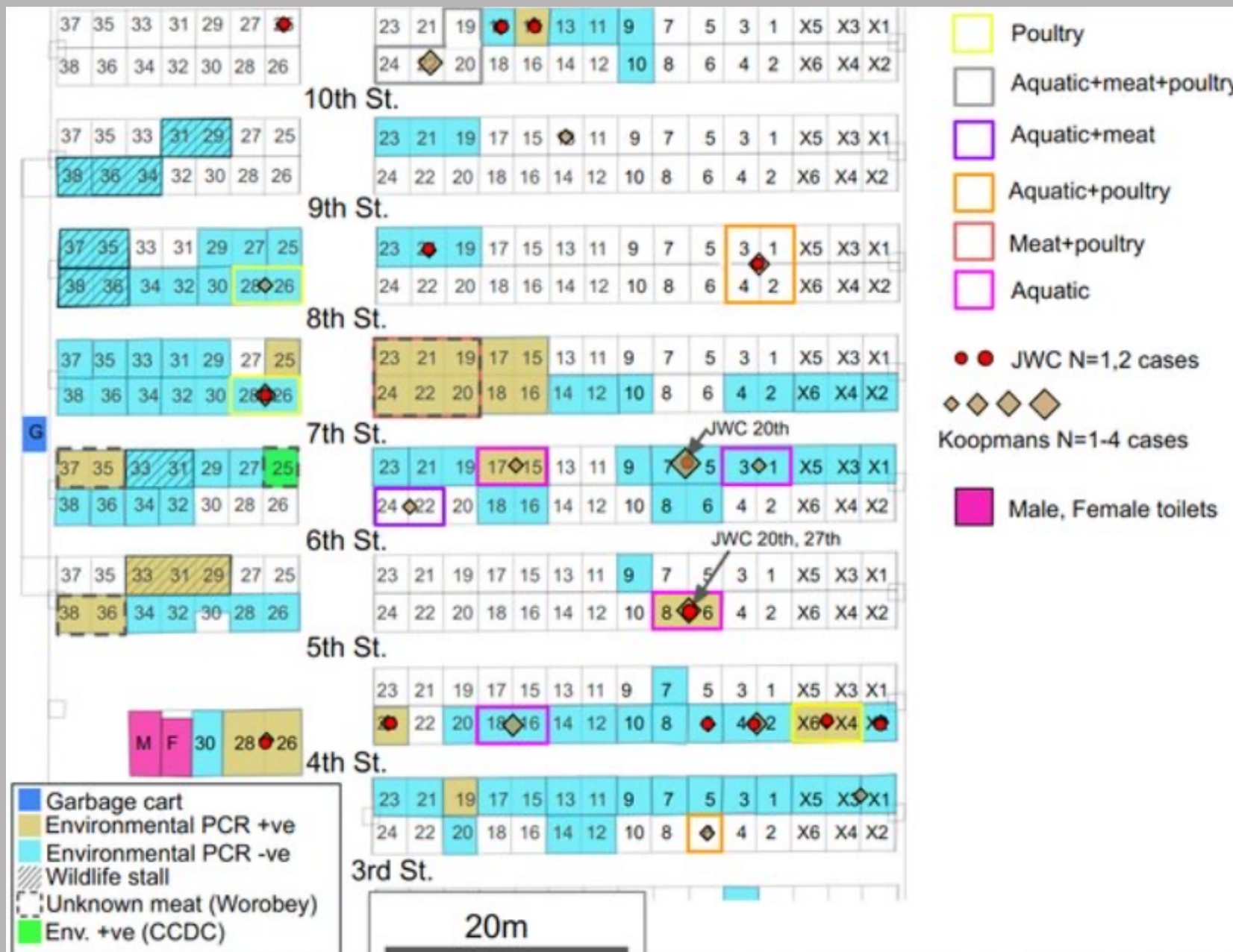
Eric asked why smallpox leaks didn't start a pandemic.

- There was only one leak causing the 1978 outbreak. 2014 was not an outbreak but discovery of unregistered vials.
- The 1978 leak didn't cause a pandemic because health authorities quarantined and vaccinated 500 people who were in contact with the index case. This was especially effective due to:
 - Limited short-range transmission: smallpox is primarily transmitted through close contact with infected individuals, and it is not as easily transmitted as respiratory viruses. The virus requires direct contact with respiratory secretions or skin lesions to spread. and does not have the ability to travel long distances through the air, limiting its spread to localized clusters of cases. This short-range transmission makes quarantine easier and more effective.
 - Vaccination: Smallpox was declared eradicated by the WHO in 1980 following a successful global vaccination campaign. By the time of its eradication, a large proportion of the world's population had been vaccinated against smallpox, providing herd immunity

In contrast, for SARS-2 the index case was likely not even aware, and no immunity existed.

Loose Ends - Expected Reads Near Infected Animals

The issue of low reads in animal stalls, raises the question of how many reads should we expect? Let's examine stalls with a sick person (red circles). That's at least what we should get on a cage of a sick animal, which can't move, and the cage is a great surface that isn't cleaned.



And indeed:

- The environmental sample with the highest number of SARS2 reads (183340) from non-cultured samples comes from stall West|11|15 which had a sick person according to the WHO map.
- Next highest (42766), is stall West|2|14, which is cut from the map.
- Next highest (3235 and another sample with 137 reads as per Crits-Christoph preprint) is stall West|4|X4-X6 with a sick person.
- Next highest (689 reads as per Crits-Christoph preprint) comes from stall West|5|6-8 with a sick person

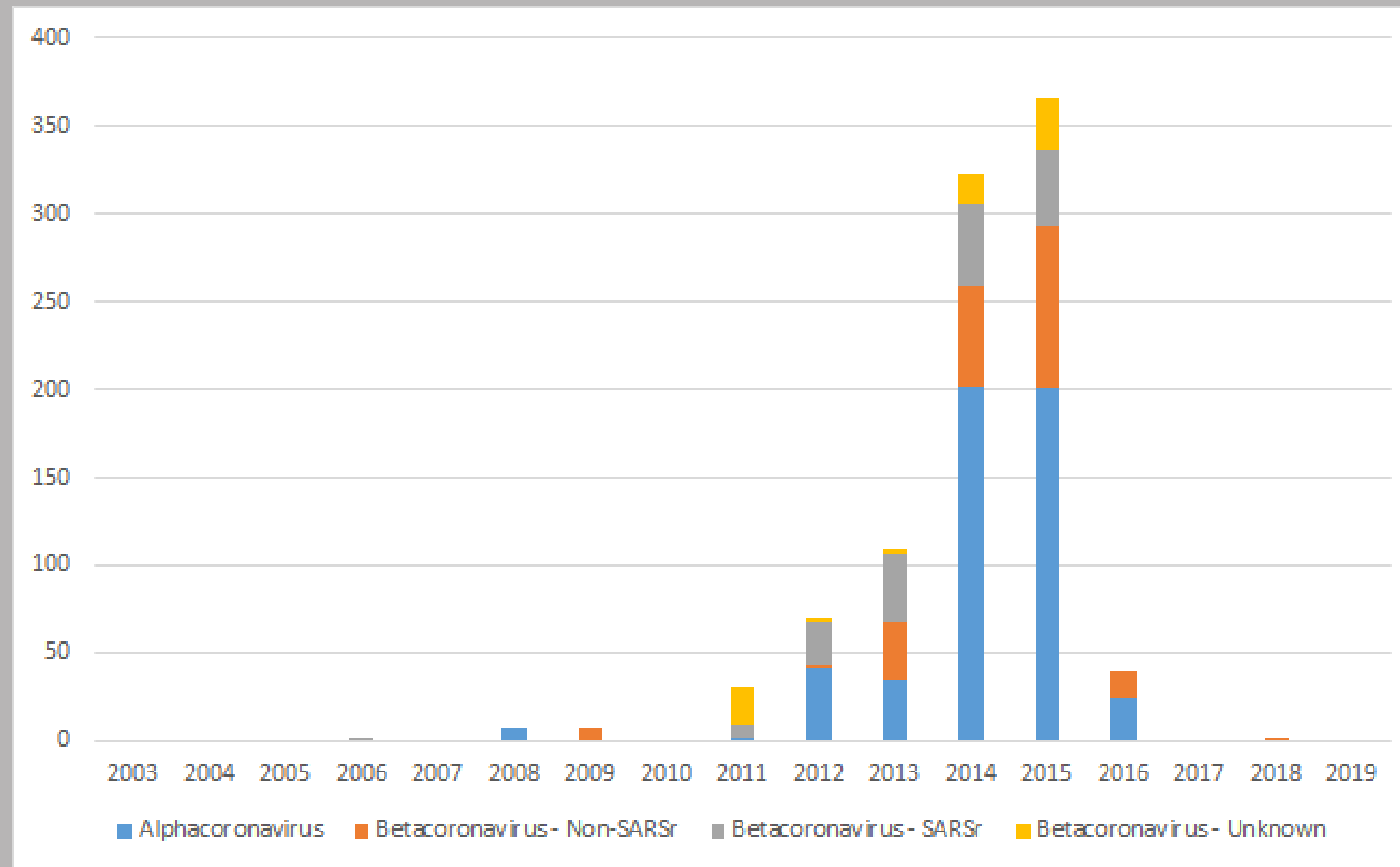
Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

- Previously, there were long delays between collection and publication. it's not “secret viruses”, just those that are still being research and not yet made public.
- The 2020 Latinne paper only published the viruses that WIV had collected prior to 2016: <https://twitter.com/franciscodeasis/status/1336491158680297477>
- RaTG13 was collected 2013 and published under embargo 2018, and released 2020.
- After that many viruses have been collected by Libiao Zhang of GIABR with whom WIV collaborated: <https://twitter.com/franciscodeasis/status/1418952588662292482>

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019



Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

In 2018, DEFUSE explicitly mentions that they wanted to run a full inventory of viruses in Yunnan caves. It is highly plausible that WIV (and GIABR) have been constantly sampling for new viruses

“Our strategy begins by a complete inventory of bats and their SARSr-CoVs at our intervention test site cave complex in Yunnan, China that harbors bats with high-risk SARSr-CoVs. We will collect data from three caves in that system (one is our intervention test site and two control sites) on monthly bat abundance and diversity, viral prevalence and diversity, individual bat viral load and host physiological markers; genomic characterization of low- and high-risk SARSr-CoV strains among bat species, sexes, and age classes; satellite telemetry and mark-recapture data on bat home range and inter-cave movement; and monitoring of daily, weekly and seasonal changes in bat populations.”

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

“To characterize spillover risk of SARSr-CoV quasispecies (QS), the Wuhan Institute of Virology team (WIV) will test bat fecal, oral, and blood samples for SARSr-CoVs by PCR. We will collect viral load data from fresh fecal pellets. SARSr-CoV spike proteins will be sequenced, viral recombination events identified, and isolates used to identify strains that can replicate in human cells. The Univ. N. Carolina (UNC) team will reverse-engineer spike proteins of a large sample of high- and low-risk viruses for further characterization. This will effectively freeze the QS we analyze at t=0. These QS_0 strain viral spike glycoproteins will be synthesized, and those binding to human cell receptor ACE2 will be inserted into SARSr-CoV backbones (non-DURC, non-GoF), and inoculated into humanized mice to assess capacity to cause SARS-like disease, efficacy of monoclonal therapies, the inhibitor GS-5734 or vaccines against SARS-CoV.”

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

PI-TA-01 Task 1: Conduct longitudinal bat sampling and ecological data collection from field sites in Southern China to obtain data for experimental studies and modeling (EHA)

Sub-Task 1.1: Apply for and obtain IACUC and ACURO approval and appropriate permits in China for bat-sample collection and field intervention pilot (EHA).

Sub-Task 1.2: Collect monthly specimens from bats at cave sites in Yunnan, China for SARSr-CoV screening and sequencing. Oral, fecal, and blood sample collected from 360 Rhinolophus spp. bats per month using live-capture and non-invasive sampling.

Deliverables: Specimens-from 3,240 bats and fecal pellets collected from high-risk reservoir populations which have been obtained with all proper permits and permissions and shipped to WIV for analysis; real-time telemetry and mark-recapture data uploaded and made available to DARPA collaborators; completed database maintained.

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

They planned to take multiple samples per bat

We will capture Rhinolophus spp. bats using harp traps and mist nets during evening flyout, collect rectal, oral, and whole blood samples (x2 per bat)

In the past, 10,000 samples generated 180 SARS-like viruses

For the past 14 years, our team has conducted CoV surveillance in bat populations across S. China, resulting in >180 unique SARSr-CoVs in ~10,000 samples.

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

SARS-like virus prevalence data from DEFUSE:

Species	n	SARSr-CoV prevalence
<i>Rhinolophus sinicus</i>	1036	10.9%
<i>R. ferrumequinum</i>	191	6.3%
<i>R. affinius</i>	518	1.2%

In Phase I we will sample 60 bats each of *R. sinicus*, *R. ferrumequinum*, and *R. affinius*, (180 bats per cave) every three months non-destructively for 18 months from our three cave sites. Given ~6-9% prevalence (n=3,304) of SARSr-CoVs in *Rhinolophus* spp. at our sites, this sample size would allow detection of 10% fluctuation in viral prevalence among sampling periods and caves. During the 2 months per quarter without physical bat trapping we will collect fresh fecal pellets by placing clean 2m² polyethylene sheets beneath roosting bats³⁷. *Rhinolophus* spp. have a 7-week gestation period,

Loose Ends - WIV Virus Collection

A repeating claim is that WIV has published its viruses and was unlikely to have many more in 2019

Based on 6 samples per bat (rectal, oral and whole blood x2), and past statistics, this project should yield hundreds of SARS-like CoVs.

This is without counting samples from fecal pellets, and other collection efforts unrelated to DEFUSE

Regardless of the specific statistics, DEFUSE was a project intended to find in nature viruses with pandemic potential. Obviously, their efforts should match that goal and they should be likely to find a few.

Updating our Market Model / HSM is a likely first cluster location

HSM is a likely first cluster or superspreading location, even if SARS2 did not cross into humans zoonotically

First, the market is larger than usually claimed

Check-ins to a seafood market are not really relevant

Stats: 50,000 sq m,
1000+ tenants

"The 40,000-square-meter South China Seafood Wholesale Market is the largest seafood wholesale market in Central China and can accommodate more than 400 operators"

"The market is the biggest wholesale outlet of its kind in central China and sells vegetables, fruit, seafood and agricultural by-products"

Traffic, Location, and Permanent Residents

Traffic and Location

HSM is located in a high-density neighbourhood and receives high traffic (10,000/d)

Having high traffic from a high-density neighbourhood means a higher chance that someone infected will come in the early days

Many permanent residents (1000)

Allows for the virus to stay and form a cluster (unlike public transport)

Environmental factors : Temperature

Temperature

The virus's stability decreases as temperature increases; it has a half-life of about 1.7 to 2.7 days at a moderate temperature of 20°C.

SARS-CoV-2 remains viable for longer periods at cooler temperatures, lasting up to 14 days at 4°C in a nutrient medium.

At a higher temperature of 40°C, the virus's infectivity drops to just a few hours on common surfaces.

At freezing temperatures like those used for cold-chain storage (below -18°C), the virus can remain infectious for at least 60 days, demonstrating greater stability and infectivity at colder temperatures.

Environmental factors : Indoor / no UV exposure

An indoor market would not benefit from the mitigating effects of UV radiation and good ventilation on Covid-19 growth

No UV radiation

Ultraviolet (UV) radiation has been shown to limit Covid-19 growth rates and is a favourite among disinfection systems against COVID-19 outbreaks.

A study found that UV radiation has a statistically significant effect on daily COVID-19 growth rates: a SD increase in UV lowers the daily growth rate of COVID-19 cases by 1% over the subsequent 2.5 wk

WHO joint-mission wrote :

“Apart from the central street, the alleys of the market are very narrow, dark and poorly ventilated, with a roof covering the whole wing”

Environmental factors : Indoor / poor ventilation

Indoor Market : Poor Ventilation

Ventilation would also be lower than in an outdoor location

Even among indoor locations, a market such as HSM stands out in terms of poor ventilation compared to a more modern building such as a shopping centre, which is likely to be much better ventilated.

The WHO joint-mission confirmed that the ventilation system had not been used since the outbreak of avian influenza in 2013 :
“The ventilation system had been closed since the live poultry trade had been stopped following the outbreak of avian influenza.”

Environmental factors : Hygiene

Hygiene measures in wet markets do not correspond to the level of cleaning and disinfection found in other possible places of transmission (universities, shopping centres...):

Poor hygiene is a major risk factor for human health in wet markets.

Poor hygiene practices have been linked to bacterial and parasitic infections in wet markets, typically due to improper hygienic routines of cutting boards, contaminated water sources, etc

Specifically for HSM, the WHO joint-mission confirmed that:

“The “sewerage” was an elaborate semi-open drainage system in poor condition running through the entire market”

Environmental factors : Organic material (1/2)

Organic matter is widespread in wet markets. This further increases the viability and transmission of the Covid-19 virus through the following channels:

Protection

Organic material can provide a protective environment for the virus. By encapsulating viral particles, shielding them from environmental factors like UV radiation and desiccation (drying out), which can reduce the virus's viability.

“Many viruses can be stabilised and protected by the surrounding organic material, such as in saliva or mucus droplets. It has been reported that the presence of substances such as bacteria, fats, proteins in the viral inoculum can additionally increase the persistence”

Environmental factors : Organic material (2/2)

Extended Survival on Surfaces

Organic material on surfaces, such as contaminated respiratory droplets or saliva, can help the virus remain viable for longer periods. The virus can bind to these materials, allowing it to persist on surfaces and potentially remain infectious if not properly cleaned and disinfected.

Increased Adherence

Organic material can enhance the adhesion of viral particles to surfaces, making it more likely for the virus to stick to objects, as well as mucous membranes in the respiratory and digestive tracts of individuals who come into contact with contaminated surfaces or aerosols.

Environmental factors : Conclusion

A number of environmental factors significantly increase the likelihood of the HSM being the site of a superspreader event compared to other busy locations

Low temperatures (and perhaps humidity) increase the lifespan of the virus

HSM's cleaning and disinfection measures are well below what is done in other frequently visited areas of Wuhan such as shopping centres and universities

An indoor market means that ventilation within the market will be poor and the natural protection afforded by exposure to UV light will be largely absent

A wet market contains many sources of organic materials that protect viral particles and give them greater adherence to fomites

Specific conditions in HSM : Mahjong halls

This study found Mahjong halls to be a major COVID-19 transmission vector

High Proportion of Total Cases

In Yangzhou, a mahjong house was the source of the largest transmission chain, with 202 cases representing 26.8% of all cases during the study period.

Demographic Concentration

The transmission chain mainly affected middle-aged and elderly individuals, who are more likely to experience severe illness and represent the majority of the players in a mahjong house, increasing their exposure risk.

Superspreading Events

The initial case in this chain was reported on July 21, with the majority of subsequent cases occurring between July 27 and August 9. This period included three distinct superspreading events, indicating that the environment facilitated multiple, intense outbreaks.

Superspreading Potential

A subgroup analysis revealed that cases in mahjong houses were disproportionately likely to lead to further spread, with 5% of cases responsible for 80% of the total transmission

These details highlight the unique combination of factors in mahjong houses—such as close contact, indoor settings, the demographics of players, and the likelihood of extended stays—that create an environment highly conducive to the spread of COVID-19.

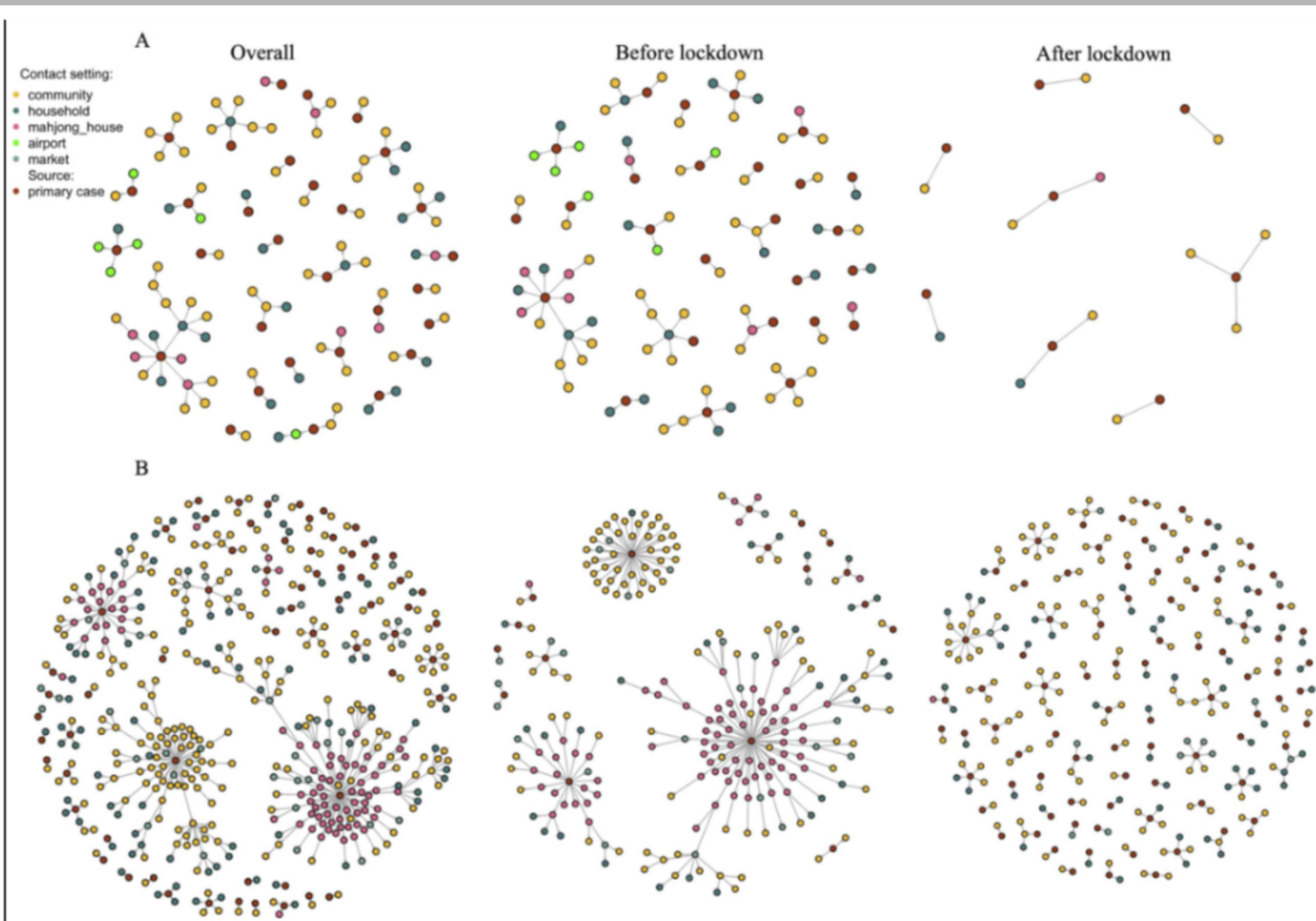
Specific conditions in HSM : Mahjong halls

In subgroup analysis, we found that cases in mahjong houses had substantial potential of superspreading, with 5% of cases seeding 80% of the total transmission (Table 2 and Fig. S5 in Appendix). Moreover, infectors aged more than 65 had relatively higher transmissibility, with a higher estimated R value compared to other age groups; infectors aged between 0 and 17 had sizable superspreading potential, with 8% of cases generating 80% of all transmission (Table 2 and Fig. S5 in Appendix).

Table 2. Estimated reproductive number (R) and dispersion parameter (k) of the negative binomial distributions, and inferred superspreading potential of Delta variants stratified by different contact settings, and age groups of seed cases.

	R (95% CI)	k (95% CI)	* Prop80% (95% CI)	** Prop0% (95% CI)
Contact Settings				
Household (n = 108)	0.58 (0.47–0.70)	13.42 (1.88-Inf)	32% (25%–36%)	57% (50%–66%)
Community (n = 266)	0.71 (0.59–0.84)	0.53 (0.37–0.76)	22% (18%–26%)	64% (57%–70%)
Mahjong house (n = 91)	0.82 (0.29–1.82)	0.05 (0.03–0.10)	5% (3%–9%)	86% (74%–94%)
Market (n = 38)	0.59 (0.39–0.81)	1.92 (0.58-Inf)	29% (18%–39%)	60% (44%–74%)

Specific conditions in HSM : Mahjong halls



Reminder: Mahjong Hall Early Cases Evidence

Testimonies from 8 people, collected by Chinese & Western media, include people working in HSM, a Wuhan journalist and the head of the respiratory department at Jinyintan hospital. Together, they say more than 35 early HSM cases were among Mahjong players.

“had heard about [..] some mah-jongg infections” source

“I heard that four people were playing at one table, and all four got sick” source

“At the end of December, we had already judged that this infectious virus was related to the South China Seafood Market because most of their patients knew each other. [..] They are either a family or people who play mahjong together.” source

“ most of the 27 infected people first reported by the government[56] had a common hobby - they like to play mahjong.” source

“Our boss was sick on December 19 (2019), sick before playing mahjong with a player who sells frozen chicken and frozen duck, who was also admitted to hospital for pneumonia.” source

“a friend of mine is in the hospital with it. [..] He thinks he might have got it while playing mahjong, reckons the air in the room was stuffy, a lot of people in there.” source

“The owner of the shop was a man in his 40s and 50s who usually likes to stay up late and play mahjong. Recently, he “didn't care” after catching a cold. He didn't go to the doctor at first. After a few days, he found that his condition was getting worse before going to the hospital. He was eventually diagnosed with pneumonia.” source

“heard on the morning of December 31 that three store owners who often played mahjong at stalls near 10th Street were hospitalized two days ago. He sent WeChat greetings to one of the bosses, who replied “infected with the virus”.” source

Examining the claim of many SSE Locations other than HSM

In the next slides we will look at the data set of 430 potential SSE locations throughout Wuhan provided by Weibo check-in data, as set out by the Worobey paper.

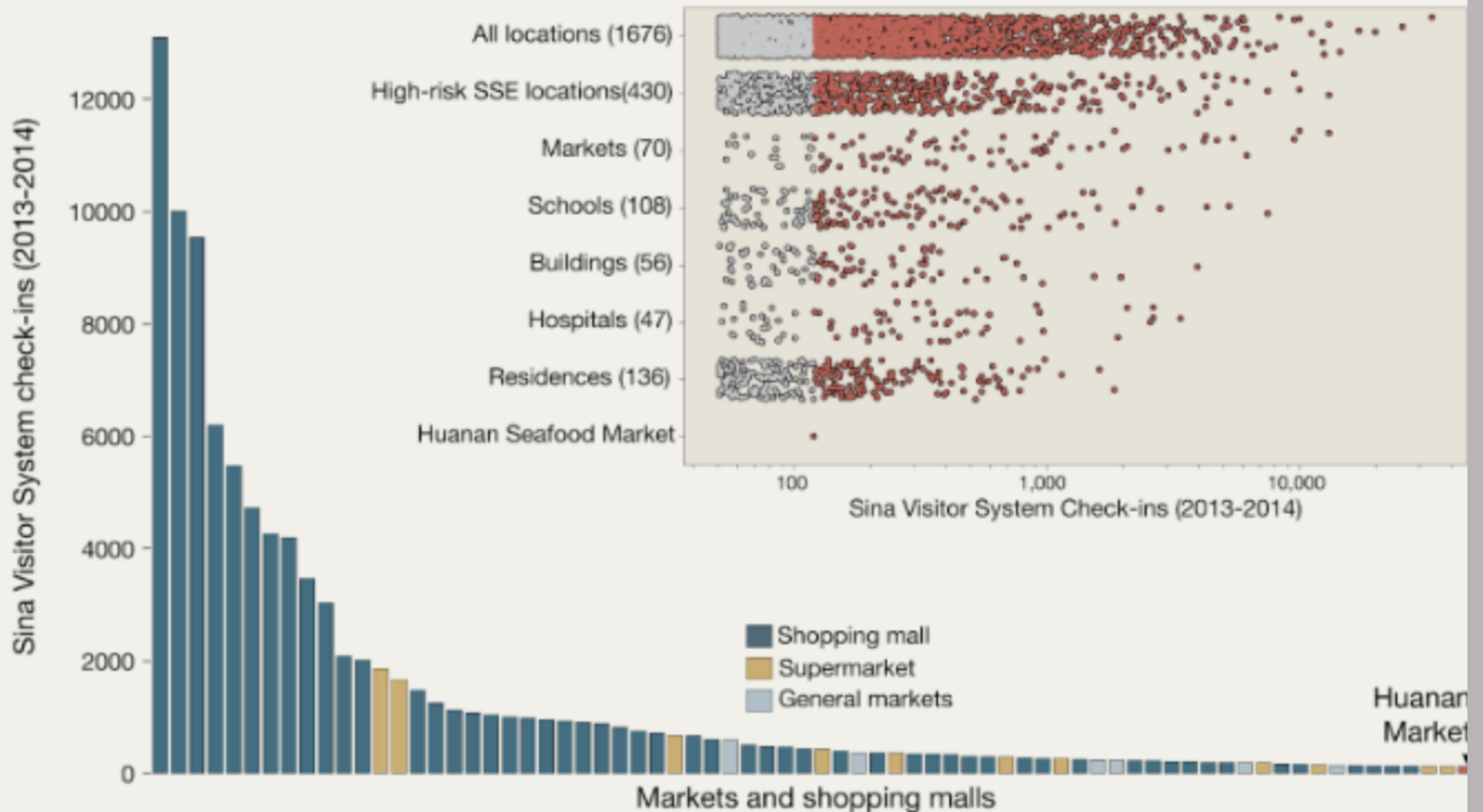
We will first look at the categories which are not 'market' and use criteria to explain why each of these categories do not meet the conditions for a likely early cluster.

Then we will examine the "markets" category, where we will show that most entries in it are not, in fact, markets, and therefore also do not provide the same fertile conditions as Huanan.

Lastly, we will search for other markets, missed by the Weibo check-in data, some of which could form an early cluster.

Thus, we show that claiming 430 more likely SSE locations than Huanan in Wuhan is grossly misleading. There are a handful at most that are comparable.

Other SSE Locations



Why are the other categories not SSE

	High traffic	Density and proximity	Permanent residents	Specific factors
Schools	Schools have a low-traffic stable population that does not change. Very few visitors.	Schools are dense environments, many classrooms with 30-40 students per classroom.	Schools have a stable population that does not change, allowing for repeated exposure.	We expect schools to have far better hygiene and ventilation, but all have classrooms, which are similar to a market in terms of proximity and person-to-person contact. Schools also have a much younger population.
“Buildings” We replace this category with an offices, skyscrapers, and hotels category. Buildings are too varied to be addressed collectively, and many locations are included in other categories anyway.	Office buildings and hotels tend to be high traffic locations.	While office buildings and hotels tend to have high populations, no single room or meeting space tends to be overly crowded.	Offices have permanent populations, hotels have staff which act as permanent populations even if the guests change over often.	These buildings likely have far better ventilation and hygiene.
Hospitals	Hospitals have a fluid population that usually experiences rapid change-over. We expect hospitals to have far fewer visitors than Huanan, however this is canceled out by the fact that many of them will be sick.	Hospitals are not usually dense for hygiene reasons.	While the patient population is fluid, there is a large permanent staff population in the hospital.	Hospitals have high hygiene and ventilation standards, and few spaces which could act similar to a market. An exception could be staff rooms, however they would adhere to stricter hygiene rules, being aware of respiratory infections.

Why are the other categories not SSE

	High traffic	Density and proximity	Permanent residents	Specific factors
Residences	Residences have a stable population that does not change. Very few visitors.	While apartment and office buildings are densely populated, the units within them are usually small and limit the amount of dense mingling of the residents.	Residences have permanent residents.	Usually well ventilated + good hygiene standards. This also has little to no weight, as the units within the buildings are not dense.
HSM				

The Buildings Category

1593	民生大厦	Minsheng Building	114.27328	30.59422	楼宇	127	Building
1303	田园商务大厦	Pastoral Business Building	114.27537	30.62019	楼宇	167	Building
1524	保利金国际广场B座	Poly Gold International Plaza B seat	114.29257	30.5761	楼宇	135	Building
1205	美奇国际	Meiqi International	114.27715	30.57292	楼宇	186	Building
1545	建银大厦	CCB Building	114.27108	30.59685	楼宇	132	Building
1437	湖北信息产业科技大厦(珞喻路)	Hubei Information Industry Technology Building (Luo Yulu)	114.3922	30.51164	楼宇	147	Building
637	武汉世界贸易大厦	Wuhan World Trade Building	114.26858	30.58027	楼宇	381	Building
641	湖北省奥林匹克体育中心办公楼	Hubei Olympic Sports Center Office Building	114.469353	30.475173	楼宇	378	Building
1215	光谷国际	Optics Valley International	114.430036	30.504692	楼宇	184	Building
1181	水果湖广场	Fruit Lake Plaza	114.34452	30.54789	楼宇	191	Building
1283	武汉天河机场综合楼	Wuhan Tianhe Airport Complex	114.220215	30.777736	楼宇	170	Building
510	友谊大厦(友谊路)	Friendship Building (Friendship Road)	114.28005	30.57607	楼宇	480	Building
1542	中百写字楼	Zhongbai Office Building	114.28957	30.58195	楼宇	132	Building
360	泰合广场	Thai Hop Square	114.26917	30.57476	楼宇	685	Building
373	楚天传媒大厦	Chutian Media Building	114.36608	30.56727	楼宇	674	Building
1340	东风汽车公司	Dongfeng Motor Corporation	114.164179	30.5019	楼宇	161	Building
434	积玉桥万达广场	Jiyu Bridge Wanda Plaza	114.310167	30.56509	楼宇	580	Building
110	武汉国际会展中心	Wuhan International Convention and Exhibition Center	114.275453	30.582323	楼宇	1971	Building
1324	武汉科技大厦	Wuhan Science and Technology Building	114.25035	30.61199	楼宇	164	Building
1590	烽火科技大厦	FiberHome Technology Building	114.39153	30.51299	楼宇	128	Building
328	中建广场	China Construction Plaza	114.33213	30.53465	楼宇	762	Building
1302	同成广场	Tongcheng Plaza	114.26565	30.59376	楼宇	167	Building
1383	宝丰时代	Baofeng Times	114.25375	30.57833	楼宇	155	Building
472	保利金国际广场A座	Poly Gold International Plaza A seat	114.29216	30.57633	楼宇	531	Building
147	新世界国贸大厦	New World International Trade Building	114.269613	30.594837	楼宇	1544	Building
1296	浦发银行大厦	Shanghai Pudong Development Bank Building	114.27457	30.59066	楼宇	168	Building

The market category

Name_POI.1	Rootclaim category	Rootclaim Description
Wuhan Wangfujing Department Store	DEPARTMENT STORE	Department store
ATM trendy department store (Simen Store)	DEPARTMENT STORE	Fashion department store
SOGO Zhuang Sheng Sogo Department Store	DEPARTMENT STORE	Department store
new world department store (Hanyang Store)	DEPARTMENT STORE	Department store
Metro	WHOLESALE	retail and wholesale food
Wuhan Economic Development Wanda	DUPLICATE	DUPLICATE
Xianglong Times Square	MALL	High end shopping mall
Walmart Supercenter (Zongguan Xihui Store)	SUPERMARKET	WALMART
new world department store (Guomao Store)	DEPARTMENT STORE	Department store
Wanda square B seat	MALL	Wanda square is shown elsewhere to be a mall
Zhongnan Commercial Building of Zhongshang Group	MALL	commercia/retail/export
new world department store (Wuchang store)	MALL	entertainment lifestyle center
Metro (Qiaokou Store)	SUPERMARKET	METRO
Zhongbai Warehouse Caidian Shopping Plaza	WHOLESALE	WAREHOUSE - similar to market
RT-Mart Supermarket	SUPERMARKET	Closed, in-door supermarket

The other 70 'markets' are not markets.



The market category



The other 70 'markets' are not markets.



future city · Euler	MALL	mixed commercial (mall) and residential
Zhongbai Warehousing (evergreen shop)	WHOLESALE	WAREHOUSE - similar to market
PARKnSHOP	MALL	Seems to be a shopping mall also under name Bajja
Wushangjian Second Shopping Mall	MALL	Shopping mall
Grass Mud Horse Building	UNKNOWN	can't find info online
Huangpi Shopping Plaza	MALL	shopping mall
Wanshang · white horse	WHOLESALE	Clothing wholesale market
Carrefour (Wusheng Road Store)	SUPERMARKET	CARREFOUR
new world department store (center store)	DEPARTMENT STORE	department store
Intime Department Store	DEPARTMENT STORE	women's clothing high end
Carrefour Hongshan Plaza Store	SUPERMARKET	CARREFOUR
Dayang Department Store (Jiangnan Road)	DEPARTMENT STORE	department store



The other 70 'markets' are not markets.



Walmart Supercenter (Xudong Pinmao Store)	SUPERMARKET	WALMART
Zhongbai Warehousing (Luoshi Road Shopping Plaza Store)	WHOLESALE	WAREHOUSE - unclear what shopping conditions are like
Wuhan Wangjiawan Moore City	MALL	shopping mall +hotel
Ocean Department Store (Wuhan Optics Valley Store)	SUPERMARKET	OCEAN
Hanshang Group twenty one shopping center (Longyang Avenue)	MALL	shopping mall
Warrare Plaza	UNKNOWN	this seems to be a middle school
Holiday Park Shopping Center	MALL	shopping mall
Yangtze River Chongwen Square	MALL	Leisure + entertainment
Carrefour (Erqidian)	SUPERMARKET	CARREFOUR
Carrefour (Shishengdian)	SUPERMARKET	CARREFOUR
Fahrenheit	DEPARTMENT STORE	department store
China Business Plaza Shopping Center	MALL	Shopping center
Carrefour (Optics Valley Store)	SUPERMARKET	CARREFOUR
Carrefour (Zhongjia Village Store)	SUPERMARKET	CARREFOUR
Zhongbaichang Tangjiadun Shopping Plaza (Tangjiadun Road)	DEPARTMENT STORE	discussed in source as central department store/mall
new world department store (Qiaokoudian)	DEPARTMENT STORE	department store
Hanzheng Street Center Mall	MALL	street mall
Wuhan Guoguang Phase II	MALL	Shopping mall
Carrefour (construction shop)	SUPERMARKET	CARREFOUR

Summary of The Contents of The Market Category

MALL	DEPARTMENT STORE	SUPERMARKET	UNKNOWN	WHOLESALE	DUPLICATE
32	16	14	2	5	1

Warehouses in the market category

We found that the category “warehouses” is most closely related to the conditions that we would find in a market, but are still far from being wet markets.

Some examples of warehouses listed:



Zhongbai Warehousing



Wuhan Baima Clothing Wholesale Market



A screenshot of the Wuhan Wholesale Network website. The page features a navigation menu at the top with links for 'Simplified C...', 'Welcome to Wuhan Wholesale Network. Please log in and register for free', 'Wuhan Whol...', 'Member Centre', 'My Order', 'Favorites', 'stall center', and 'A must-read'. Below the navigation is a search bar with the text 'Please enter the stall name/house number/abbreviation to search for the st' and a 'Search' button. There are also icons for 'Factory direct', 'Dropship', 'Worry-free', and 'Lightning delivery'. A banner below the search bar reads 'Follow the official WeChat public' and 'For more hot-selling information, search the public account on WeChat: Wuhan', accompanied by a QR code. Below the banner, the text 'Wan Shang White Horse' and 'WANSHANG·BAIMA' is displayed. At the bottom, there is a section for 'Main business: women's casual sports suits' with several empty product slots.

Wuhan Baima Clothing Wholesale Market is the largest brand clothing wholesale market in Central China and Hubei Province. The first floor is the world of boutique clothing; the second to sixth floors are the boutique women's clothing wholesale center, the seventh floor is the men's clothing and leisure franchise area, the eighth floor is the casual wear franchise area, and the ninth floor is the brand men's clothing franchise area, which gathers nearly all the famous national brands of clothing.

Hanyang Metro store

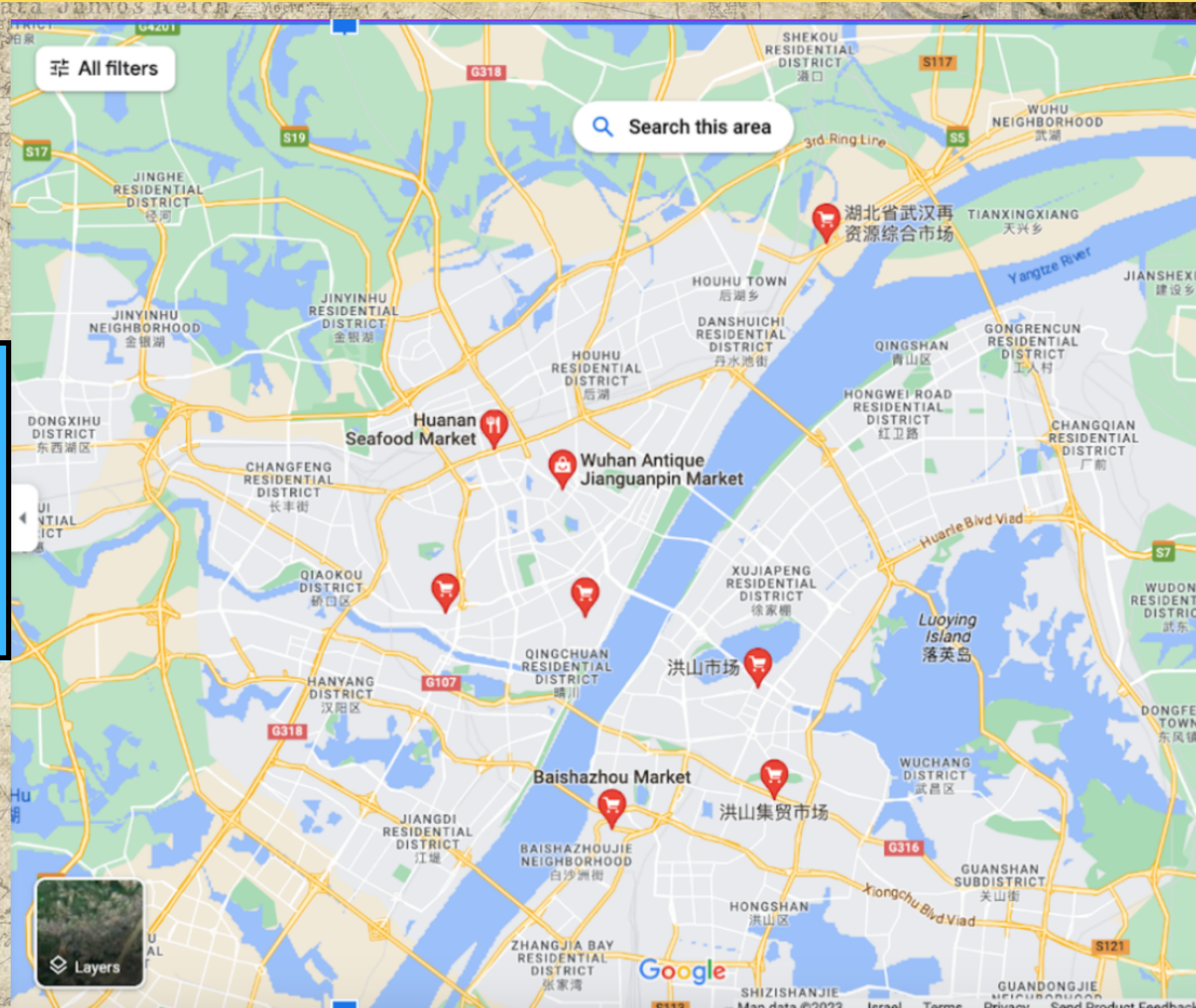


On December 20, Metro Wuhan Hanyang store and Panlong store opened simultaneously, ending 2017 perfectly. At this point, Wuhan will become the city with the most intensive distribution of Metro in Central China with a layout of four stores in one city, indicating that Metro will accelerate its in-depth expansion into Central China with Wuhan as the center and continue to improve its sales network layout.

“Markets” are far from meeting the early cluster criteria that Huanan met.



Alternative ways of comparing HSM to other markets



In a google map search of 傳統市場 (translation of “wet market”) near Wuhan, 8 results are shown:

Occasionally, this search will only yield 3 results: HSM, Baishazhou, and the recycling comprehensive market.

Huanan Seafood Market

**Wuhan Antique
Jianguanpin Market**

**Antique market - no/little
organic material**

**Jiangnan District Huanghe
vegetable market**

Labelled a vegetable market but a food safety supervision noted 7 categories of food, including meat and dairy products. It is not clear what percentage is made up of meat products.

**Hongshan
market**

Described as the “people’s vegetable market” which recently underwent an upgrade, including serious upgrades to hygiene standards. This could, of course, be propaganda but was undergone and published in 2017, such that there is little to be gained from such propaganda.

“According to reports, Jiahe Fresh Food Market has previously undergone a 2.0 upgrade. When entering the market, in addition to the regular hardware facilities upgrades... As the focus of the renovation, fly-proof curtains were installed at the entrances of public toilets in the market, “six-step hand washing methods” were posted, and hand sanitizer was provided. The market arranged dedicated personnel to be responsible for the daily cleaning and disinfection of public toilets, and kept them clean in a timely manner. There are no flies or odor in the public toilets. Beigang Jiahe Fresh Food Market Co., Ltd. Card”

**“Neat stalls,
clean floors, a
bright
environment”**

Wutie Hongshan Market

Described in link as a 'comprehensive shopping mall'

Baishazhou Market

One of the four markets identified as selling wildlife

Covers 700 acres.

Wuhan Recycling Comprehensive Market

Very unclear what this location is, searching its chinese name offers links such as this (link in title) - discussing hubei and wuhan's technological and/or ecological policies. Unclear if this is even a market.

Wuhan Department Store Culture Wholesale Station

A department store selling entertainment goods and sporting equipment.

Wuhan Antique Jianguanpin Market



Jiangnan District Huanghe vegetable market



Hongshan market



Next outbreaks - Empirical Data (1/2)

Location of outbreak	Start of outbreak (2020)	Nb of cases	Size of outbreak	Source / Index case of infection
Xinfadi (Beijing)	Jun-2020	335	Major	Frozen salmon import ?
Dalian	Jul-2020	144	Major	Imported frozen food
Qingdao	Oct-2020	14	Small	Imported frozen food
Kashi (aka Kashgar)	Oct-2020	[428-430]	Major	Imported aviation container
Tianjin	Nov-2020	10	Small	Imported frozen food
Shanghai	Nov-2020	3	Small	Overseas aviation container
Manzhouli	Nov-2020	25	Small	Close contact with confirmed patients

Next outbreaks - Empirical Data (2/2)

Location of outbreak	Start of outbreak (2020)	Nb of cases	Size of outbreak	Source / Index case of infection
Qingdao 2	Dec-2020	2	Small	Cold-chain food
Chengdu	Dec-2020	14	Small	Garbage from imported cases
Beijing	Dec-2020	[3 - 41]	Small	Close contact with confirmed patients
Dalian 2	Dec-2020	69	Small	Cold-chain transportation
Shenyang	Dec-2020	38	Small	Pre-symptomatic individual who traveled abroad
Shijiangzhuang	Dec 2020 - Jan 2021	1129	Major	Contact with a patient from abroad
Northeast Region	Dec 2020 - Jan 2022	1650	Major	Visit of an infested city

Sources : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8322611/> and <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8139231/#s0055>

NB : It is possible that there are a few more small outbreaks not covered in these studies

Next major outbreaks and their link to markets

Xinfadi

(June 2020 / 335 cases)

Main SSEs:

Xinfadi market

**MARKET
INVOLVED**

“All the cases were somehow linked to the market : Among the 368 persons isolated and treated, 272 (73.9%) had epidemiologic links to the market—169 (46%) were market workers and 103 (28%) were visitors. The remaining 96 (26%) were close contacts of the other cases”

Dalian

(July 2020 / 144 cases)

Cold-chain processing company, Taoyuan market

**MARKET
INVOLVED**

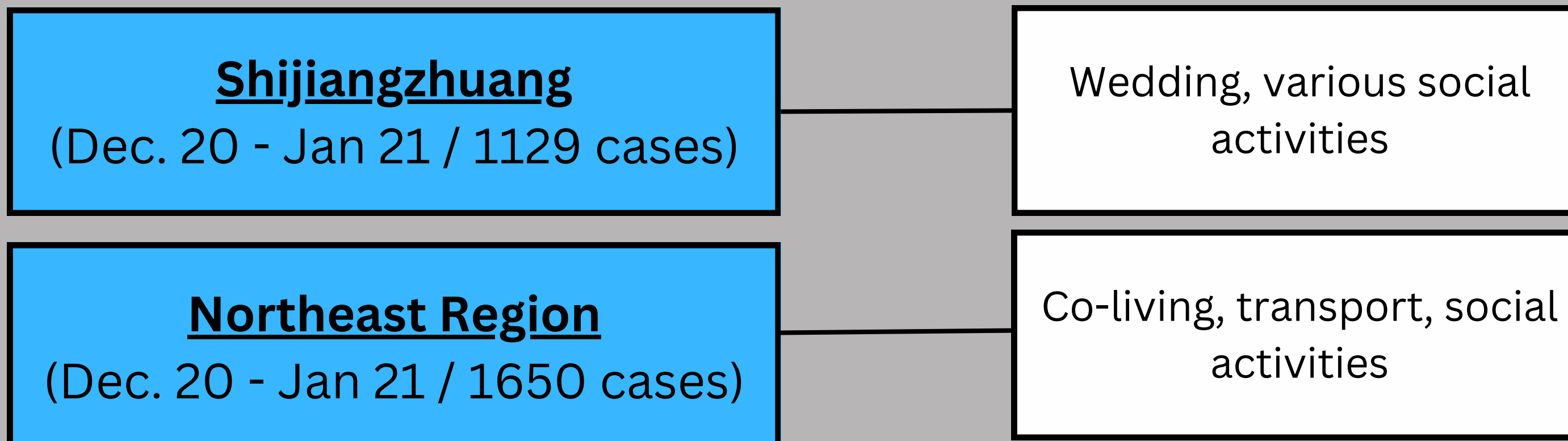
Infected people, reported since Aug 20, had visited the Taoyuan market, a well-known market with fresh vegetables, meat, seafood and groceries.

"Some cases since Aug 26 are owners or customers of the market. Therefore, they belong to the same outbreak, introduced by infected people into the Taoyuan market, which then became an amplifier and spread out," said Meng Jun, deputy director of the Dalian center for disease control and prevention

Next major outbreaks and their link to markets



“the Kashgar outbreak was initiated by an asymptomatic individual, which later on spread to other individuals as a result of eating together, co-habitation, co-working, shopping at farmer’s market, attending weddings or social gatherings, and visiting the hospital during the asymptomatic period”



Next major outbreaks and their link to markets

3 of the 5 next major outbreaks, the earliest ones, all involved a market

Additionally, the last two (December/January 2021) outbreaks occurred during the Delta variant wave, where infectivity was higher thus making markets less unique as a potential superspreader location.

Interesting to note that, as discussed earlier in the SARS outbreaks, a market with wildlife is not even such a common source for zoonotic spillovers - restaurants are much more common.

Responding to claim that SSE listings do not prominently feature markets

Peter brings 3 listings of SSEs, here is list 1 :

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

Responding to claim that SSE listings do not prominently feature markets

List 2:

SARS CoV 2	January 19, 2020	Ningbo, China	Bus	30	Female	64	www.nytimes.com/2020/03/01/health/coronavirus-bus-china.html
	January 24–28, 2020	France	Resort	11	Male	53	www.theguardian.com/world/2020/feb/10/super-spreader-brought-coronavirus-from-singapore-to-sussex-via-france
	February 10, 2020	Daegu, South Korea	Church	38	Female	61	www.theguardian.com/world/2020/feb/20/south-korean-city-daegu-lockdown-coronavirus-outbreak-cases-soar-at-church-cult-cluster
	Feb 26–27, 2020	Boston, MA	International conference	97	ND	~100	[4]
	Monday, March 2, 2020	New York, USA	Hospital (multiple)	90	Male	50	www.nytimes.com/2020/03/10/nyregion/coronavirus-new-rochelle-pneumonia.html
	Tuesday, March 10, 2020	Washington, USA	Choir practice	52	ND	ND	[80]
	June 17-20 2020	Georgia, USA	Summer camp	260 ^a	ND	Teenage	[81]
	Monday, August 17, 2020	Paju, South Korea	Coffee shop	56	Female	Mid 30s	www.businessinsider.com/56-got-coronavirus-south-korea-starbucks-mask-wearers-did-not-2020-8
	May 27-29, 2020	Utah, USA	Daycare facility	5	ND	Adult	[71]

Responding to claim that SSE listings do not prominently feature markets

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

List 3 :

Responding to claim that SSE listings do not prominently feature markets / Problems with the listings

1. Problems with the lists provided

There are many issues with the lists provided

- **These lists omit known crucial SSEs**
 - such as Xinfadi or Dalian in China or other SSEs outside of China (see slide below for examples) and omit full geographical areas like Latin America
- **For example, the 1st SSE in the list is a choir practice in Washington on March 10th.**
 - At that time, there were 313 confirmed cases & 39 deaths (given 1.5% fatality rate & the early stage of the pandemic then, it means there were 3000+ cases).
 - Many places will work as an SSE in a city with many infected people running around (for example, choirs have many factors that can lead to clusters: older - median age was 69 - singing, proximity, 2.5 hours together in a closed room), but the probability that such a small group of people will be one of the first few cases in a city is much smaller.

Responding to claim that SSE listings do not prominently feature markets / Problems with the listings

- **The selection in list 2 is not representative.**
 - It contains 9 SSEs. The sources are a mixture of articles and scientific papers. The list is far from complete.
- **In list 3, the data must be adjusted.**
 - The SSE with the highest number of cases is classified as "worker dormitories", it includes 1690 infected people, but this regroups 17 sites. As such, it gives only ~100 infected people per site.
 - Therefore, among the top 7 sites (after dividing by the number of sites), 5 are "Food Processing Plant" (1029 - 534 - 260 - 260 - 260 infected people).
 - These are similar to wet markets in their internal conditions, although they don't have the incoming traffic. However, the largest case from Germany involves 6500 workers, which is significant traffic.
 - This list strengthens markets as early clusters.

Responding to claim that SSE listings do not prominently feature markets / Problems with the listings

- **These lists describe SSEs happening while multiple parallel infections are being imported, which is very different from an early cluster**
 - This allows removing the key requirement of “high incoming traffic” which allows for an infection at early stages, when infection numbers are low.

Responding to claim about markets / More market outbreaks

2. There are actually more market examples, including outside of China

- **In Latin America**, markets played a major amplifying role in the pandemic:
 - Venezuela capital Caracas's largest produce market, the Coche Wholesale Market, was at the center of Caracas outbreak
 - In Peru, Lima's huge wholesale markets were qualified as "enormous hubs of infection" and "probably the biggest vector of infection".
 - Similar decisive impact of markets in other South American countries like Brazil (with São Paulo's CEAGESP wholesale market).
- **In Thailand** in mid-December 2020, after 6 months of near zero infections, more than 1,300 cases were traced to a seafood market in Samut Sakhon, a coastal province near Bangkok
- **In Singapore**, a cluster of nearly 1200 cases was linked to a fishery port (where seafood was sold to the public) and wet markets. This too happened months after zero Covid.

Summary - Factors making wet markets likely early clusters

- **When there are widespread infections or multiple imports of the virus, as most countries had in February-March 2020, the importance of the "high incoming traffic" factor decreases because there are many opportunities to import the virus.**
 - And indeed, we see a clear pattern where wet market cases are prominent after zero covid periods.
 - Few countries had zero covid periods, so there were fewer opportunities to see market clusters.
- **Behavioral factors may also come into play**
 - once the airborne nature and human-to-human transmission of the virus is known, people may behave differently in markets (more cautious: wearing masks...) than in smaller indoor gatherings (choir, churches...)

Summary - Factors making wet markets likely early clusters

- **Wet markets in the full sense of the term, meeting all the criteria we identified, are mainly found in East Asia and to a lesser extent in Latin America.**
- **Specifically for HSM, it further had the mahjong room, which seemed to be a major accelerator.**

Simulating HSM Infections

To illustrate the importance of these unique factors, let's plug a few numbers and see what result they produce after a few weeks.



Simulating HSM Infections

A lab worker is infected, and starts an infection chain

The virus spreads through Wuhan and reaches 128 people

One of them is likely to visit HSM over a few days:

Population 12,320,000

10,000 HSM visitors/day

Probability that 1 of them is from the 128:

$$128 * 10000 / 12,320,000 = \sim 10\%/day$$

Probably 2x higher due to high density around HSM = $\sim 20\%/day$

This is the first major factor differentiating HSM from other locations - a very high likelihood of getting an infection in the early days.

Simulating HSM Infections

The HSM conditions make an infection likely and growth rate higher than Wuhan average.

Ventilation

Density

Organic Materials

Temperature

Let's say 2x faster. Probably much more within HSM itself, but we are also counting "market related infections".

This is the second major factor differentiating HSM from other locations - higher transmission, which due to the exponential nature of a virus, can result in major differences within just weeks.

Simulating HSM Infections

The importance of these conditions was demonstrated in the Xinfadi outbreak

pang et al 2020: “To probe the origin of the infection, we analyzed the spatial distribution of infected employees in XFDM. Strikingly, 20.9% (122/584) of employees working in the basement of the XFDM trading hall (XFDM-TH) were positive for SARS-CoV-2, which is significantly higher than those of other areas in the market (1.7%, 47/2727, $\chi^2 = 363.29$, $P < 0.001$). “

Despite Xinfadi being primarily a vegetable market, most of the infections (47%) were from seafood vendors.

<https://academic.oup.com/nsr/article/7/12/1861/5936602>

Simulating HSM Infections

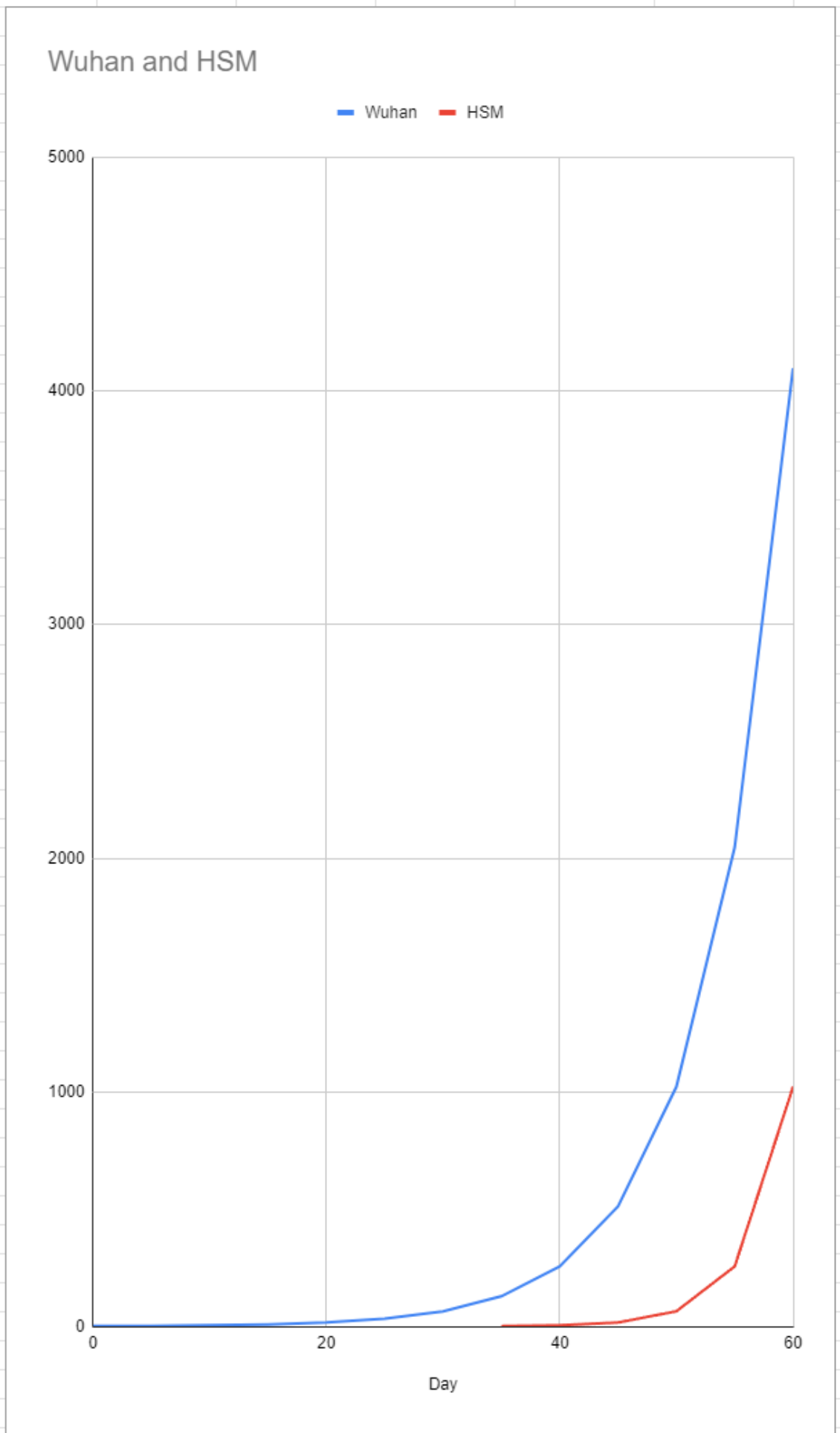
WHO reported roughly 120 cases that have no connection to the market, 60 that do. We will assume this is accurate, even though it is likely biased towards market cases. So we're looking to reach a ratio of 4:1 in infections. We're starting at 128:1, so we need 32x, which is 5 doublings - can be easily reached within a few weeks. Specifically:

The older population gets more symptoms, say 2x.

- We're starting at 128:1
- After 5 doublings, it's 4096:1024
- With 1 in 32 hospital/clinic visits found by WHO, and 1 in 16 in HSM, we get
- 128:64 - as reported by WHO

This brings us to the December numbers. At that point the market is saturated and can no longer grow fast. All of Wuhan continues to grow (perhaps a bit faster due to lineage B domination or other mutations), reaching lockdown at 100,000s of infections, and a few more after lockdown - producing the seropositivity numbers of April

Simulating HSM Infections



Simulating HSM Infections

Bottom line: This makes “HSM is the first cluster” weak as evidence. In the market model presented in session 1, having an early cluster where wildlife was sold was an interesting coincidence, but was overwhelmed by the evidence against wildlife involvement at HSM. Now it’s weak even before the evidence. Once considering the evidence, it is certain that HSM is not the source, even if Covid origins are zoonosis.

Specifically, under zoonosis, HSM is fairly likely to be the first cluster whether or not wildlife brought the virus to HSM. And since the evidence is strong against patient zero in HSM, HSM is very likely not the source.

Reminder of evidence against the HSM as origin

- Lineage B is dominant in the market, whereas lineage A is ancestral. We would expect A to be dominant in market if it were the epicenter.
- Majority of cases could not be connected to the market, including early cases with no plausible connection (Connors, Chen).
- No positive animal samples in the market or farms supplying it.
- No evidence of positive infections among animal vendors, and infections are distributed uniformly.
- Positive environmental samples are not positively correlated with wildlife stalls.
- Positive environmental samples with wildlife mtDNA are not positively correlated with SARS2.

Where are the early cases outside HSM?

We maintain that the earliest cases were not the ones that were recorded from the 10th December in the market, but rather were the result of low-level transmission outside of the market until that date. Early cases were thoroughly investigated in [a lengthy report put together by Dr. Gilles Demaneuf](#). Summary of its conclusions:

Conclusion:

Limitations of the official 2019 cases

Observations:

The outbreak most likely started in September to mid-November 2019:

- The first official market case goes back to only December 10, 2019.
- In contrast, on top of genetic clock arguments (not discussed here), there are also epidemiological ones to support a Sep to Nov 2019 Wuhan emergence (at the market or elsewhere), starting with the November 2019 cases strongly hinted to by the SCMP leak (a leak already validated for its Dec 2019 onset cases), but also extending to odd cases such as Connor Reed.

The market filled many ideal conditions for an superspreader event:

- middle aged and elderly population.
- mahjong games in the evening in a fairly enclosed room (to which many of the earliest known cases took part), situated just above the point of maximum density of positive environment samples.
- transport hub nearby, with also recorded visits from outsiders to the mahjong room.

The market also filled many ideal conditions for the detection of the outbreak there:

- near the very best top tier-3A hospitals in Wuhan, reached as some initial cases eventually progressed through the cheaper but much less reactive small local clinics and secondary hospitals, which for many low income early patients would have instead been the most realistic medical care option before the announcement of the outbreak,
- a PUE surveillance hospital nearby (i.e.: **the Wuhan Central Hospital which broke the news of the outbreak**),
- a specific attention, in at least one of these key hospitals, to **clusters of PUEs linked to markets** following SARS (i.e.: **SARS veteran Zhang Jixiang, the official whistleblower of Hubei Provincial Hospital of Integrated Western and Chinese Medicine**).

Where are the early cases outside HSM?

Conclusion:

Limitations of the official 2019 cases

Observations (cont):

By contrast:

- Early cases happening far away from the market, without any obvious link to any wildlife, would by definition be much less likely to raise any suspicion at all, especially in the middle of a flu season, and in the absence of any sizable cluster.
- This is indeed exactly what happened in the case of Mr Chen (onset 16 Dec), who was first treated in a Jianxia hospital 30km away on the opposite side of Wuhan, who did not go to any wet market, lived and worked exclusively locally, and was only diagnosed when he transferred across the river to Wuhan Central Hospital where a relative worked.

These implicit outbreak recognition biases were soon followed by explicitly reporting ones:

- From ~1 Jan 2020 to at least 24 Jan 2020, cases without a market link were systematically repressed, in the sense that they could not even be entered as a 'suspected' infection in the CDC NNDRS (an entry which was a requirement to get the CDC to test the patient samples).
- By contrast, cases with a market link could at least be entered, even if they were then typically prevented in other ways from progressing to a 'confirmed' or 'diagnosed' infection until 24 Jan 2020 (so that the official numbers would still not change).
- These reporting biases only abated once human transmission was eventually officially recognized.

In the end, the time allocated to a proper retrospective search for 2019 cases was very limited:

- Only about 4 weeks were spent by Chuanhua Yu and his team up to 25 Feb 2020 to do the official retrospective search.
- That search was not just for 2019 cases but covered **all cases** to 7 Feb 2020 or so (date of sufficient availability of enough RT-PCR kits). There simply was not much time to work on 2019 cases.

Where are the early cases outside HSM?

Conclusion:

Limitations of the official 2019 cases

Observations (cont.)

November 2019 cases were likely identified:

- the nine Nov 2019 in the SCMP leak are very likely real and were in the NNDRS database at end Feb 2020,
- as it happened, Chuanhua was quoted in the Health Times mentioning some likely Nov 2019 cases, but had to retract his statement just after the [gag order](#) on the CDC on 25 Feb 2020.

Connor Reed is another likely Nov 2019 case:

- 25 Nov onset case with no known or likely connection to the Huanan market,
- he was likely misdiagnosed as some other condition when he did an outpatient visit to Zhongnan Hospital on 6 Dec 2019,
- his samples collected that day were eventually tested by Zhongnan Hospital, and came back positive by 16 January,
- his confirmed case could not be entered in the NNDRS database at the time since there was a total reporting block from 5 to 17 Jan (the 'two sessions'), during which there was officially not a single new COVID-19 case in the whole of China.
- His case is not included in the WHO data either.

Connor Reed:

Connor was a 26-year old English teacher in Wuhan. His illness began on **November 25**, 2019 with symptoms similar to a cold. Over the following 11 days he developed fever, extreme fatigue, cough and trouble breathing.

The difficulty in breathing became serious enough that he took a taxi to Zhongnan hospital on **December 6**, 2019 to seek clinical treatment. On admission, a doctor diagnosed him with pneumonia and sent him for further tests which lasted 5-6 hours before being sent home with antibiotics to recover.

Connor chose not to take the prescribed antibiotics and slowly recovered on his own using tiger balm as a self treatment. Overall, his symptoms were consistent with a respiratory infection including COVID-19.

His visit to Zhongnan hospital was before December 30, 2019 and so the discovery of the new virus had not yet been made public. Once the kits became available (after January 12, 2020), Connor called Zhongnan hospital on **January 16** to get the results of the PCR test of his December 6 samples, which turned out to be positive for SARS-CoV-2.

based on: [Diary of Connor Reed's illness](#), Daily Mail, 5 Mar 2020

Connor Reed, a case from 25th November

One of the earliest COVID-19 cases was Connor Reed, a British national who worked in Wuhan (No known connection to HSM) and contracted the virus on November 25th, 2019. Reed sought medical treatment 11 days later on December 6th and was diagnosed with pneumonia. On January 16th, a retroactive PCR test of his December 6th sample came back positive for Covid-19.



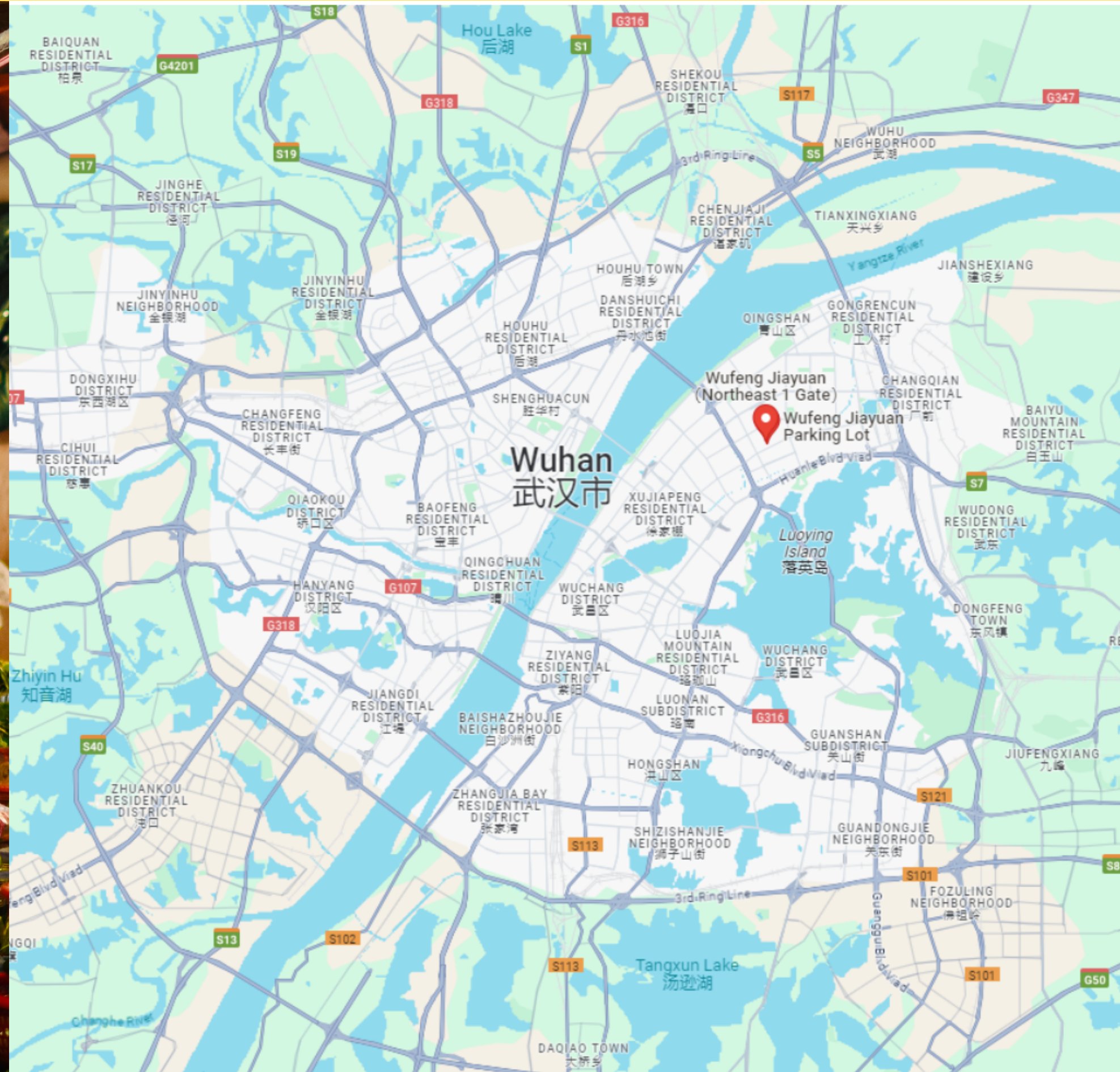
Geolocating Connor Reed



https://www.researchgate.net/figure/The-Gate-of-Wufeng-Jiayuan-Phase-II-development-Source-Baidu-Street-Maps_fig1_355373689

<https://www.facebook.com/LADbible/videos/720271455375784/>

Geolocating Connor Reed



Peter's description of Connor's case in Session 1

British teacher (Connor Reed) living in Wuhan. Might have gotten sick November 25th.

Reported in a [tabloid article \(Dailymail\)](#)

Describes cold like symptoms for a week, followed by much worse fever and pneumonia

Says that he shopped at the seafood market, thinks he caught the virus there.

Claims his cat died from covid.

Likely just a tabloid story, but if accurate, that would just confirm the virus was at the market weeks before December 10th.

Claim 1: British teacher (Connor Reed) living in Wuhan. He might have gotten sick November 25th.

Reported in a tabloid article (Dailymail)

- **This can't be dismissed as some tabloid story. Connor gave dozens of video interviews including to the Guardian. This is a very reliable story pointing to the same direction as Chen (who is also on the south side of the river with no HSM connection) - The virus spread over Wuhan long before the HSM outbreak.**

Claim 2: Connor Reed describes cold-like symptoms for a week, followed by much worse fever and pneumonia and says that he shopped at the Huanan market, thinks he caught the virus there.

- **He never visited HSM. He visited a fish market where he lived, about an hour away from HSM. He said he could have gotten infected there but there's no way to tell.**

Peter's description of Connor's case in Session 1

Claim 3: he said that his cat died from covid.

- He said the cat got feline coronavirus, and this is a coincidence.

Claim 4: Never went to the hospital, so naturally he wasn't included in the government reports.

- He went to Zhongnan University Hospital, which is on the south side of the river. The hospital informed him 40 days later that he was positive for SARS-CoV-2. Referring to testing of his sample (2:58).

Conclusion: Connor Reed is a confirmed November case with no connection to HSM. This is strong evidence that further reduces the probability that HSM was the source.

Session 1 Recap

In the first session, we discussed the location of the outbreak, focusing on 4 main evidences:

Priors

Wuhan Location

**SARS2
matches
WIV activity**

The Market

Priors and Location

Estimating directly the rate of emergence, rather than the ratio between the hypotheses.
Definition: Probability per year of a bat coronavirus pandemic starting in Wuhan.

Zoonosis

- We have SARS and MERS which were not infectious enough to cause a pandemic and HKU-1 which wasn't lethal enough.
- So a generous estimate is once every twenty years, or 5% a year.
- 1.5% for it to happen in Wuhan

Total: 0.075% per year

In a lab leak the following need to occur:

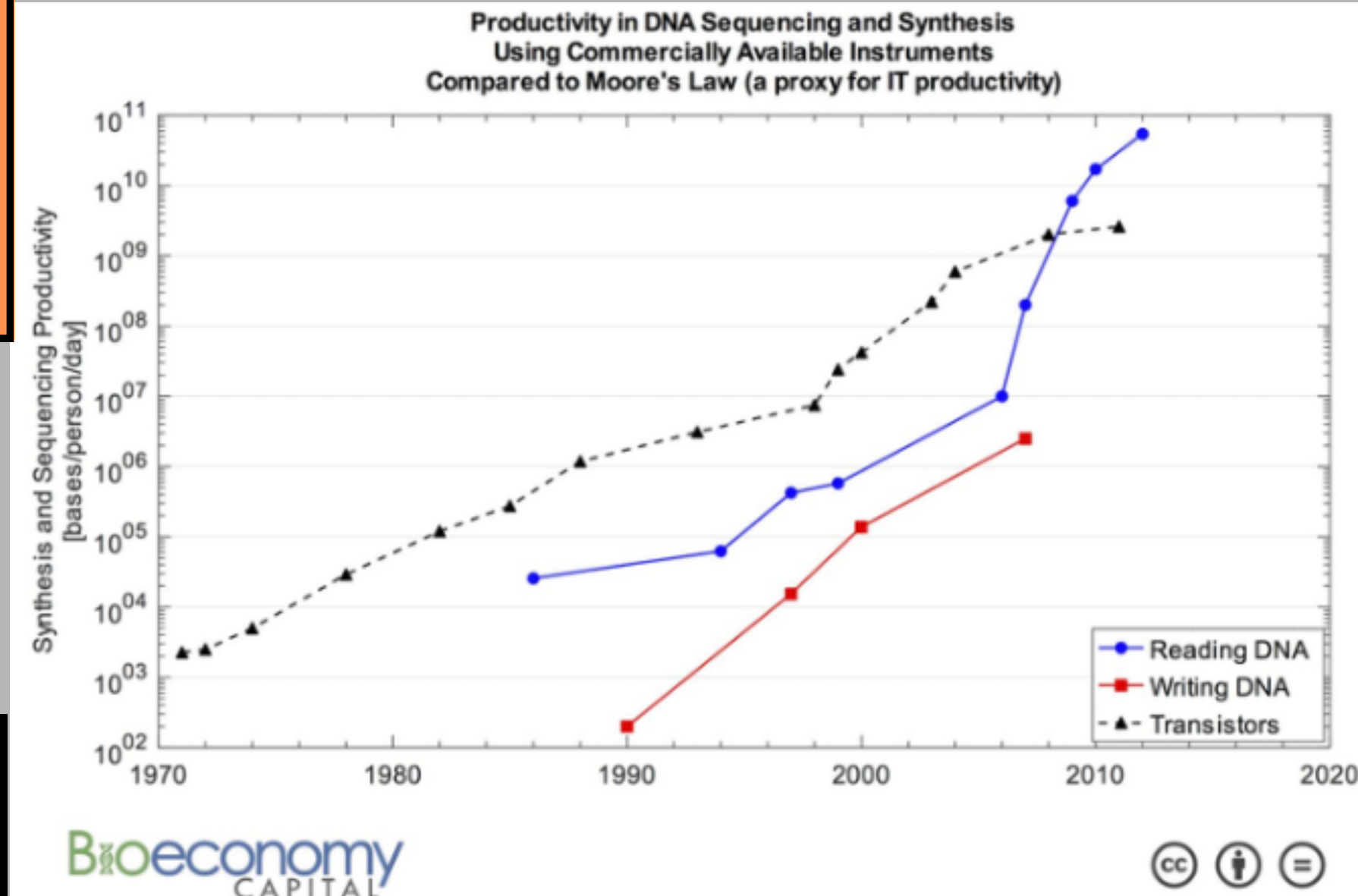
- Despite DEFUSE rejection, WIV started a similar project: 40%
- This involves screening for human ACE2 match. Given BANAL-52 is identical to SARS2 in that aspect, and WIV had 180 unpublished viruses, it's likely they would find a good match: 50%
- They would then optimize it for human infection (e.g. adding an FCS).
- Given that this work was done in BSL-2 a lab worker infection is quite likely: 15% per year
- Given infectivity, this would start a pandemic in Wuhan: 40%

Total: 1.2% per year

Final Result 94% Lab Leak

Location and GoF growth

Under zoonosis: In 2020, Wuhan's population stood at 12.3 million / Total relevant East Asian urban population 1.66 billion. 2x for large cities.



Calc: $12.3\text{m} / 1.66\text{bn} * 2 = \text{Wuhan's Share of east asia urban population} = 1.48\%$.

Under lableak: GoF research into bat coronaviruses has increased exponentially

Matching WIV activity

Prior	Low	Impartial	High
SARS1	0.3	0.5	0.7
Flu	0.15	0.2	0.25
Major pandemics	0.100	0.125	0.150
Geomean	16.51%	23.21%	29.72%
Adjust for GoF and growth	0.5	1	1.5
	0.08254818122	0.2320794417	0.4457941465
Adjusted	9.00%	23.21%	38.81%
Location Wuhan			
Zoonosis	1.48%	1.48%	1.48%
Lab Leak	10%	20%	30%
Ratio	6.74	13.48	20.22
Adjust for BSL-2	2	4	6
Total Wuhan	13.48331126	53.93324503	121.3498013
	1.11	12.52	54.10
Updated	57.14%	94.22%	98.72%
<u>Matching WIV activity</u>			
FCS	1.5	2	3
Human ACE2	3	5	10
N-glycans	1	1.5	2
Total WIV Activity	4.5	15	60
	5.008602695	187.7519609	3245.821866
Updated	85.71%	99.59%	99.98%

Best explanations under Zoonosis:

- FCS - This is what enabled the jump
- ACE2 and N-glycans - Coincidence and cryptic transmission

**Weighted:
96.9%**

The Market

The market is mostly ignored as evidence, as it is unlikely to be the source.
The new market model solidifies this conclusion from Session 1

Weighted:
95%

Prior	Low	Impartial	High
SARS1	0.3	0.5	0.7
Flu	0.15	0.2	0.25
Major pandemics	0.100	0.125	0.150
Geomean	16.51%	23.21%	29.72%
Adjust for GoF and growth	0.5	1	1.5
	0.08254818122	0.2320794417	0.4457941465
Adjusted	9.00%	23.21%	38.81%
Location Wuhan			
Zoonosis	1.48%	1.48%	1.48%
Lab Leak	10%	20%	30%
Ratio	6.74	13.48	20.22
Adjust for BSL-2	2	4	6
Total Wuhan	13.48331126	53.93324503	121.3498013
	1.11	12.52	54.10
Updated	57.14%	94.22%	98.72%
Matching WIV activity			
FCS	1.5	2	3
Human ACE2	3	5	10
N-glycans	1	1.5	2
Total WIV Activity	4.5	15	60
	5.008602695	187.7519609	3245.821866
Updated	85.71%	99.59%	99.98%
Market	0.5	1	1
	2.504301348	187.7519609	3245.821866
Updated	75.00%	99.59%	99.98%

Session 2 (Genetics) Probabilities

- **FCS - Ignored (despite no precedence in sarbecovirus)**
- **12nt clean Insert**
 - **If basing on frequency of large insertions, probably over 1000x.**
 - **Similar estimate if looking at the coincidence of the only long insertion happening to be in the most important feature of the virus.**
 - **Best explanation is there is some unknown reason why an FCS specifically should emerge with a long insertion. Years of discussions have yielded no such suggestion. Estimated at 50x, Low of 20x.**
 - **CGGCGG - Best explanation is the first CGG is random, and the second was a duplication event (more likely given the insert). 10x.**
- **Leading Proline**
 - **Could be inspired by MERS or the PAA sequence in bat coronaviruses.**
- **Why insert RRA and not RAR (for a more canonical RARR)?**
 - **Others have done it and they could be testing PAA -> PRA -> PRRA.**
- **In any case, hard to say any lab action is unreasonable, as it's hard to cover all the possibilities (See slide "Loose Ends - Unequal Conditional Probabilities").**

Session 2 (Genetics) Probabilities

<u>Genetics</u>			
12 nt clean insert from unknown source	20	50	100
CGGCGG			
Zoonosis	0.0026	0.0026	0.0026
Lab Leak	0.027777777778	0.027777777778	0.04
Ratio	10.68376068	10.68376068	15.38461538
PRRAR and "out of frame"	0.3	0.4	0.5
Total Genetics	64.1025641	213.6752137	769.2307692
	160.5321377	40117.94037	2496786.051
Updated	99.48%	100.00%	100.00%

**Weighted:
99.9%**

Session 3: Other Evidence

This session we will review the evidence that didn't fit into the categories of either of the first two sessions, which nonetheless informs our probabilistic analysis.

No intermediate host found

We expect to have an intermediate host as the virus is not effective in bats.

- We also expect that for a global emergency there would be far stronger search efforts.
- We additionally have exponential advantages in sequencing.

City	Date of first case in local fresh outbreak	Confirmed cases	Asymptomatic infection cases	Source of infection	Database
Beijing	June 11	335		Cold seafood	Link 1; Link1.2
	December 14	3		Close contact with confirmed patients from overseas	Link 2
Dalian	July 9	118	26	Imported frozen food	Link 3
	December 15	41	28	Cold-chain transportation	Link 4
Qingdao	October 11	2	12	Imported frozen food	Link 5
	December 1	0	2	Cold-chain food	Link 6
Kashi	October 24	78	350	Imported aviation container	Link 7
Tianjin	November 8	10		Imported frozen food	Link 8
Shanghai	November 9	3		Overseas aviation container	Link 9
Manzhouli	November 21	23	2	Close contact with confirmed patients from overseas	Link 10
Chengdu	December 7	14		Garbage from imported cases	Link 11

All 11 outbreaks in 2020 are imported. Not one wild life case.

Comparing to Sars1 - Intermediate Host

Comparing to SARS, multiple animals were found with ancestral strains. Especially interesting is this study.

4 of 6 civets sampled in a live-animal retail market in Shenzhen during SARS were PCR+.

The animals “originated from different regions of southern China and had been kept in separate storehouses before arrival to the market.” (Note this relates to all 25 animals sampled. It’s possible civets were from a single source, although the next finding makes that unlikely)

3 of the 4 were phylogenetically distinct, with up to 18 nt difference, and were distinct from human viruses who had a 29 nt deletion.

This is interesting in two aspects:

- The prevalence of a virus during a true zoonotic outbreak. One market housed three different strains, with 4 of 6 civets infected, indicating widespread infections in wild civets.
- The diversity demonstrates how true multiple spillovers in the same market would look like - probably not just 2 mutations apart.

No intermediate host found

Lockdowns and the closing of the market may have made tracking down intermediate hosts harder

However, the interest in finding the host is not merely an academic endeavor. It has real world utility and informs the emergency response.

**Finding the host is crucial to prevent more spillovers in the early days of a pandemic.
Especially given that China was trying to reach zero infections at the time, this could only have been done by finding the intermediate hosts and ensuring humans don't come into contact with them in other markets.**

No intermediate host found

- There is a reverse dependency between claiming two spillovers and the evidence of no intermediate host found
- The more spillovers there are, the more animals should be infected, the less likely we are to not find an intermediate host, or have any other spillover anywhere in the world.
- Especially when we see multiple outbreaks in China from imported food/contacts, during Zero COVID.

This pushes for a Best Explanation of one spillover

No intermediate host found

This is an important thing to evaluate given that we already chose the best explanation that the market is not the source. This means we give much less weight to not finding a source in the market's suppliers specifically.

Coronaviruses also jump from wild animals into livestock

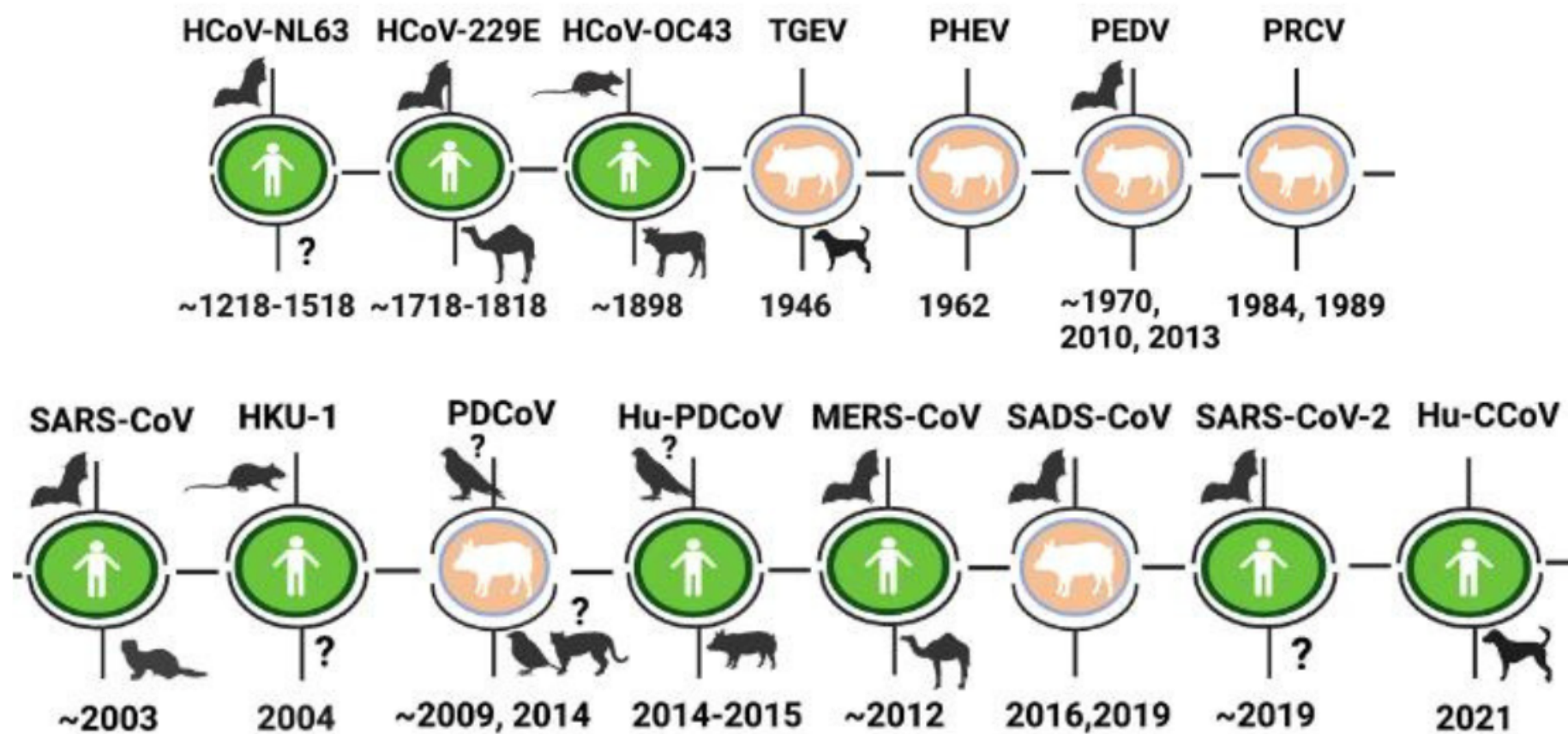


Image from [Keusch et al 2022](#). These are emergence dates. Top animal is the host species, bottom is the intermediate species.

Sometimes we find the intermediate host, like with SARS in 2003. Sometimes we can't find it, like with HKU-1 in 2004.

Overall it seems very unlikely not to find an intermediate host under these specific conditions. However, there aren't enough priors so we assign it a ratio of just 2 to 4

WIV Behavior

There were some irregularities regarding the sequencing of RaTG13

While RaTG13 was fully sequenced in 2018, a paper published by WIV researchers in February 2020 implied that RaTG13 had been sequenced only after the COVID-19 outbreak (“We carried out full-length sequencing on this RNA sample”). However, the WIV paper referenced RaTG13 without mentioning that it was previously known and sequenced as BtCoV/4991.

Long after the relationship was pointed out, an addendum published in November 2020 by Shi and her co-authors confirmed that BtCoV/4991 is indeed identical to RaTG13.

WIV Behavior

```

LIGAEHVNNNSYECDIPIGAGICASYQ : 675
LIGAEHVNNNSYECDIPIGAGICASYQ : 675
LIGAEHVDTSYECDIPIGAGICASYH : 661
LIGAEHVDTSYECDIPIGAGICASYH : 661
LIGAEHVDTSYECDIPIGAGICASYH : 662
LIGAEHVNASYECDIPIGAGICASYH : 648
LIGAEHVNASYECDIPIGAGICASYH : 652
  
```

S1/S2



SARS-CoV-2	GICASYQTQT-NSP	RRAR	SVA
SARS-CoV	GICASYHTVS	-LL----	R STS
MERS-CoV	SLCALPDTPSTLT	RSVR	SVP
PANGOLIN-CoV	GICASYQTQT-NS	----	R SVS
BAT-CoV	GICASYQTQT-NS	----	R SVA

PAI

Later, when publishing the SARS2 RaTG13 alignment, they cut it right before the FCS, where the very obvious insert would appear.

Behavior of WIV

The database of bat and rodent viruses with 80,000 samples went offline and then went back on for a short while in September 2019. While the 80,000 sample database seems to have gone offline for an unrelated reason, it is weird that WIV did not find a way to bring it back or otherwise share it with colleagues in such a time.

WIV claimed it was cyber-attacked but that fails to explain why they didn't share it in another way

Behavior of WIV

- **The database going down in September is not interesting because it is too early**
- **However, why not provide as much access to virologists as needed? This could save lives and clear the WIV of involvement.**
- **The WIV claimed a cyber attack on the database. This is strange and unsatisfactory as there are other ways to share the database**

Behavior of WIV

The WIV's behavior is far from what you would expect from virologists trying to help in a major pandemic.

- **Best Explanations:**
 - **Under lab leak, these two behaviors are expected because they would have the actual backbone from which they built SARS2 in those databases**
 - **For zoonosis, the best explanation would be academic competition, but that doesn't seem convincing in those circumstances.**
- **So we give this a factor of 1.5 to 2.**

Behavior of WIV

On the other hand, WIV workers have shown some relatively normal behavior during the start of the pandemic

- **Shi Zhengli went to a conference in Singapore on December 10th and they had a team dinner on January 15th**
- **This is not expected after such a major incident where extra care and frenzy is expected**

Behavior of WIV

Best explanations for WIV workers under lab leak :

- **For the December conference our best explanation is that they weren't aware of the leak at that time**
 - **For most of the leaks - especially with a respiratory virus - patient zero is not even aware of it.**
- **For January 15, the best explanation would be that:**
 - **(i) only a few people directly involved were aware of the leak, and**
 - **(ii) they were either in denial or pretending to behave normally**

WIV Ignoring FCS, RaTG13, the conference etc. are together given a factor of 1.5 to 2

Chinese government response

The official Chinese response was not transparent, though not particularly surprising even if the virus developed zoonotically

Members of the team participating in a WHO mission tasked with designing and proposing scientific studies have highlighted a concerning observation. They noted that Chinese authorities primarily provided their data analysis, rather than sharing the raw numerical data. Unfortunately, the team's expectations for detailed investigations into blood samples collected before December 2019 and a comprehensive list of animals sold at the Huanan market were not met, as Chinese scientists did not conduct these crucial studies.

Chinese officials confirmed that they had destroyed samples of the novel coronavirus, claiming it was done "to prevent the risk to laboratory biological safety and prevent secondary disasters caused by unidentified pathogens."

At first, China rejected the proposal for an independent international investigation into the origins of the coronavirus, which was to be led by the World Health Organization (WHO). They argued that such requests were politically motivated and asserted that focusing on an investigation would divert both attention and resources from their ongoing efforts to combat the pandemic.

Chinese government response

Strong limitations were imposed on the WHO team that was eventually allowed in, a year after the initial outbreak.

China secured veto rights over participants and insisted its scope encompass other countries as well.

The team didn't receive raw data regarding the earliest infections.

WHO team members said they didn't have the mandate, expertise, or access to investigate a potential lab leak.



Chinese government response

Peter Daszak says the WHO team had full access to everything they asked for.

However, he is motivated to focus on a zoonotic source, as he himself supported gain-of-function research, including coronavirus research and partnered with the WIV

While this is likely behavior under lab leak, it is not too unlikely under zoonosis, if China wants to point to a source outside China

No Contact Tracing

No published contact tracing efforts

We don't have any studies in which Chinese authorities tried to trace contact back, up to the earliest possible infection.

- This could have found patient zero, which could have led to the animal, and closed the case
- This would also help in preventing future spillovers, not just in investigating the origin of COVID-19.
- This would be a relatively cheap effort mostly based on interviewing a few hundred patients and contacts, until possibly reaching patient zero. No such effort was published.

This likely means either:

- a) They know what the result would be.
- b) They did it and they don't want to publish the results.

No Contact Tracing

For comparison, here is how it looked in SARS1:



Patient 1 had the earliest case, identified by retrospective case searching. He lived with his wife and four children in Foshan city and became ill on November 16, 2002. He had not traveled outside Foshan in the 2 weeks before his illness and had no contact history, but he had prepared food including chicken, domestic cat, and snake. He was part of a cluster of five patients, including his wife (42 years old, onset December 1), a 50-year-old aunt who visited him in the hospital on November 20 (onset November 27), and the aunt's 50-year-old husband (onset November 30) and 22-year-old daughter (onset December 4) ([Figure A1](#), Cluster A). None of patient 1's four children were ill. Subsequent clusters in Foshan included a food handler who infected a family member and two healthcare workers ([Figure A1](#), Cluster B) and a food handler who infected several healthcare workers ([Figure A1](#), Cluster C).

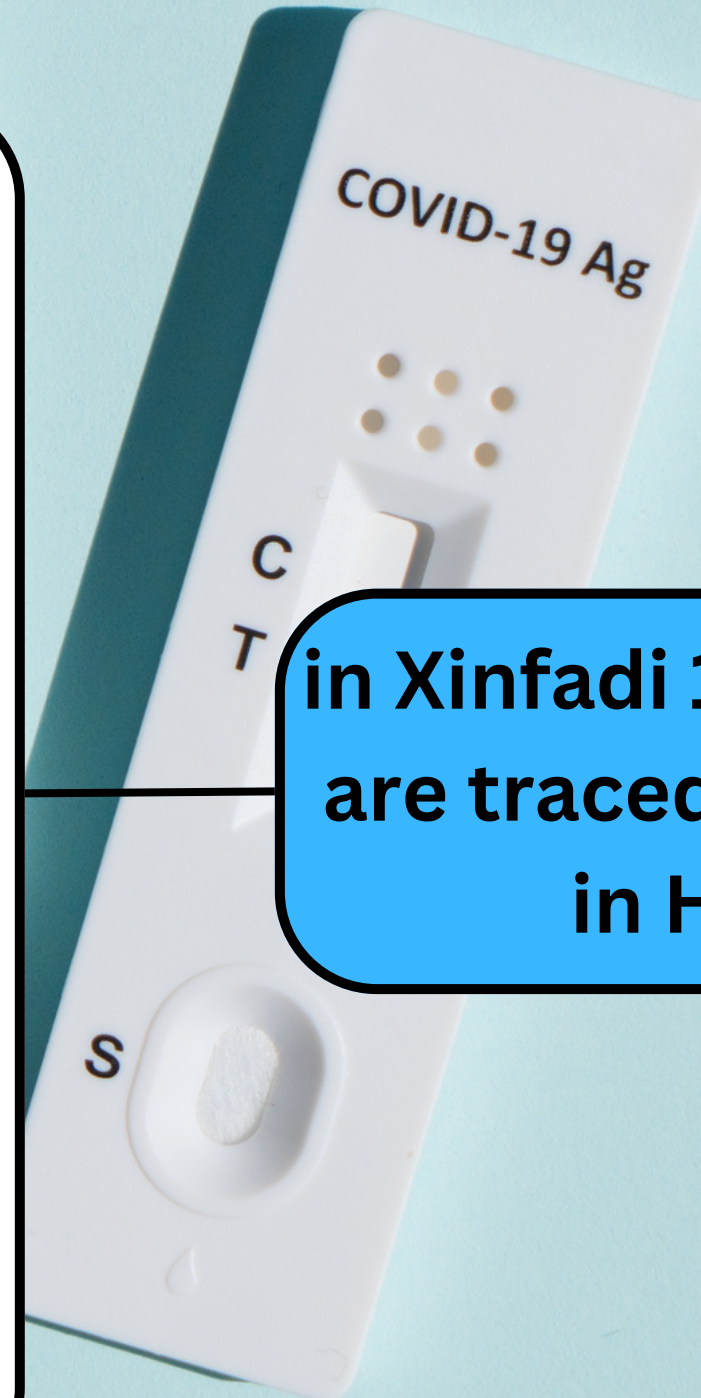
Patient 2 lived in Heyuan but worked as a restaurant chef in Shenzhen. His work was mainly stir-frying and did not involve killing animals. His animal contact history is unknown. He returned to Heyuan after becoming ill, was admitted to the local hospital and transferred to Guangzhou 2 days later. He infected a work colleague (41-year-old man, onset December 16), six healthcare workers in Heyuan (onset December 24–28), and a physician who accompanied him in the ambulance from Heyuan to Guangzhou (28-year-old man, onset December 25). Patient 3, from Jiangmen, had no contact history, no history of animal contact, and no known forward transmission. Patient 4 worked as a chef in a Zhongshan township restaurant, where he prepared steamed dishes and had contact with snakes, civet cats, foxes, and rats. He infected his 30-year-old wife (onset January 3), a 39-year-old male friend who visited him in the hospital (onset January 5), and a physician (35-year-old man, onset January 4). Patient 4 was one of two patients responsible for infecting at least three healthcare workers (onset January 4–7). Patient 5, from Guangzhou, had no history of animal contact other than with a pet guinea pig that died a month before his symptoms began. He infected a hospital intern (onset January 14) and six other healthcare workers (onset January 14–22) at Guangdong Traditional Chinese Medicine Hospital. Patient 6, from Shenzhen, had visited Hong Kong on January 14, the day before symptom onset. However, he had no contact history or contact with animals. He infected a work colleague (43-year-old man, onset January 29) and died 14 days after becoming ill. Patient 7, from Zhaoqing, was the only female index case. She traveled to Guangzhou 2 weeks before becoming ill, although she could not recollect contact with anyone with symptoms of SARS. She worked at a market but did not sell animals. She infected her 16-year-old son (onset February 3) and a physician (25-year-old woman, onset January 31).

No Contact Tracing

When taken seriously, complete tracing is possible


“Between 15 June and 10 July, a total of more than 10 million citizens, and 5342 environmental samples were screened. Eventually 368 qRT-PCR positive cases were confirmed (Fig. S1A), of which 169 (45.9%) had a history of working in XFDM. Of the visitors to XFDM between May 30 and 12 June, 103 (28.0%) were diagnosed. The remaining 96 (26.1%) patients had contact with the infected employees or visitors. These findings suggested a single outbreak source in Beijing”

in Xinfadi 100% of the infected are traced to Xinfadi, whereas in HSM only 33%

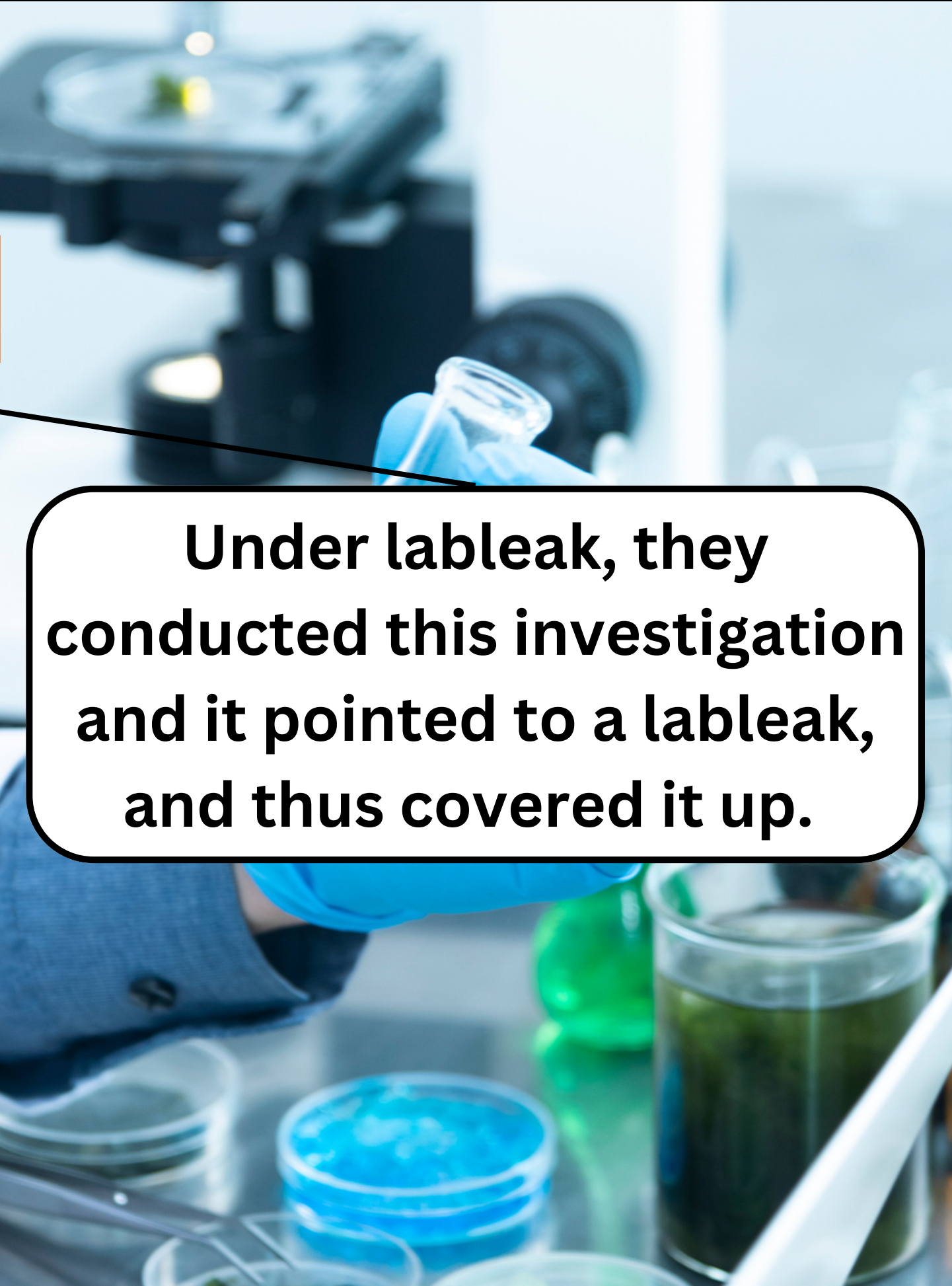



No Contact Tracing

Best Explanations



Under zoonosis, they did make such an effort and they weren't successful, didn't find anything of substance.



Under lab leak, they conducted this investigation and it pointed to a lab leak, and thus covered it up.

Missing evidence intro

Accounting for missing evidence is crucial in a robust probabilistic analysis. In some cases the fact that evidence is expected but doesn't appear has significant probabilistic strength. In this case, the following missing evidences are important:

**No
whistleblower**

**No info from
intelligence
agencies**

**No known WIV
infections**

**No published
backbone**

Missing Evidence: No Whistleblower

- So far no one from WIV claimed responsibility or blew the whistle. Not even anonymously.
- Given that we already accepted the best explanation of a few people involved and given the implications and gravity of the consequences, it is fairly reasonable for people not to come out.

**We therefore give this
a factor of 0.4 to 0.5.**

Missing evidence / No info from intelligence agencies

- The intelligence agencies have not provided any information.
- As Peter discussed in session one, there were some claims of sick people at WIV

Based on our experience in previous analyses, we give little weight to anonymous government sources.

On the other hand, WIV is not a body that we expect to be heavily covered by intelligence agencies.

WIV is a civilian, not a military virology institution.
So no info from intelligence agencies is not surprising, factor of 0.5 to 0.7.

Missing evidence / No known WIV infections

No known WIV infections

In addition to the US intelligence report which is not given much weight, there is the case of Huang Yanling who WIV responded suspiciously to questions about her disappearance

Given that SARS2 is not too harmful to young people and the most likely patient zero would probably be a young PhD or postdoc, it's definitely possible that the first infections left no significant trace.

Factor of 0.5 to 1

No published backbone

No published backbone for SARS 2. But given what we've discussed before, this is expected

As previously discussed, WIV was likely involved in wide bat virus collection efforts, which are only published years later

Since recent sequences are unpublished, it is not surprising the SARS2 backbone is unpublished

This study, co-authored by Yuri Deigin, found four sequences of an unpublished HKU-4 with MERS inserts

document the use of a previously unpublished HKU4 reverse genetics system in apparent MERS-CoV related gain-of-function research. Our study also emphasizes the importance of improved biosafety protocols in sequencing centers and coronavirus research facilities.

That specifically shows gain of function research on an unpublished backbone

Behavior			
No contact tracing	1.5	2	4
Not sharing genomes	1.5	2	4
RaTG13, Ignoring FCS etc	1.5	2	4
No whistleblower	0.4	0.5	0.6
No leaks / intelligence	0.5	0.7	1
No published backbone	0.3	0.4	0.5
No known WIV infections	0.5	1	1.5
No intermediate host	2	4	10
	0.2025	4.48	288
	32.50775788	179728.3729	719074382.7
	97.50%	100.00%	100.00%
Weight	20%	60%	20%
Final	99.50%		

Summary of probabilistic calculations

We provide three estimates: low, medium and high. The medium and high estimates are way above 99.9%.

Only on very conservative estimates we do not have certainty here, but even in this case we have a fairly high probability for a lab leak

We merge the three estimates at a ratio of 20-60-20:

- 60% represents the fact that most of the time we do reasonable estimates and the medium numbers are correct**
- 20% accounts for the possibility of having a strong bias towards one of the hypotheses.**

Summary of probabilistic calculations

Averaging this way, we get a final number which is 99.5% probability for a lab leak.

- **Since the calculation is itself proven in probability theory, to refute the number one would have to show which estimates are way off; whether there are unaccounted dependencies; other evidence, etc.**
- **There are a lot of things that would need to be wrong to get below 50%. Their combined strength should be over 1:100,000.**
- **Nothing that was discussed during the sessions can come close to this range.**

We can therefore conclude with certainty that SARS2 originated from a gain of function laboratory leak.

Let's now move to a more intuitive understanding of what this number means

Intuitive Understanding of the Model

Up to now we have discussed and weighed each piece of evidence separately, giving each the best explanation under both hypotheses.

The best explanations are far more probable under the lableak hypothesis.

Now we will try to make the probabilistic model more intuitive

What Needs to be Explained?

The idea of probabilistic inference is to quantify which hypothesis better fits our world experience. The hypothesis that requires less unusual explanations / coincidences to explain the evidence is the more likely one.

Lab leak needs to explain only two things:

Priors: the assumption is that zoonoses happen more often.

Explained well by growth of GoF, BSL-2 highly likely to leak

The market being the first early cluster

Explained well by the HSM's location, size, and conditions

Zoonosis needs to explain all of these things:

Wuhan: The outbreak starts next to a major bat Coronavirus lab doing gain of function and was specifically involved in the DEFUSE proposal, which is likely to produce this exact virus.

After correcting for the size of Wuhan and its location, there is no explanation why this would be the place other than a coincidence.

What Needs to be Explained?

FCS

- First FCS in a sarbecovirus: Ignored (conservatively), since some unique feature is expected for a new pandemic
- Clean 12nt insertion: No explanation given other than coincidence (“weird stuff happens in nature”).
- CGGCGG: The best explanation we could find is just that there's something weird that we don't know here.

No explanation given other than coincidence

All the evidence against the market

- No positive animals in the market or its sources, or anywhere
- The negative correlation in wildlife vendors samples
- No animal vendors infected
- Dominance of the non-ancestral lineage B

All the evidence against a spillover

- The early genetic variance points to a quick localized jump rather than multiple jumps over a large area, as we saw in SARS1, MERS, HIV, Minks.
- No other spillovers (by reporting or by phylogeny)
- No intermediate host

Chinese and WIV behavior cancels out

What Evidence Supports Zoonosis?

Zoonosis only has:

- Prior that is based on little experience, and is overwhelmed by a high likelihood that WIV would leak a virus like SARS2
- “Dispositive” evidence that is always based on complex modeling that tries to obscure obvious evidence.
 - With many errors detected
 - Motivated researchers
 - Many assumptions, wherein one wrong assumption collapses the conclusion.

Complex models are not bad, just risky. And very suspicious when that's all you have. Reminiscent of UFO evidence.

Summary of intuitive explanation

Overall, we see that zoonosis has a lot of evidence for which they have no explanation other than an extraordinary coincidence, while the lab leak hypothesis explains all the evidence quite easily.

Best timeline of events under lab leak

In 2019, after DEFUSE rejection UNC and WIV started similar projects in parallel. GoF research has been growing exponentially, and the rejection of the proposal doesn't dampen the interest in the project's aims.

The UNC MERS cleavage paper

Someone in WIV used their database to screen for human ACE2 matches

unpublished viruses

Found a good match

Amino Acids identical to BANAL-52 in RBD

Added an FCS

- DEFUSE cleavage paragraph
- 12 nt clean insertion
- CGGCGG
- PAA

Best timeline of events under lab leak

Possibly checked n-glycans for higher infectivity

- DEFUSE n-glycans
- N-glycan change from day 1
- Study on n-glycan effect on infectivity in lung cells

The work was done in BSL-2

- Website
- Prior paper
- SZL interview

- Given SARS-CoV-2 infectivity, someone is likely to contract it at BSL-2
- An infection chain starts in Wuhan



Best timeline of events under lab leak

Reached the HSM (as lineage B), where it grew faster than the rest of Wuhan, forming the first early cluster

- All evidence against wildlife
- Lineage B
- The market model

Few people were aware of this (possibly only the researcher, their direct bosses etc.).

- WIV continued normally
- No whistleblowers or leaks
- No known WIV infections.
- Suppressing release of virus databases that contain the backbone.

When the pandemic started, no evidence of zoonosis could be found

- No intermediate host
- Many outbreaks following covide zero, but non from wildlife

Few senior Chinese officials know it's a leak and block investigation

- No contact tracing
- Manipulating WHO

Best timeline of events under zoonosis

A bat coronavirus with good human ACE2 binding jumped to an intermediate host

BANAL-52 is an example of such a virus

There it acquired an FCS through a clean insertion

- **CGG by chance, then duplication**
- **There was a Proline in there**

Best timeline of events under zoonosis

N-glycan removed due to selective pressure.

**Virus gets to Wuhan (via animal or person)
Two options:**

HSM is not the epicenter (more likely)

1. The animal infected someone in Wuhan with lineage A
2. It spread through Wuhan
3. Mutated to lineage B
4. That infected the market

HSM is the epicenter

1. An animal was brought to HSM. It was a single animal, not from a farm.
2. No other infections
3. It then infected someone with lineage A, he left the market and infected outside
4. It then infected someone who mutated to lineage B and spread in the market

The intermediate host is not found and did not infect anyone else anywhere

Parallel vs Serial | OR vs AND

Another way to look at it intuitively is to compare arguments that have multiple parallel lines of reasoning, to those that are a series of sub-claims

- The former is invalidated when all arguments fail. This is when we multiply probabilities, and can thus reach high certainty
 - after carefully checking for dependencies, using conservative estimates, and accounting for the possibility of mistakes.
- The latter is invalidated when one argument fails. This is when you must choose the lowest probability in the chain.
 - This is the best explanation for the evidence under the competing hypothesis. We have to meticulously look for the weakest link, and estimate that alone.

Parallel vs Serial | OR vs AND - Examples

Serial example - Pekar, Worobey models

- If any of these fail, the evidence is irrelevant:
 - Data collection was unbiased
 - Worobey uses check-in data
 - Pekar uses data with a mistake in the first A patient (the familial cluster)
 - Both use early cases with many biases towards HSM
 - Data was not filtered or manipulated
 - Both use early case data relying exclusively on a source that needs to cover-up a lab leak (if one happened).
 - Part 1 of the model has no errors
 - Worobey incorrectly classified Wuhan locations to superspreader locations.
 - Part 2 of the model has no errors
 - Worobey assigned $p=.034$ to a lineage case near the market based on a wrong assumption of linear relation with density, when in fact high density affects the exponential growth rate
 - ...
 - There is no bug in the code that implements the model
 - The Pekar erratum
 - The result was interpreted correctly
 - Pekar interpreted this to be a Bayes factor for multiple introductions, but forgot to account for the stage at which 2 lineages were created in the intermediate host
 - All of the opposing evidence must be false
 - proCoV2 18060 is a reversion
 - Connor is not Covid
 - Chen somehow has an HSM connection

Parallel example - FCS

- Any of these legs can fall and it's still very strong evidence
 - First in SARS-like viruses
 - Clean 12-nt insertion
 - CGGCGG / non-viral source



Meta-inference

Opposition's probabilistic inference mistakes

1. Strawmanning (not choosing the best explanations):

- 10,000 for the market, instead of thinking what is the best explanation it would form an early cluster
 - We happened to find very strong reasons, that coupled with the evidence against wildlife involvement make it a Bayes factor of 1
 - But even if we haven't, this issue is highly complex, involving human behavior, unreliable data, biology, politics. The best explanation under lab leak must be that there is some unknown reason.
 - This is what we did with the FCS, instead of claiming a 1 in a billion coincidence, even though there we have actual statistics on long insertions - much better certainty than virus dynamics in a city.
 - An example of a naive calculation of the FCS
 - 50 in millions insertions. 1:20000 for it happening
 - probability of falling exactly in the S1/S2 junction: 1:10000
 - probability of coming from some unknown non-viral source: 1:100
 - Total 1:2E10, just for this evidence.
- The more complex the issue is, and the less experience and statistics we have on it, the lower the Bayes factor that can be claimed.
- A number like 10,000 should never appear in a robust probabilistic analysis. It can only appear in scientific experiments that were carefully designed to remove all confounding variables.
- If you have a reliable 1000x factor in your analysis (and not as a result of hundreds of trials), there is no need for probabilistic analysis - that is the true hypothesis.

Meta-inference

Opposition's probabilistic inference mistakes

2. Forgetting dependencies:

- Multiplying $p=0.034$ for the two A cases being “near the market”
- It's probably wrong, with no specifics on one of the cases (that's a strawmanning mistake)
- Assuming it's true, there is a strong dependency between the two.
- The two patients are not independent events that can be multiplied. It could be simply because that area happened by chance to have more early infections, which would explain the two cases, as well as why HSM (and not another wet market) was the early cluster.

Meta-inference

Opposition's probabilistic inference mistakes

3. Not adjusting for filtering from multiple opportunities:

- Evaluating A20 as another lab worker going straight to the market.
 - By that time, late December, there are many cases in Wuhan. Only one of them needs to arrive in the market to deposit this sample.
 - That's likely to happen, but was given a 2000 factor.
- Evaluating probability of a single lab leak event,
 - The research was likely going on for months in BSL-2, giving plenty of opportunities for a leak.

Meta-inference

Opposition's probabilistic inference mistakes

5. Cherry picking - Ignoring evidence that doesn't fit, and evaluating only what does fit

- Evaluating “the market has A and B”, not evaluating why 16 out of 16 cases are B, and the only A is environmental with 2-3 mutations.
- Evaluating coincidence of “cluster where wildlife is”, but where are all the other spillovers near wildlife?
- Evaluating “one wildlife stall a bit more positive”, but not evaluating that there are no sick vendors, the reads are low, and the case pattern over time is completely random - exactly as expected if there was another driver other than wildlife e.g. the mahjong room.

Meta-inference

Opposition's probabilistic inference mistakes

6. Focusing on the low conditional probability for one hypothesis, rather than the ratio between the conditional probabilities of all hypotheses (the bayes factor).

- BBC reporting on WTC7 collapse ahead of time.
- Two polytomies 2 mutations apart
 - The evidence supports one introduction, but even if we assume two - having them 2 mutations apart, at the same time and location is indeed a low conditional probability event for lab leak, but it is even lower for zoonosis. The Bayes factor should increase lab leak, not decrease it.
- Evaluating the probability of finding a BANAL-52 like virus specifically.
 - Must use the same evidence in both hypotheses: i.e. also reduce zoonosis to choose BANAL-52-like from all the viruses with pandemic potential (after an FCS is added)
 - Interestingly, for nature the search space is wider, because it doesn't have to be an hACE2 match, like it is in DEFUSE.
- Same issue with the S1/S2 vs S2' choice



Meta-inference

Opposition's probabilistic inference mistakes

7. Double counting

- 1:99 prior, after building a model from scratch for emergence, which includes the odds of a leak)
- The reason for the two polytomies is that B got amplified in the market, which was already accounted for elsewhere. The Pekar model should have added the involvement of an amplification location, which would have resulted in a higher likelihood of encountering two polytomies.
 - “This pattern indicates a very common phenomenon in the spread of an infectious disease, known as an infection cluster or outbreak cluster. Infection clusters occur at a specific location with a sudden burst of infected cases during a short time interval”
 - <https://www.sciencedirect.com/science/article/pii/S1672022920300620>

Meta-inference

What Rootclaim does differently

In contrast we:

- **Look at all the evidence without cherry-picking.**
- **Steelman everything**
 - **We truly steelman, not give 100% instead of 50% as a steelman, and then assign factors of 1000x and 10000x for complex arguments involving human behavior and biology, based on unreliable data sources, with multiple points of failure, and lots of uncertainty on the exact details.**
- **Careful look for any missed dependencies and double counting**
- **Leave a lot of buffer for our own mistakes, so the conclusion is robust**

Contradictory high Bayes factors

Having high Bayes factors for and against an hypothesis is very unlikely. While you may find several >100x factors for the true hypothesis, finding one for a false hypothesis is by definition rare.

- **Example:**

- H1 - a die is biased to give 6 at 50%
- H2 - the die is biased to give 1 at 50%
- Throwing a six 5 times in a row is $1/32$ under H1 and $1/100000$ under H2.
- It is therefore very unlikely to have reliable reports of 5 ones in a row AND of 5 sixes in a row. One of these reports is false, or a result of filtering from 100,000s of throws.
 - More specifically, if we have clear reliable video evidence of 5 ones in a row, and a testimony “my friend said he rolled 8 sixes in a row”, the latter is likely unreliable despite claiming a far higher p-value. We need to choose the best explanation, which is the one with the higher conditional probability (e.g. a false report = 5% vs 8 sixes = $1/100,000,000$).

A high Bayes factor for the wrong hypothesis happens usually due to either a) multiple comparisons (you checked hundreds of things and one gave 100x), or b) unreliability (i.e. the best explanation is not the 100x but some mistake / manipulation which would happen at a far higher rate)

In our case the strong lab leak evidence is reliable and very unlikely under zoonosis, but the “strong” zoonosis evidence is all based on sharpshooter fallacies, garbage-in-garbage-out, manipulated data, modeling mistakes etc.

Simulating the debate to infinity

- The debate has reached a point where we're discussing lots of minor details.
 - This will only get worse as we continue
 - Perhaps we will be debating whether the company that prints the label used on the swabs that collected A20 has a 4% or 0.5% error rate, based on two different ways to count past cases.
- So much evidence and details may create a feeling that “nothing can be concluded”, and much more time is needed.
 - Especially as our opponent has proven much more efficient in detail collection
- The main insight of Rootclaim is that inference is far more important than evidence. In most disagreements, people continue hunting for evidence long after the existing evidence can already provide >90% confidence using robust probabilistic inference.
- Let's imagine what will happen if we spend more time on this.
- We'll find more evidence. Unlike Peter, we did not do a thorough review of studies and data, and mostly relied on quantifying the main arguments by each side. The little time we did invest yielded some interesting findings:
 - The study showing 4 of 6 positive civets in a random market during SARS.
 - Connor Reed
 - The factors making HSM a likely early cluster
 - We'll have the time to go through all the early cases analysis by Gilles Demaneuf and find why early cases outside the market are lacking.

Simulating the debate to infinity

- The opposition has already invested enormous resources in finding evidence to support their case, so we're unlikely to get any new significant evidence supporting zoonosis.
- We'll find more mistakes in the complex modeling studies.
 - After painstakingly going through all the details, we may find one that still has a strong finding (no argument in the series failed). But who's to say it won't be one of the lab leak studies we ignored, like the restriction sites one?
 - Note: Our decision not to do that is due to our experience with their low reliability. That may have created the wrong impression that we're going against some scientific consensus.
 - More cases where a factor claimed to be 200x turns into a 4x (Pekar polytomies model)
 - We may find more biases in the early cases and sequences
 - We may find more biases in the environmental sampling

Simulating the debate to infinity

- **And more reasons why their models are anyway irrelevant:**
 - **More reasons why HSM is a likely first cluster**
 - **More reasons why that corner in HSM was positive and why sampling was biased (remember these are people)**
 - **Just found the ventilation issue two days before sending the final presentation.**
 - **More reasons why 2 polytomies are reasonable (although not much weight left there given their two catastrophic mistakes)**
- **We'll find more problems in information submitted by the opposition**
 - **Describing Connor Reed as an unreliable story that is anyway connected to HSM**
 - **Cutting the stall data from the mahjong samples**
 - **A few more we may get into in responses**
- **Peter will find more Rootclaim mistakes that have little effect, because we already chose the best explanations and accounted for uncertainty.**
 - **The visitors among early HSM cases were regular, but still they're not involved in wildlife**
 - **There aren't many early clusters other than the market, but we already gave that a modest 2-4x due to our uncertainty there.**
 - **And also discovered why it is not interesting that it's an early cluster.**

Simulating the debate to infinity

- **We'll find underestimates in our conservative numbers:**
 - **A mistake in steelmanning FCS to 50. That's only the probability of some explanation existing, it would still be at most 10% that it would happen, so 500.**
 - **Not discounting for wildlife trade in Wuhan. 1.5% should be 0.5%**
 - **Not accounting for no more spillovers (only no intermediate host). Could potentially be very strong evidence once thoroughly investigated.**
- **There will be no new explanation for simple evidence that has already been thoroughly investigated from all directions.**
 - **Why the FCS looks that way,**
 - **Why it started in the most likely city in the world for a bat coronavirus lab leak (UNC wouldn't do it in BSL-2)**
 - **Why it looks like a DEFUSE product**
 - **Why there is no intermediate host or multiple real spillovers (not same location, time, 2 mutations apart)**

The Bottom Line

- **The evidence for zoonosis is surrounded by a lot of complexity and uncertainty. Further investigation may strengthen it or weaken it.**
- **The lab leak evidence is robust and won't change much.**
- **Since it's unlikely there will be strong evidence for mutually exclusive hypotheses, it is much more likely that as uncertainty is reduced, the zoonosis evidence will suffer more.**
- **There is no need to continue discussing the evidence. Looking at the pattern of the existing evidence provides a very clear picture of what is the most likely hypothesis.**