

The Huanan market was the epicenter of SARS-CoV-2 emergence

Michael Worobey^{1*}, Joshua I. Levy², Lorena M. Malpica Serrano¹, Alexander Crits-Christoph³, Jonathan E. Pekar^{4,5}, Stephen A. Goldstein⁶, Angela L. Rasmussen⁷, Moritz U. G. Kraemer⁸, Chris Newman⁹, Marion P. G. Koopmans^{10,11}, Marc A. Suchard^{12,13,14}, Joel O. Wertheim¹⁵, Philippe Lemey^{16,17}, David L. Robertson¹⁸, Robert F. Garry^{17,19,20}, Edward C. Holmes²¹, Andrew Rambaut²², Kristian G. Andersen^{2,23*}

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.

²Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA.

³W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA.

⁴Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093, USA.

⁵Department of Biomedical Informatics, University of California San Diego, La Jolla, CA 92093, USA.

⁶Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.

⁷Vaccine and Infectious Disease Organization, University of Saskatchewan, Saskatoon, SK S7N 5E3, Canada.

⁸Department of Zoology, University of Oxford, Oxford, UK.

⁹Wildlife Conservation Research Unit, Department of Zoology, The Reanati-Kaplan Centre, University of Oxford, Oxford, OX13 5QL, UK.

¹⁰Pandemic and Disaster Preparedness Centre Rotterdam/Delft, The Netherlands

¹¹Department of Viroscience, Erasmus University Medical Center, Rotterdam, Netherlands.

¹²Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA.

¹³Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA.

¹⁴Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA.

¹⁵Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA.

¹⁶Department of Microbiology, Immunology, and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium.

¹⁷Global Virus Network (GVN), Baltimore, MD, USA.

¹⁸MRC-University of Glasgow Center for Virus Research, Glasgow, G61 1QH, UK.

¹⁹Tulane University, School of Medicine, Department of Microbiology and Immunology, New Orleans, LA, 70112 USA.

²⁰Zalgen Labs, LCC, Germantown, MD, 20876 USA.

²¹Sydney Institute for Infectious Diseases, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia.

²²Institute of Evolutionary Biology, University of Edinburgh, King's Buildings, Edinburgh, EH9 3FL, UK.

²³Scripps Research Translational Institute, La Jolla, CA, USA.

*Corresponding author. Email: worobey@arizona.edu (MW), andersen@scripps.edu (KGA)

Abstract

Despite strong epidemiological links and the documented presence of SARS-CoV-2 susceptible animals, the role of the Huanan Seafood Wholesale Market in the COVID-19 pandemic remains controversial. Using spatial analyses we show that the earliest known COVID-19 cases diagnosed in December 2019 were geographically distributed near to, and centered on, this market. This distribution cannot be explained by high densities of elderly people at greater risk of symptomatic COVID-19. This pattern was stronger in cases without, rather than with, identified epidemiological links to the Huanan market, consistent with SARS-CoV-2 community transmission starting in the surrounding area. By combining spatial and genomic data, we show that both the two early lineages of SARS-CoV-2 have a clear association with the Huanan market. We also report that live mammals, including raccoon dogs, were sold at the market in late 2019 and geospatial analyses within the market show that SARS-CoV-2-positive environmental samples were strongly associated with vendors selling live animals. Together, these analyses provide dispositive evidence for the emergence of SARS-CoV-2 via the live wildlife trade and identify the Huanan market as the unambiguous epicenter of the COVID-19 pandemic.

One sentence summary: Geographical clustering of the earliest known COVID-19 cases and the proximity of positive environmental samples to live-animal vendors suggest that the Huanan Seafood Wholesale Market in Wuhan was the site of origin of the COVID-19 pandemic.

Introduction

On 31 December 2019, the Chinese government notified the World Health Organization (WHO) of an outbreak of severe pneumonia in Wuhan, Hubei Province of unknown etiology (1–4). Of the initial 41 people hospitalized with pneumonia by 2 January 2020 and later confirmed to be infected with a novel coronavirus – severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) – 27 (66%) were epidemiologically linked to the Huanan Wholesale Seafood Market (hereafter “Huanan market”) in the Jianghan District of central Wuhan (2, 5, 6). Importantly, initial diagnoses of pneumonia cases of unknown etiology in Wuhan hospitals between 18 and 29 December 2019 were free from ascertainment bias as they were made based on signs and symptoms before the Huanan market was identified as the shared risk factor for the preponderance of unexplained pneumonia cases (5). Indeed, it was the clinicians and administrators at the Hubei Provincial Hospital of Integrated Chinese and Western Medicine who noticed that four of the first seven cases they diagnosed worked at the market (5, 7). They promptly notified public health authorities of the potential importance of the Huanan market on 29 December (2, 7, 8). A review of all cases notified to a national disease surveillance system by hospitals in Wuhan as part of a joint WHO-Chinese study into the origins of SARS-CoV-2 (‘WHO mission report’) confirmed that a large proportion of the earliest known COVID-19 cases were linked to the market. This in-depth retrospective clinical case-finding concluded that 55 hospitalized COVID-19 cases with symptom onset in December 2019 had a link to the Huanan market, out of the 168 for whom exposure history to this market was available (33%) (9).

The first documented Coronavirus disease 2019 (COVID-19) cases in 2020 were focused in the residential community surrounding the Huanan market, including cases that were not epidemiologically linked to the market (5, 9, 10). Although not clear from the WHO mission report (9), animal surveys have shown that live mammals were on sale at the Huanan market and at three other markets in Wuhan, including species considered to be wildlife but which are often now farmed as well as wild-captured (11). Importantly, several of the animal species present at the Wuhan markets are known to be susceptible to SARS-related coronaviruses (SARSr-CoVs), such as SARS-CoV (hereafter ‘SARS-CoV-1’ for clarity) and SARS-CoV-2 (12–14). During the early stages of the COVID-19 pandemic, animals sold at the Huanan market were considered by Chinese public health and scientific authorities as the probable source of the unexplained pneumonia cases (15–22) (**File S1**). This suspicion led to the decision to close and sanitize the Huanan market on 1 January 2020 – in line with protocols in place for suspected zoonotic disease outbreaks to prevent further animal-to-human pathogen transmission – with environmental samples also being collected from vendors’ stalls (9, 15) (**File S1**).

Despite the clear epidemiological linkage of the earliest known COVID-19 cases to the Huanan market, this observation alone does not establish that the COVID-19 pandemic originated there. In particular, while many of the earliest COVID-19 cases were linked to the market, not all December cases had an identified link. While this is likely because the extent of asymptomatic/mild cases was high (23), obscuring transmission histories, it has led to conjecture that the Huanan market only represented human-to-human transmission chains, in the context of potentially high numbers of undocumented infections, and not a zoonotic spillover event. Despite the cases reported on 27 December, 2019 being a family cluster, leading the diagnosing physician to conclude that a human-to-human-transmissible virus was involved (5), initial reports from household surveys conducted by the Wuhan CDC reported a low secondary attack rate (3). However, the appearance of travel-associated COVID-19 cases early in January 2020 confirmed substantial under-ascertainment of cases in China, most likely due to the broad range in clinical presentations (24). In line with the expected confusion during the early stages of an emerging disease outbreak, incomplete and conflicting information has been released on some of the earliest cases. For example, a putative index case with first symptoms reported on 1 December, who evidently did not have exposure to the market, in fact did not fall ill with COVID-19 until late-December (9). Note, here we define ‘index case’ as the earliest documented COVID-19 case, as distinct from the ‘primary case’, which is the very first human infection(s) that likely remain(s) unidentified (25). Similarly, another putative index case, again with no known link to the Huanan market and reported to have had an onset on 8 December 2019, may not have become ill with COVID-19 until 16 December (5). This leaves a vendor at the Huanan market, who fell ill on 10 December, as the earliest index case with a clearly established onset date (5), although it is unlikely that this individual was the primary case.

In this study, we use epidemiological, genomic, commercial, photographic, location, social mobility and survey data – from a range of sources – to investigate the hypothesis that the COVID-19 epidemic in Wuhan began at the Huanan market. We conclude that the Huanan market was indeed the epicenter of COVID-19 emergence. We demonstrate that December 2019 COVID-19 cases were geographically distributed unexpectedly near to, and centered on, the Huanan market, irrespective of whether or not they worked at, had visited, or were knowingly linked to someone who had visited this market in late 2019. Furthermore, of those cases epidemiologically linked to the market, the overwhelming majority were specifically linked to the western section of the Huanan market, where most of the live-mammal vendors were located. Validating this spatial link between live animals and human COVID-19 cases, we show that positive environmental samples distinctly associated with animals clustered within a small area of the Huanan market where live mammal sales were most concentrated. We report that vendors at the Huanan market sold SARS-CoV-2-susceptible live mammals, including Asian raccoon dogs (*Nyctereutes procyonoides*; hereafter “raccoon dogs”), hog badgers (*Arctonyx albobularis*), and red foxes (*Vulpes vulpes*), in November and December 2019. Finally, we show that two early lineages of SARS-CoV-2 show a clear association with the Huanan market. Collectively, these results provide incontrovertible evidence that there was a clear conduit, via susceptible live mammals, for the zoonotic emergence of SARS-CoV-2 at the Huanan market towards the end of 2019. This spillover of a bat SARS-CoV via a transiently-infected conduit animal population mirrors the market-based origins of the first SARS virus, SARS-CoV-1 (26, 27) and underscores the high likelihood of further cross-species transmission associated with the live animal trade (28, 29).

Results

High-density clustering of early COVID-19 cases around the Huanan market

The residential locations of many early COVID-19 cases in Wuhan cluster around the Huanan market (5, 9, 10) including those from December 2019 who had neither worked at, nor recently visited, the Huanan market, or knowingly been in contact with anyone who had (5). Given this, and the fact that the Huanan market is the only site in Wuhan for which clear epidemiological evidence points to a putative origin, we sought to quantitatively investigate whether the early COVID-19 pandemic had an epicenter at this site.

We used descriptive and statistical analyses to assess the December cases' distance to, and centering on, the Huanan market. We based our spatial analyses on data summarized in the WHO mission report on the origin of the COVID-19 pandemic (9). This report documented 174 COVID-19 cases in Hubei province in December 2019, identified after careful examination of reported case histories. Although geographical coordinates of the residential locations of the 164 cases who lived within Wuhan city limits were not provided, detailed maps allowed us to reliably extract the latitude and longitude coordinates of 156 (see **Methods and Figs. S1-S11**). With missing cases most likely falling close to the Huanan market (see **Methods**), most of the following analyses are conservative with respect to hypothesis testing.

We investigated the densities of the case locations across Wuhan using a kernel density estimate (KDE) based on all 156 December 2019 COVID-19 cases for which we extracted location information (**Fig. 1A and B**). In contrast to the sections below where we statistically investigate the degree to which the December cases' distance to, and centering on, the Huanan market departs from expectation, here we explore the density of cases qualitatively. This analysis revealed that although cases in December 2019 occurred across a wide area, a majority clustered in central Wuhan near the west bank of the Yangtze River, with a high density near to, and surrounding, the Huanan market (**Fig. 1A**). We found that the relatively small area within central Wuhan with the highest case density is in the direct vicinity of the Huanan market, with the market falling within the bounds of the highest density area in the KDE (**Fig. 1B**), a result we found to be robust to KDE bandwidth (**Fig. S12**). This association was present irrespective of epidemiological contact history, as we found that December cases with no linkage to the Huanan market clustered just as strongly around the market (**Fig. 1C**). These findings strongly suggest that community transmission of SARS-CoV-2 began at, or near, the Huanan market.

Next, we investigated whether the clustering of December cases contrasts with clustering of COVID-19 cases during January and February 2020, when SARS-CoV-2 was widely disseminated across Wuhan (**Fig. 1D, E**). These later case locations have been mapped using cell phone location data from individuals using a COVID-19 app on Sina Weibo, a social media platform in China (30). Our analysis of these data revealed that, unlike the clustering observed in December (**Fig. 1A-C**), the spatial distribution of COVID-19 cases in January and February closely mirrored areas within central Wuhan with high population densities of older people, >60 years of age (30) (**Fig. S13**). These results correspond closely with the official case counts from China's national infectious disease surveillance system at the neighborhood level (31).

Finally, we compared the spatial pattern of December COVID-19 cases in Wuhan with population density patterns in the city from data we acquired from worldpop.org. As with the Weibo COVID-19 cases (**Fig. 1D**), the population of Wuhan is highly dense on both the west and east bank of the Yangtze River in central Wuhan (**Fig. 2A**). This is in contrast to the observed clustering of COVID-19 cases in December 2019, which are highly concentrated on the west bank (**Fig. 1A-C**). Combined, these analyses reveal how the spatial patterns of COVID-19 cases shifted between late 2019, when the outbreak began (32), and early 2020, as the epidemic spread widely across Wuhan: December COVID-19 cases had a clear association to the Huanan market that was unrelated to population density or demographics, which is unlike the pattern observed during later stages of the epidemic in January and February.

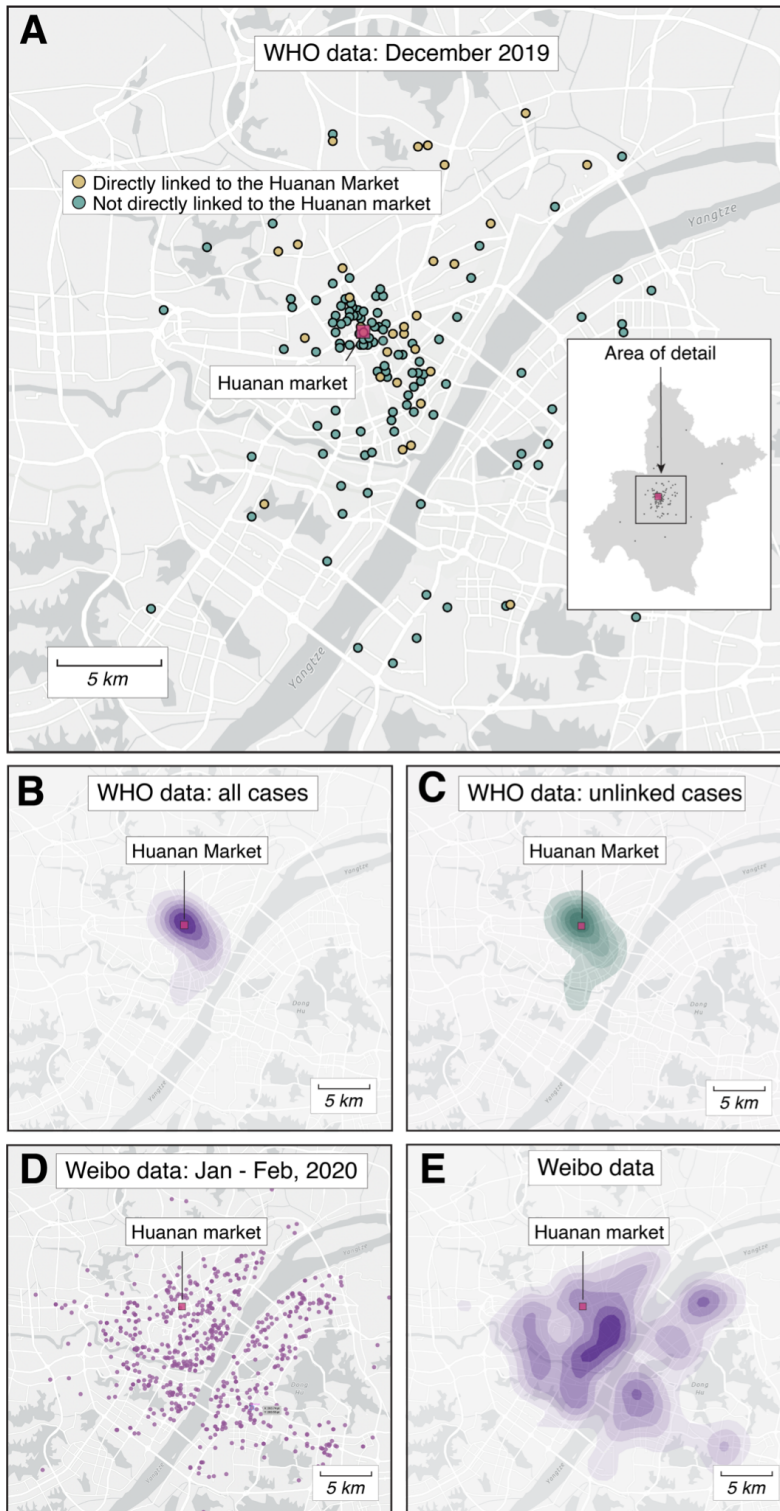


Figure 1. Spatial patterns of COVID-19 cases in Wuhan in December 2019 and January-February 2020. **A.** Locations of the 156 cases we extracted from the WHO mission report (9). Those reported to have had a history of exposure at the Huanan market are shown in yellow and those not found to have had a link indicated in teal. The inset shows a map of Wuhan showing the area detailed in the main panel, and the December 2019 case indicated with gray dots. (No cases are obscured by the inset.) In both the inset and the main panel the location of the Huanan market is indicated with a red square. **B.** Kernel density estimate using all 156 COVID-19 cases locations from December 2019. **C.** Kernel density estimate using the 123 COVID-19 cases locations from December 2019 that were unlinked to the Huanan market. **D.** Locations of 737 COVID-19 cases from Weibo data dating to January and February of 2020. **E.** Kernel density estimate using 737 COVID-19 cases locations from Weibo data.

Early COVID-19 cases occurred near the Huanan market

Having investigated the spatial patterns of early COVID-19 cases in this way, we next sought to statistically investigate two related but distinct questions: (1) did COVID-19 cases in December 2019 live unexpectedly near to the Huanan market, and if so, (2) were these locations also centered on the market? To do so, we compared the locations of the COVID-19 cases in December of 2019 with null models based on either area-specific residential population densities in Wuhan (see below), or the COVID-19 case density later in the epidemic (Weibo data). The ‘centered on’ question is important because it would be possible for cases to fall close to a particular location but nonetheless be centered on a different location: if the earliest cases resided somewhat close to the Huanan market but were systematically clustered

around some other site (**Fig. 2B**), that would be strong evidence against an origin of the Wuhan outbreak at the Huanan market. We are in effect asking whether the spatial patterns of cases in December reflected an outbreak that was already cryptically widespread in Wuhan, such that any association with the Huanan market would merely reflect a workplace spreading event rather than the site of origin.

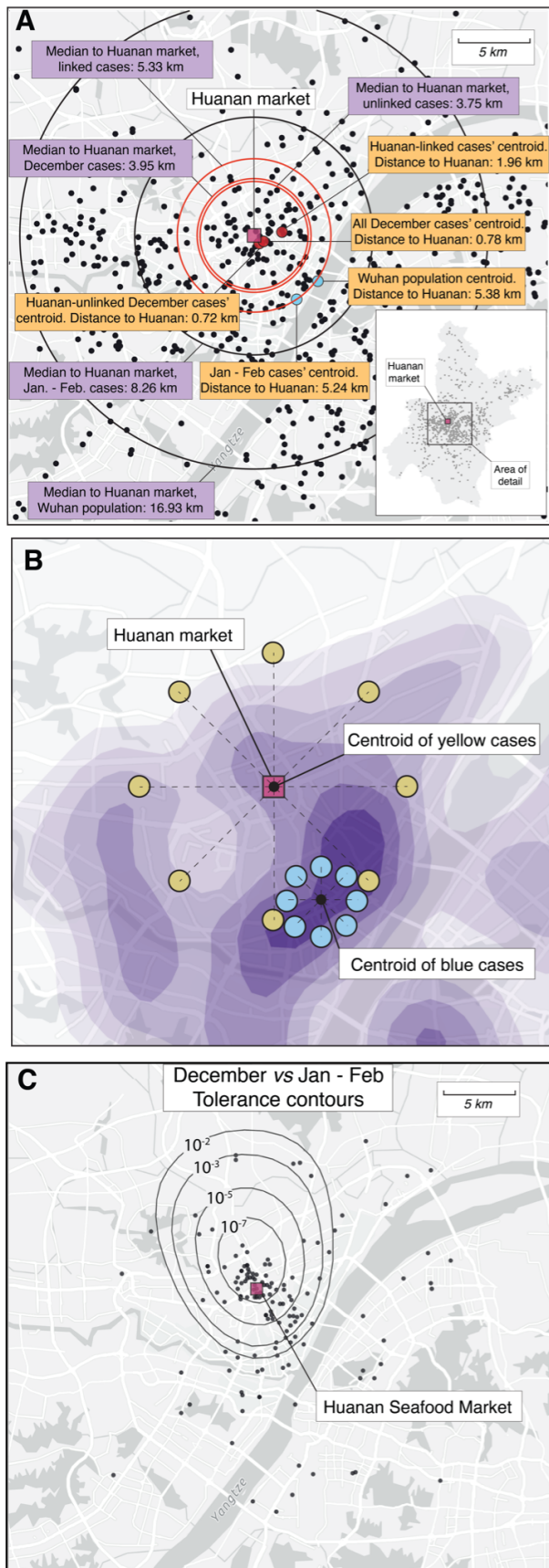


Figure 2. Spatial analyses findings. **A.** Inset: Map of Wuhan city, with gray dots indicating the center of the 100m by 100m square visited in the sampling process (1000 samples total). The area of detail depicted in the main panel is highlighted and the Huanan market indicated. Dots indicate 1000 sampled locations in Wuhan city from the worldpop.org population density data (weighted random sample). Main panel: Black dots indicate locations in central Wuhan sampled from worldpop.org. The median distance between points sampled from the population density and Huanan market is indicated by the larger black circle (radius $r = 16.93$ km), and the median distance between the Weibo January-February COVID-19 locations and the Huanan market by the smaller black circle ($r = 8.26$ km). These larger distances contrast with the median distances between COVID-19 cases in December and the Huanan market, in red circles, with the following data groupings moving from the largest to smallest: cases epidemiologically linked to the Huanan market ($r = 5.33$ km); all December cases ($r = 3.95$ km); cases unlinked to the Huanan market ($r = 3.75$ km). The centroids of the Wuhan population density data and the Weibo location data from January and February are represented by blue circles, and the distance between each centroid and the Huanan market is shown. The centroids of the December case locations are shown by red circles and the distance from the Huanan market to the ‘all cases’, ‘linked cases’ and ‘unlinked cases’ centroids are indicated in the respective text boxes. Note that the black and red concentric circles all indicate the median distance between a set of individual locations and the Huanan market (how far they are from the market). The distances from the various centroids to the Huanan market, on the other hand, are indicated in the text box associated with each centroid (how centered a set of locations is on the market). **B.** Schematic showing how cases can be near to, but not centered on, a specific location. Part of the KDE of the Weibo COVID-19 case location data from January and February is shown (in purple; see **Fig. 1E**), with two hypothetical spatial distributions of cases from December shown (yellow and blue circles). In both hypothetical distributions, cases are approximately the same median distance from the Huanan market (4 km): The blue cases and distribution show how cases that tend to be quite near the Huanan market can nevertheless not be centered on it, as the yellow cases are (see black circles at intersection of radii for each distribution). A pattern like the blue cases, which are systematically centered around a point (the centroid) quite far from the Huanan market – and indeed fall in an area with a high density of elderly people and observed cases in January and February – would suggest that the epicenter of the outbreak was not within the Huanan market, while a centering pattern like the yellow cases would be consistent with a Huanan market origin. **C.** Tolerance contours based on relative risk of COVID-19 cases in December, 2019 versus cases in January – February 2020 from the Weibo data. The dots indicate approximate residential locations of cases with onset in December of 2019. The contours represent the probability of observing the density of cases in December within the bounds of that contour if the December cases had been drawn from the same spatial distribution as the Weibo cases from January and February, 2020.

To address the ‘distance to’ question we used the same worldpop.org population density data from Wuhan, and location-based Weibo COVID-19 case data from January and February 2020 (30) as above. This was to test whether the distances of the December COVID-19 cases were shorter than expected compared to a sampling of the residential locations of the Wuhan population as a whole or compared to the locations of people with COVID-19 in the city in early 2020. To do so, we compared the locations of the COVID-19 cases in December of 2019 with null models based on either the worldpop.org or Weibo data.

We found that the weighted median distance of Wuhan’s population to the Huanan market (*i.e.*, the median distance between residential location and the Huanan market for all residents of Wuhan) was 16.93 km, whereas the median distance to the Huanan market of the Weibo reported COVID-19 case locations in January and February was 8.26 km. In comparison, we found that the median distance observed between all reported December COVID-19 cases (‘all cases’) and the Huanan market was just 3.95 km, that for only those cases linked to the Huanan market (‘linked cases’) was 5.33 km, and that for only those cases without a link to the market (‘unlinked cases’) was 3.75 km (**Fig. 2A and Table 1**).

Table 1. Median distances and centroids to the Huanan market.

Data	Null (median distance to Huanan market, centroid distance to Huanan market)	Statistic	
		Median distance to Huanan market	Centroid distance to Huanan market
156 cases, both linked and not linked to the Huanan market	Worldpop.org (16.93 km, 5.37 km)	3.95 km ($p < 0.001$)	0.78 km ($p = 0.002$)
	Weibo (8.26 km, 4.82 km)	3.95 km ($p < 0.001$)	0.78 km ($p = 0.002$)
123 cases, not linked to the Huanan market	Worldpop.org (16.93 km, 5.37 km)	3.75 km ($p < 0.001$)	0.72 km ($p = 0.002$)
	Weibo (8.26 km, 4.82 km)	3.75 km ($p < 0.001$)	0.72 km ($p = 0.002$)
33 cases, linked to the Huanan market	Worldpop.org (16.93 km, 5.37 km)	5.33 km ($p < 0.001$)	1.96 km ($p = 0.020$)
	Weibo (8.26 km, 4.82 km)	5.33 km ($p < 0.007$)	1.96 km ($p = 0.064$)
9 cases, linked to the Huanan market, lineage B	Worldpop.org (16.93 km, 5.37 km)	9.11 km ($p = 0.047$)	1.62 km ($p = 0.015$)
	Weibo (8.26 km, 4.82 km)	9.11 km ($p = 0.659$)	1.62 km ($p = 0.049$)
2 cases, not linked to the Huanan market, lineage B	Worldpop.org (16.93 km, 5.37 km)	1.90 km ($p = 0.002$)	1.63 km ($p = 0.015$)
	Weibo (8.26 km, 4.82 km)	1.90 km ($p = 0.005$)	1.63 km ($p = 0.050$)
2 cases, not linked to the Huanan market, lineage A	Worldpop.org (16.93 km, 5.37 km)	1.17 km ($p < 0.001$)	0.95 km ($p = 0.003$)
	Weibo (8.26 km, 4.82 km)	1.17 km ($p < 0.001$)	0.95 km ($p = 0.014$)

We generated a null expectation for the geographical distribution of COVID-19 cases in December 2019 using Wuhan’s population density distribution from worldpop.org. We randomly sampled pixels of the data, weighting the data such that a region with a higher population density was proportionally more likely to be sampled than a region with lower density (see **Methods**). We found that it would be exceedingly unlikely to observe case locations in December so close to the Huanan market if they simply reflected population densities across Wuhan and the virus had been cryptically circulating widely throughout Wuhan’s population prior to its detection via cases linked to the Huanan market. This was true whether we considered all COVID-19 cases (median 3.95 km; $p < 0.001$), linked cases (median 5.33 km; $p < 0.001$), or unlinked cases with no epidemiological association to the Huanan market (median 3.75 km; $p < 0.001$; **Fig. 2A**).

Next, we generated null distributions using Weibo case location data from January and February and obtained similar results. We found that all December cases, as well as both unlinked and linked cases, lived closer to the market than expected by chance ($p < 0.001$, $p < 0.001$ and $p < 0.007$, respectively). Given potential imprecision in location extraction in the December COVID-19 case data, we performed sensitivity analyses by resampling each case within a circle of 1.0 km radius from its center point and found that our results were robust to this level of imprecision – which is far greater than possibly could have arisen during our location extraction process (see **Methods and Table S1**). In addition, we assessed the impact of missing data by uniformly sampling additional locations from the 25% KDE for all 156 December cases (see **Methods**), and found our results to be robust to this (**Table S1**). These results confirm those from the previous section: that early COVID-19 cases were highly focused at, or near to, the Huanan market. They also demonstrate that the clustering of early cases is best explained by proximity to the market and not population density or demographics. Indeed, we found that the density of elderly people at higher risk of COVID-19 symptoms were considerably higher elsewhere in Wuhan (30, 33) (**Figs. S13 and S14**).

Early COVID-19 cases were geographically centered on the Huanan market

As mentioned above, if the COVID-19 pandemic began at the Huanan market, then December cases should not just occur near the market but would also be expected to be centered on it (**Fig. 2B**). To investigate this, we calculated the centroid of all December COVID-19 cases, as well as December cases with or without a direct epidemiological link to the Huanan market. In agreement with our previous analyses, we found that case centroids were remarkably close to the market, with the ‘all cases’ centroid located 0.78 km away from the Huanan market, the ‘linked’ cases centroid 1.96 km away, and the ‘unlinked cases’ centroid just 0.72 km away (**Fig. 2A**). In comparison, using worldpop.org data, we calculated the centroid of Wuhan’s population to be located 5.37 km away from the Huanan market, while the centroid of Weibo COVID-19 cases from 2020 was 4.82 km from the market (**Table 1**).

Next, we tested whether it would be unexpected to observe such a degree of centering on the Huanan market if the virus had been cryptically spreading widely within Wuhan prior to its detection at the Huanan market. We therefore obtained a weighted random sample of 1,000 points across Wuhan from the worldpop.org population data to generate a set of density-informed potential emergence sites. Each point marks a potential early center of spread of SARS-CoV-2 in the city, analogous to the centroids inferred for the observed COVID-19 case data. (**Fig. 2A**, black dots). Using this null distribution, we found that distances between the Huanan market and the centroids of all December cases, cases linked to the Huanan market, and cases unlinked to the market were shorter than expected based on Wuhan’s population density distribution ($p = 0.002$, $p = 0.02$ and $p = 0.002$, respectively; **Table 1**).

Next, we repeated the same analysis using the Weibo COVID-19 cases data from January and February 2020 and found that these December data centroids were also significantly closer to the Huanan market than expected by chance ($p = 0.002$, 0.064 and 0.002 for ‘all’, ‘linked’, and ‘unlinked’ cases, respectively). The Weibo data reveal that by January and February many cases were observed in highly populated areas of the city, in particular areas with a high density of older people (**Figs. S13**). Notably, these areas are distinct from the area of highest COVID-19 case density, surrounding the Huanan market, in December 2019 (**Fig. 1**). Similar to our ‘distance’ analyses from the previous section, we found that these results were robust to missing data and uncertainty in location extraction from the maps in the WHO mission report (9) (see **Methods and Table S1**).

To investigate the location, shape, and spatial extent of the region of increased density near the Huanan market in December 2019, as well as whether there were any other regions of Wuhan with increased case density, we performed a spatial relative risk analysis (**Fig. 2C and see Methods**). We tested for locations where the densities of WHO-reported cases from December (both linked and unlinked) were significantly greater than expected given a null distribution based on the Weibo case data from January and February. In line with our previous findings, we found that the Huanan market is located within a well-defined, isolated region with an aberrantly high density that is expected to be observed in less than 1 in 10 million samplings of the null distribution (relative risk analysis results in **Fig. 2C**, see null distribution in **Fig. 1D**). We did not find any other regions in Wuhan of significantly increased density. Hence, it was not only that case that

December COVID-19 cases were close to, and centered on, the Huanan market; the density of cases in the immediate vicinity of the market was also highly statistically significant, and this was the only such area, showing that the market was the epicenter of SARS-CoV-2 emergence and early COVID-19 disease transmission.

Both SARS-CoV-2 lineages ‘A’ and ‘B’ were geographically associated with the Huanan market

Two distinct lineages of SARS-CoV-2 (named ‘A’ and ‘B’ (34)) have co-circulated globally since early in the COVID-19 pandemic (35) and genome sequences from lineage B, but not lineage A, have been reported from cases and/or environmental samples associated with the Huanan market (**Table S2**) (9). To investigate whether these early SARS-CoV-2 lineages were closer to, and more centered on, the Huanan market than expected if the virus had been circulating widely in Wuhan in December 2019, we performed an integrated analysis of early SARS-CoV-2 genomic data with case location data.

We linked SARS-CoV-2 genome lineage information (‘A’ or ‘B’) and location information for thirteen COVID-19 cases with onset in December 2019 – eleven of lineage B, of which nine were linked to the Huanan market and two were unlinked, and two of lineage A, both of which were unlinked to the market. With the exception of one lineage A case, for which location data was available on where the case stayed in the five days preceding symptom onset (36), locations were residential sites we recovered from the WHO mission report and our geocoding process (see **Methods and Figs. S1-S10**). Because no lineage A genome sequences have been reported from human cases associated with the Huanan market, we hypothesized that these lineage A genomes may not show a significant geographical association with the location of the market. Moreover, based on the results above showing that cases that were epidemiologically linked to the Huanan market had a weaker geographic association with the market based both on ‘distance to’ and ‘centering on’ analyses than cases that were unlinked to the market, we hypothesized that lineage B cases linked to the Huanan market may be less near to, and also less centered on, the market than lineage B cases that were unlinked to the market.

Our analysis revealed that the nine Huanan market-linked lineage B cases were a median of 9.11 km away from the Huanan market, marginally closer than expected compared to the Wuhan population density null distribution (median = 16.93 km; $p = 0.045$), but not closer than the Weibo cases from 2020 were to the Huanan market (median = 8.26 km; $p = 0.659$) (**Table 1**). Strikingly, however, our analyses showed that the two unlinked lineage B cases were much closer to the Huanan market than expected (median = 1.90 km) compared to both population density and Weibo case null expectations ($p = 0.002$ and 0.005 , respectively). The centroids of both the nine market-linked and two market-unlinked lineage B cases were closer to the Huanan market than expected compared to both Wuhan’s population density patterns (whose centroid-to-market distance was 5.37 km) ($p = 0.015$ and $p = 0.015$, respectively) and to the spatial distribution of cases in early 2020 (whose centroid-to-market distance was 4.82 km) ($p = 0.049$ and $p = 0.050$, respectively; **Table 1**).

The lineage A viruses in our analysis are the two earliest known lineage A genome sequences (see **Methods**), both from cases epidemiologically unlinked to the Huanan market, making them directly comparable to the two Huanan market-unlinked lineage B viral genomes mentioned above. Remarkably, we found their median distance to the market was just 1.17 km, much shorter than expected based on the null distribution from Wuhan’s population density patterns (median = 16.93 km; $p < 0.001$) and also much shorter than Weibo case locations from January and February of 2020 to the market (median = 8.26 km; $p = 0.002$; **Table 1**). We also note that the median distance of 1.17 km of the lineage A genomes to the market was less than that of the unlinked lineage B genome cases, which was 1.90 km, and much less than that of the Huanan market-linked lineage B cases, which was 9.11 km.

Similarly, we found that – with a centroid-to-market distance of 0.95 km – the lineage A genome cases were more centered on the Huanan market than expected either from Wuhan’s population density patterns (centroid-to-market distance = 5.37 km, $p = 0.003$) or from Weibo COVID-19 case location (centroid-to-market distance of 4.82 km, $p = 0.014$; **Table 1**). Indeed, with a centroid-to-market distance of 0.95 km, we found that the lineage A genomes’ centroid was closer to the Huanan market than that of the nine market-linked lineage B genomes (1.62 km) and that of the pair of unlinked lineage B genomes (1.63 km). In sum, these analyses – unlike what we had hypothesized – show that at an early

time point in the epidemic both the ‘A’ and the ‘B’ lineages of SARS-CoV-2 were geographically associated with the Huanan market. These findings suggest that both lineages may have spilled over at the Huanan market during the beginning stages of the COVID-19 pandemic in Wuhan, with the limited geographic dispersal of lineage A compatible with a more recent zoonotic event.

SARS-CoV-2-susceptible live mammals were sold at the Huanan market during November and December 2019

In addition to selling fresh and frozen seafood, poultry, and other commodities, the Huanan market was among a handful of markets in Wuhan that consistently sold live, wild-captured or farmed wildlife in the years and months leading up to the COVID-19 pandemic (11). An investigation co-conducted by one of us (C.N.) prior to the detection of the pandemic (from May 2017 to November 2019) (11) found that 38 wildlife species, including 31 protected or illegal-to-trade species, were sold by live-animal vendors in Wuhan. That study reported wild-caught and farmed wildlife species were sold at the Huanan market, the Baishazhou market, the Dijiao outdoor pet market and the Qiyimen live animal market, but it did not report sales by specific market or specific month. We build on that work here by breaking down which mammalian species were sold live at the Huanan market during the crucial months of November and December 2019. It was within this two-month window that separate cross-species transmission events, leading to SARS-CoV-2 lineages A and B, are estimated to have occurred (companion paper by [Pekar et al.](#)).

Table 2. Live mammals traded at the Huanan market, November and December 2019

Species (Susceptibility*)	Family (Susceptibility*)	Order (Susceptibility*)	Observed at Huanan market, November 2019	Observed at Huanan market, December 2019**
Raccoon dog (<i>Nyctereutes procyonoides</i>) (Y)	Canidae (Y)	Carnivora (Y)	Y	Y
Amur hedgehog (<i>Erinaceus amurensis</i>)	Erinaceidae	Eulipotyphla	Y	ND***
Hog badger (<i>Arctonyx albogularis</i>) (Y)	Mustelidae (Y)	Carnivora (Y)	Y	ND
Asian badger (<i>Meles leucurus</i>)	Mustelidae (Y)	Carnivora (Y)	Y	ND
Chinese hare (<i>Lepus sinensis</i>)	Leporidae (Y)	Lagomorpha (Y)	Y	ND
Chinese bamboo rat (<i>Rhizomys sinensis</i>) (Y)	Spalacidae (Y)	Rodentia (Y)	Y	ND
Malayan porcupine (<i>Hystrix brachyura</i>)	Hystricidae	Rodentia (Y)	Y	Y
Chinese muntjac (<i>Muntiacus reevesi</i>)	Cervidae (Y)	Artiodactyla (Y)	Y	Y
Marmot (<i>Marmota himalayana</i>)	Sciuridae	Rodentia (Y)	Y	Y
Red fox (<i>Vulpes vulpes</i>) (Y)	Canidae (Y)	Carnivora (Y)	Y	Y
Siberian weasel (<i>Mustela sibirica</i>)	Mustelidae (Y)	Carnivora (Y)	N****	ND
Pallas's squirrel (<i>Callosciurus erythraeus</i>)	Sciuridae	Rodentia (Y)	N	ND
Masked palm civet (<i>Paguma larvata</i>) (Y)	Viverridae (Y)	Carnivora (Y)	N	ND
Coypu (<i>Myocastor coypus</i>)	Echimyidae	Rodentia (Y)	N	ND
Mink (<i>Neovison vison</i>) (Y)	Mustelidae (Y)	Carnivora (Y)	N	ND
Red squirrel (<i>Sciurus vulgaris</i>)	Sciuridae	Rodentia (Y)	N	ND
Wild boar (<i>Sus scrofa</i>) (Y)	Suidae (Y)	Artiodactyla (Y)	N	ND
Complex-toothed flying squirrel (<i>Trogopterus xanthipes</i>)	Sciuridae	Rodentia (Y)	N	ND

*Based on live susceptibility findings, serological findings, or ACE2 binding assays. See Table S3 for details and associated references.

**From photographic evidence from 3 December, 2019. (See Supplemental Materials.)

***ND = no data.

****Animals listed as “No” were, however, present at Wuhan markets during the 2017-2019 study period (11)v



Figure 3. Photographs from inside the Huanan market. A-C: photographs taken by a concerned citizen on 3 December, 2019, posted on Weibo and reported by CNN (37). **D, E:** photographs taken by one of us (E.C.H.) in the western section of the Huanan market on 29 October, 2014. Note that the raccoon dogs appear to be local, wild-caught common raccoon dogs rather than farmed raccoon dogs and that their plush coats are consistent with those observed in the winter. Note the red fox in the top right of panel **B**.

Based on data collected during the earlier study (11), we report that (1) multiple mammalian species that are susceptible to SARS-CoV-2 – and thus plausible intermediate hosts of its progenitor viruses – were sold live at the Huanan market in November of 2019 (Tables 2 and S3), and (2) despite a general slow-down in live animal sales during the winter months in Wuhan markets, raccoon dogs were consistently available throughout the year, including at the Huanan market (Tables 2 and S3). Raccoon dogs are noteworthy among the live mammal species traded at the Huanan market in November 2019 because they were associated with the emergence of SARS-CoV-1 (26) and have been shown to be both susceptible to infection with SARS-CoV-2 and capable of transmitting the virus (12–14). Other species for which there is evidence of SARS-CoV-2 susceptibility were also present at the Huanan market in November (Tables 2 and S3): hog badgers (*Arctonyx albogularis*), Chinese bamboo rats (*Rhizomys sinensis*) and red foxes (*Vulpes vulpes*), the latter of which have been shown to shed SARS-CoV-2 (38). In addition, species were present that belong to mammalian families or orders that are also susceptible to SARS-CoV-2: Asian badgers (*Meles leucurus*), Malayan porcupines (*Hystrix brachyura*), Chinese

hares (*Lepus sinensis*), Chinese muntjac (*Muntiacus reevesi*), and marmots (*Marmota himalayana*) (**Tables 2 and S3**). Hence, multiple candidate species were present within the Huanan market, although we cannot be certain which was the most likely intermediate species for SARS-CoV-2. Notably, other plausible intermediate host species (based on susceptibility data; **Tables 2 and S3**) were not observed live at the Huanan market in November 2019, including masked palm civets (*Paguma larvata*), American mink (*Neovison vison*), and wild boar (*Sus scrofa*). However, since animal surveillance was not continuous during the month of November (11), it is possible that these species could still have been present at times other than when observations were made, and therefore cannot be completely ruled out as putative intermediate hosts.

Confirmatory evidence of live mammals for sale at the Huanan market also comes from previously reported photographs dated to late 2019 and first posted on Weibo on 3 January, 2020 (see **Supplementary Text** and (37)) (**Fig. 3A-C**). We found that all of the live mammal species photographed in early December – raccoon dogs, Malayan porcupines, Chinese muntjacs, marmots and red foxes – are on the list of those on sale in November 2019 (**Tables 2 and S3**). Although we cannot independently verify their provenance, a CNN reporter spoke directly with the source of the photographs and obtained copies of the image files and shared them with us. The source stated to the reporter, who communicated to us (see **Supplementary Text**), that the photographs were taken at the Huanan market on 3 December 2019. We note that the floor tiles, floor drain grates, and concrete wall blocks in them are virtually identical to those observed by one of us (E.C.H.) in the western section of the Huanan market in October 2014 (**Fig. 3D, E**), corroborating that they were taken at the Huanan market. This photograph of caged raccoon dogs (**Fig. 3E**), albeit from five years before the emergence of SARS-CoV-2 (39), corroborates the sale of SARS-CoV-2-susceptible wildlife on the western side of the market at a stall location we found was selling live mammals illegally in late 2019 (**Table S2**).

Following the first cases of COVID-19 in Wuhan being linked epidemiologically to the Huanan market, public health officials closed and disinfected the market on 1 January 2020 to prevent further animal-to-human infections (**File S1**). Some 457 samples from 188 individual animals corresponding to 18 mammal species were screened for active SARS-CoV-2 infection via qRT-PCR from “*within and outside Huanan Market*”, and no positive SARS-CoV-2 samples were identified (9). Moreover, on the order of 80,000 samples from mammals across China were tested for SARS-CoV-2, yielding no positive findings (9). Crucially, most of these ~80,000 were from (1) species that do not appear to be susceptible to SARS-CoV-2, such as chickens and cattle, (2) archival samples collected well before the beginning of the COVID-19 pandemic, (3) samples collected in far-flung regions of China, or (4) samples collected in captive animals not part of the supply chain of live mammal markets in Wuhan. They are therefore of little to no relevance (see **Supplementary Text**).

In fact, the WHO mission members were told that no unlicensed or live-trapped wild animals had been for sale at the Huanan market and that “*no verified reports of live mammals being sold around 2019 were found*” (9). Notably, none of the live (known to be susceptible) mammals from species we identify here as present at the Huanan market in November and/or December 2019 have been reported to have been tested for evidence of SARS-CoV-2 infection. The only live mammals ‘from’ the market among the 188 appear to have been animals such as stray cats, dogs, snakes, rabbits, and mice. Indeed, of the 188 animals tested, 167 were those species or hedgehogs, pigs, chickens, salamanders, crocodiles, turtles, fish and sheep (9). Only 21 individuals from traded mammalian species likely susceptible to SARS-CoV-2 infection were tested – but these were dead, of unknown procurement date, and refrigerated or frozen, and therefore highly unlikely to be a source of human infection: six “*bamboo rat*”, six “*muntjac*”, six “*badger*”, two “*wild boar*” and one “*weasel*” (9). Some wild mammals were screened (negative) for active SARS-CoV-2 infection by qRT-PCR, but none were sampled from Hubei province. In addition, bamboo rats, porcupines, and wild boar from farms in Hubei supplying the Huanan market, sampled during February and March 2020, also showed no sign of active infection (9). Not a single raccoon dog from the Huanan market, live or dead, was reportedly tested for SARS-CoV-2, nor were those from the wild populations or the farms in Hubei that supplied them to the Huanan market (9). Unfortunately, to our knowledge, no serological testing was conducted on animals from within the market or from farms supplying it, nor on farm workers or animal traders in those supply chains. Indeed, both the Huanan market and the farms in Hubei province that supply it (40)

were rapidly shut down before such sampling occurred. Hence, it is apparent that by the time the Huanan market was closed on 1 January, 2020, and animal sampling at the market began, the SARS-CoV-2-susceptible live mammals that we know had been on sale there in the preceding months (**Tables 2 and S3**) were no longer present (9).

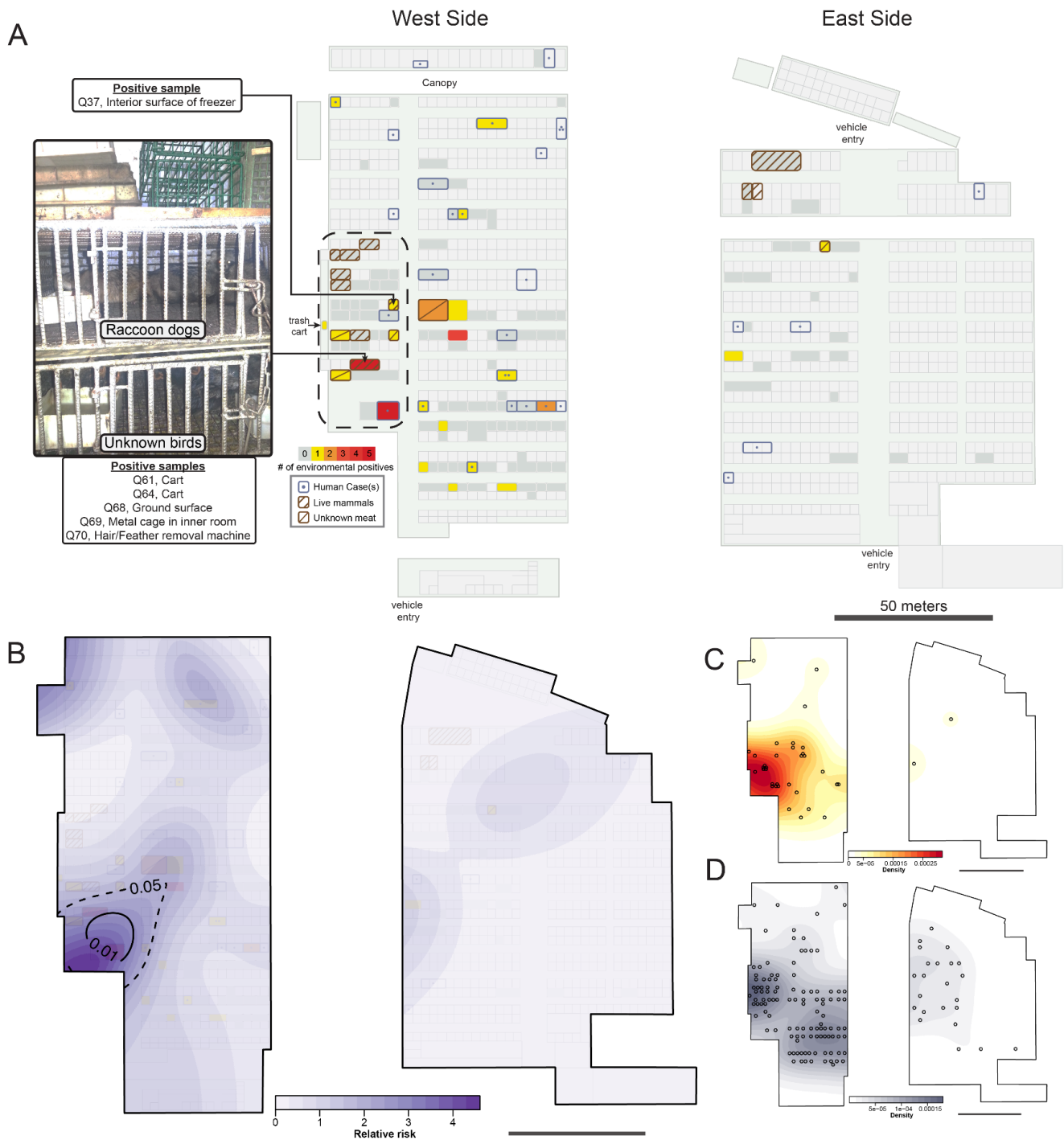


Figure 4. Map of the Huanan Wholesale Seafood Market. A. Aggregated environmental sampling and human case data from Huanan Market. Captions (left) describe the types of SARS-CoV-2 positive environmental samples obtained from known live animal vendors. Image (left) of raccoon dogs in a metal cage, on top of caged birds, taken in business with five positive environmental samples. Rectangle with dashed outline is used to denote the ‘wildlife’ section of the market. **B.** Relative risk analysis of positive environmental samples. Tolerance contours enclose regions with statistically significant elevation in density of positive environmental samples relative to the distribution of sampled stalls. **C.** Distribution of positive environmental samples. Sample locations (centroid of corresponding business) and quantity are shown as black circles. **D.** Distribution of businesses investigated with environmental sampling, with sampling locations shown as black circles. See Table S6 for details on environmental samples that were SARS-CoV-2-negative.

SARS-CoV-2-positive environmental samples were associated with live animal sales at the Huanan market

The analyses above show that the Huanan market was the epicenter of SARS-CoV-2 emergence and that susceptible animals were sold during the crucial months of November and December 2019. It has previously been reported that environmental samples from the Huanan market tested positive for the presence of SARS-CoV-2 and that COVID-19 cases were primarily located in the western part of the market (9). To investigate environmental sample positivity and potential association with live animal sales within the Huanan market, we analyzed a data set from the official Chinese Center for Disease Prevention and Control (CCDC) report dated 22 January 2020 (**Table S4 and File S1**). This report was described in Chinese News media in January and February 2020 (15, 16, 18, 19) and made publicly available in June 2020 (41). A total of 585 environmental samples were taken from various surfaces in the Huanan market between 1 and 12 January 2020 by the CCDC (**File S1**) (15, 16, 18, 19, 41). As reported towards the end of January 2020 by Chinese news media (16, 19, 22, 42) (**File S1**) and the CCDC (15, 18, 20, 21, 41) (**File S1**), 31 of the 33 environmental samples that tested positive by qPCR or next generation sequencing (NGS) were located in the western area of the market (**Fig. 4, Table S2 and File S1**). The CCDC report (15) states: “it was discovered that wild animal trade exists in the Huanan Seafood Market’s West District, especially in Lanes 7 and 8 toward the interior of the market, where multiple wild animal vendors are located” and concluded that “it is highly suspected that the current epidemic is related to the trade of wild animals”. This conclusion was repeated by the director of the CCDC in January (20) and February 2020 (21) (**File S1**).

To investigate the potential role of the live animal trade in SARS-CoV-2 spread in the Huanan market, we built on the initial analysis in the WHO mission report (9) by integrating public online maps and photographic evidence, data from public business registries (**Table S5**), information about which live mammals species were sold at the Huanan market in late 2019 (**Tables 2 and S3**), and the CCDC report (**File S1**). We reconstructed the floor plan of the market (see **Methods**) and, by using the business registries of vendors at the market (**Table S5**), as well as an official report from summer 2019 fining the owners of three of these businesses for illegal sale of live mammals (43), we were able to identify an additional five stalls likely selling live or freshly butchered mammals, or other unspecified meat products in the western section of the market (**Fig. 4 and Table S2**) that were not labeled as such in the WHO mission report (**Table S2**). Thus, in addition to one live animal sales stall with environmental positives identified in the WHO report (9), we found evidence for at least two, and possibly as many as six such positive stalls, all closely clustered in the southwest corner of the western section of the market (**Fig. 4 and Table S2**). In addition, the CCDC report (15) provides further clarity by describing the number of positive samples per stall and the surface that had been swabbed to obtain each sample (**Table S2 and File S1**). The WHO mission report described 718 samples from 134 stalls that were obtained from the market (9), resulting in 40 positive samples across 21 stalls (**Table S4**); details for 33 of these positive samples were documented in the earlier CCDC report (**Table S2 and File S1**).

To study the distribution of positive environmental samples and December 2019 human COVID-19 cases within the Huanan market, we analyzed the sampling location data from the CCDC report (**Fig. 4**). As also reported in Chinese news media at the time (16, 19, 22, 42), 31 of the 33 (**Table S2**) positive environmental samples came from the western section of the market, where the majority of COVID-19 cases were associated, including all of the cases before, or on, 20 December (**Fig. 5**). By integrating information from the WHO mission report (9) on stalls that were subjected to environmental testing, but did not have any positive samples (**Table S6**), as well as two stalls with positive environmental samples listed in the WHO, but not in the CCDC report (**Table S7**), we developed a comprehensive map of the location and quantity of positive environmental samples at the Huanan market (**Figs. 4A and S15**).

Of particular salience, we found that five SARS-CoV-2-positive environmental samples were taken from a single stall known to be selling live mammals in late 2019 (**Table S2**). This level of positivity was matched only by a stall adjacent to it, also with five positive samples (**Table S2**). Importantly, the objects that tested positive from the stall selling live mammals showed clear associations with animal sales: a metal cage that was situated in a back room, two carts (of the kind frequently used to transport mobile animal cages) and a hair/feather remover (**Table S2**). No human COVID-19 cases were reported there (9, 15). Remarkably, this was the same stall where one of us (E.C.H.) observed live raccoon dogs

housed in a metal cage in 2014, stacked directly on top of a cage of live waterfowl or poultry (39) (**Fig. 3E**). A common trash cart directly outside of this specific lane was also found to be positive (**Figs. 4A, S16 and Table S2**). Finally, in another stall we identified as involved in the sale of live animals, a positive sample was obtained from inside a freezer (**Figs. 4A, S16 and Table S2**). While there were yet other environmental samples found throughout the market, most were directly associated with external surfaces, objects, or locations operated or visited by COVID-19 cases (**Fig. 3**) and therefore likely from human infections. Taken together, findings reported in this section are highly suggestive of infected animals having been present at the Huanan market at the beginning of the COVID-19 pandemic.

SARS-CoV-2-positive environmental samples showed clear clustering within the area of the Huanan market with live animal sales

Having found a clear association between animal sales and environmental sample SARS-CoV-2 positivity at the Huanan market, we next investigated the clustering of these samples within the market. To identify factors predictive of sample positivity, we used a binomial generalized linear model (GLM) to test the effect of proximity to known human cases and/or vendors selling live mammals on the frequency of positive samples at each business. While the WHO report indicates that approximately 5.4 samples were taken per stall (718 samples/134 stalls), neither the WHO (9) nor CCDC report (15) provides information on the number of samples from each individual stall, so we considered a range of possible quantities of total samples taken per stall ($N=5-10$, **Table S8**). For all tested values of N , we found that both distance to the nearest vendor selling live mammals and distance to the nearest human case were independently predictive of environmental sample positivity ($p=0.004$ and 0.014 , respectively for $N=6$, **Table S8**). However, we found that the odds of detection ($P(\text{detected})/P(\text{not detected})$) decreased by 2.8% per meter away from a live mammal vendor and by 6.6% per meter from a known human case (for $N=6$, **Table S8**), indicating that estimated effect on sample positivity due to proximity to live mammal vendors was farther-reaching than that of stalls with human cases. When we included vendors known to sell meats of unknown type, we found that the effects remained ($p=0.003$ and 0.008 , for distance to nearest live mammal/meat stall and distance to the nearest human case, respectively for $N=6$; **Table S9**). Crucially, these trends were not apparent in the stall-wise positivity data in the WHO report, but by integrating additional details on the number of positive samples per stall available in the CCDC report (**Table S2**), we found that proximity to vendors selling live mammals explains with high confidence the presence of SARS-CoV-2-positive environmental samples in the Huanan market.

To investigate clustering of SARS-CoV-2 positive environmental samples within the Huanan market, we performed a spatial relative risk analysis to determine any regions of the market with an increased density of positive environmental samples relative to a null model of all sampled stalls. We found evidence of only one such area, an isolated region in the southwest area of the market with a significantly increased density of positive samples ($p < 0.05$) that encloses two vendors associated with the sale of animals, including one known vendor of live mammals and another selling meat of unknown type (**Fig. 4B**). While environmental sampling of the market was incomplete and spatially heterogeneous, our analysis takes into account the empirical environmental sampling distribution, which was biased towards ‘stalls related to December cases’ as well as ‘stalls that sold livestock, poultry, farmed wildlife’ (9), in our null hypothesis (**Fig. 4C-D**).

Finally, to identify the spatial patterning of human cases in the market, we plotted cases as a function of their date of symptom onset noted in the WHO report, (9) as well as one additional case that was well described in the media (5), but missing from the WHO report (**Fig. 5A and Table S4**). This revealed that cases detected between 10 and 20 December, all from the western side of the market, formed two distinct spatial clusters, with one further north and the other more central (**Fig. 5B**). Although later cases diffused more widely across the market, the same two clusters on the western side of the market form the dominant centers of case density (**Fig. 5C**). In marked contrast to our findings from the environmental surveillance (**Table S2**), only two of the 31 human cases were observed in the southwestern section of the market selling live animals (**Fig. 4A**), with no clear clustering. This is not surprising given the high mobility of humans within the market relative to traded mammals, and given that only a small minority of infections become serious enough to lead to hospitalization (23).

Overall, these findings indicate that SARS-CoV-2-positive environmental samples were strongly associated with the sale of live mammals, particularly in the southwestern corner of the Huanan market where positive environmental samples were likely derived from infected animals. Additional sporadic positive environmental samples throughout the market were likely due to human infections.

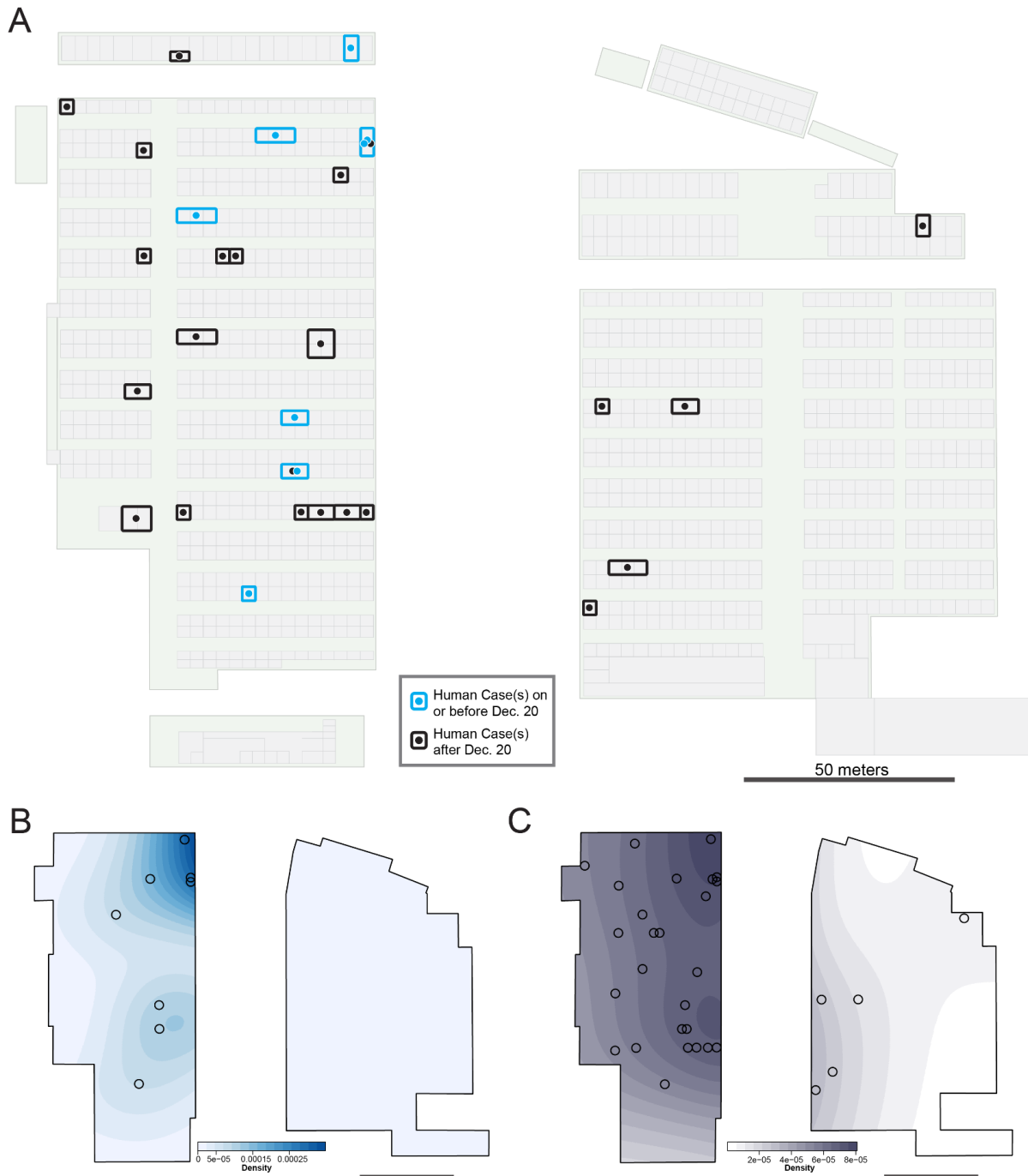


Figure 5. Location and timing of human cases in Huanan Market. **A.** Outline colors correspond to the timing of the first known case in each business. Individual case timing is denoted by marker color and shown within the outlined business. **B.** Distribution of known cases on or before December 20th, 2019. Locations of each case are shown as a black circle. **C.** Distribution of all known human cases in Huanan Market. See Table S7 for details on SARS-CoV-2 positive human cases with the Huanan market.

Discussion

Multiple lines of evidence establish the Huanan Seafood Wholesale Market in Wuhan as the site where the COVID-19 pandemic originated in late 2019: (i) SARS-CoV-2-susceptible animals, including raccoon dogs, were sold at the market in

November and December 2019; (ii) vendors known to have sold raccoon dogs in earlier years – and known to have sold illegal live mammals in late 2019 – yielded a large number of SARS-CoV-2-positive environmental samples, including several objects clearly associated with animals; (iii) positive environmental samples in the market were concentrated in the southwest corner of the western section of the market, the same area where most live mammals were traded; (iv) most cases among vendors within the market occurred in the western section where live mammals were sold; (v) a large proportion of the earliest known cases were identifiable as individuals who worked at, visited, or were linked to somebody who visited the market (5); (vi) this epidemiological link to the Huanan market is genuine and not due to ascertainment bias caused by special focus on the market as a possible site of cases (5); (vii) lineage A – and not just lineage B – viruses were circulating near to and centered on the Huanan market in the early stages of the outbreak, suggesting multiple spillovers may have occurred at the market; (viii) the spatial pattern of cases in December being so close to and centered on the Huanan market cannot be explained as arising by chance given population density patterns in Wuhan; (ix) this pattern holds when considering only cases that had no history of exposure at Huanan market, demonstrating that community transmission began in the direct vicinity of the market; (x) only by January and February 2020 did the spatial pattern of cases reflect that of the population density patterns in Wuhan precluding an earlier period of general transmission; and (xi) in a city of 11 million people there are thousands of sites (office buildings, factories, places of worship, universities, bars, restaurants, schools, etc.) that would be at least as likely to be the site of the initial cluster of a respiratory disease as the western section of the Huanan market – which measures only about 150m by 70m (see **Fig. S16**) – if the pandemic had not involved the trade in live mammals.

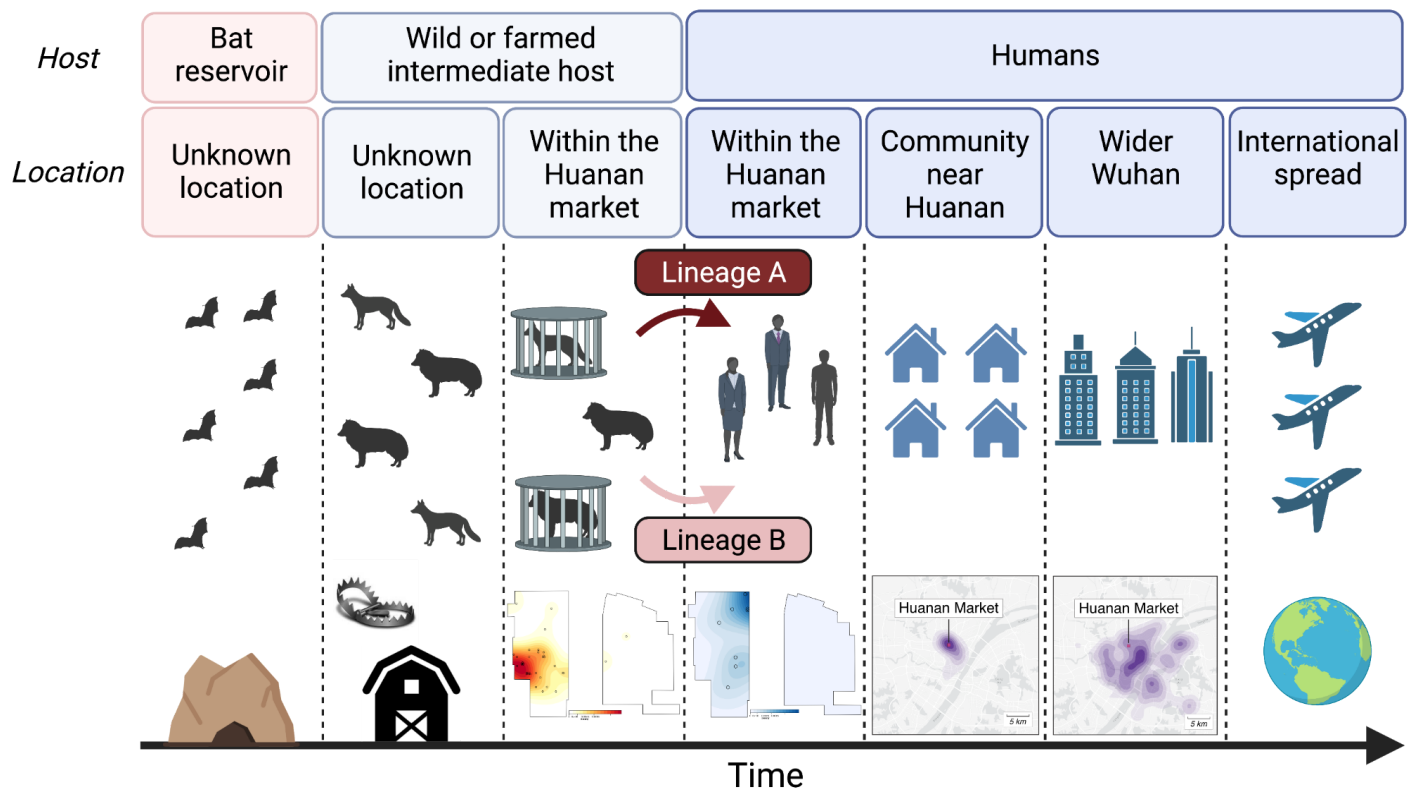


Figure 6. A schematic diagram of key milestones in the emergence of SARS-CoV-2. Bats were the reservoir of the ancestor of SARS-CoV-2, and they likely infected a wild or farmed intermediate host with this virus. Our analyses, and other evidence, indicates that the virus spread to other animals, and these were transported to the Huanan market. There were likely multiple infected animals at the Huanan market, leading to SARS-CoV-2 positive environmental samples (Fig. 4C) and at least two introductions of SARS-CoV-2 (*i.e.*, lineages A and B) into humans (companion paper by [Pekar et al.](#)). Human cases clustered in (Fig. 5B) and near to (Fig. 1B) the Huanan market, with SARS-CoV-2 quickly spreading to wider Wuhan (Fig. 1E) and then nationally and internationally. Created with Biorender.com.

No other location in Wuhan was, at the time or retrospectively, identified as being plausibly epidemiologically linked to the December 2019 COVID-19 cases (10). For example, a major transit hub, the Hankou Railway Station, is located near the Huanan market; but if mass transit was the route of entry into Wuhan, it would still require the establishment of infection in the live mammal populations in the Huanan market to be consistent with the results presented here. Likewise, the Wuhan CDC (WCDC) is very close to the market's location, but there was “no storage nor laboratory activities on CoVs or other bat viruses preceding the outbreak” at that site (9) and no epidemiological evidence of COVID-19 cases linked to the WCDC in December 2019 or earlier (9). Beyond the Huanan market, no other proposed or hypothesized origin narrative has been supported by any data (10, 44).

In a related study, we estimate that the primary case of SARS-CoV-2 likely occurred in late-November 2019 (and not before early November) and infer separate introductions of SARS-CoV-2 lineages A and B from animals into humans; a single-origin scenario of these two lineages has a probability of just 3.6% (companion paper by [Pekar et al.](#)). Combining these findings with the collection of evidence we report here, including that both lineages show a strong geographical association with the Huanan market early in the Wuhan epidemic, allows us to conclude that multiple origins of SARS-CoV-2 most likely occurred specifically within the Huanan market, beginning in November 2019, and perhaps extending into December. Such a pattern is consistent with multiple cross-species transmissions of SARS-CoV-2 from mink to workers on individual mink farms (45) and from infected Syrian hamsters to humans in the pet trade (46). The sustained presence of a source of repeated introductions into the human population over a period of time in late 2019, such as infected, live, mammalian wildlife at the Huanan market, is necessary to explain the origins of SARS-CoV-2 in light of the data provided here and in other work (5, 10, 32, 44, 47).

A key observation here is that SARS-CoV-positive environmental samples in the Huanan market were concentrated in the southwest corner of the western section where most live mammals were traded (**Fig. 4**), with multiple positive samples taken from one stall known to have sold live mammals. Notably, the NGS data from three Huanan market environmental samples were described in the WHO report (9) and virus consensus sequences submitted to GISAID; however, they only came from stalls that did not sell live mammals (**Table S2**), but which did report human COVID-19 cases. Critically, none of the raw metagenomic data from the six environmental samples originating from stalls that did sell live animals, but which did not report human cases, has been made available on an open database, even though the CCDC report makes it clear that these NGS data exist.

We cannot independently verify the authenticity of the CCDC report (15) (**File S1**) as it has only become public only via a media leak in June 2020 (41). However, the main results were reported by both state-run Chinese news media (16, 19, 22, 42) (**File S1**) and the CCDC (18, 20, 21) (**File S1**) and are fully consistent with those described in the CCDC report, including the number of overall specimens and positive samples. Given lack of detail, we also do not know the specific scheme employed for the environmental sampling at the Huanan market and it is possible that biases may exist. Further details on exact sampling mechanisms would be required to fully resolve such issues, but we note that our results are robust to significant sampling biases and these results should also be considered in the larger context of the strong association of early cases with the Huanan market.

For every one of the COVID-19 cases who became ill enough to be among the 174 hospitalized patients with illness onset in December of 2019 (9), there were likely more than ten milder cases that went unnoticed (5, 48). Clearly, therefore, we have not ascertained every early case in this pandemic (or any other). However, it is remarkable that, by the time of COVID-19 symptom onset of the Huanan market vendor constituting the earliest known case (10 December, 2019 (5)), there were likely only about ten, and less than 70, people infected with the (future) pandemic virus (companion paper by [Pekar et al.](#)). In less than two years it has become clear that (1) *Rhinolophus* bats in Laos harbored SARS-CoVs with a functional core – the receptor binding domain of the Spike protein – almost identical to that of SARS-CoV-2 (49), (2) early COVID-19 cases were both epidemiologically and geographically linked to the Huanan market in Wuhan, and (3) this was a market with sales of susceptible, live mammals at which SARS-CoV-2-positive environmental samples were

strongly associated with live mammal stalls known to sell these animals. We arguably now have as clear a picture of the zoonotic origin of SARS-CoV-2 as for any of the other eight coronaviruses known to infect humans (**Fig. 6**).

Note added in proof: A recent preprint from Gao *et al.* (50) confirms the authenticity of the CCDC report and reported the presence of the ‘A’ lineage of SARS-CoV-2 from an environmental sample at the Huanan market, consistent with a separate introduction of this lineage at the market. Further, this preprint reported additional positive environmental samples in the southwestern area of the market selling live animals. These findings are strongly consistent with those presented here, corroborating our conclusion that the Huanan market was the epicenter of SARS-CoV-2 emergence and the site of origin of the COVID-19 pandemic.

Acknowledgements

We thank Michael Standaert, Bonnie LaFleur, @babarlephant and a CNN reporter for assistance. We thank worldpop.org for making population density data from Wuhan freely available and GISAID submitters for generously sharing SARS-CoV-2 genomic data under GISAID regulations. We thank both the patients and the clinicians and researchers whose data made this research possible.

Funding

This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. 75N93021C00015 (MW). JIL acknowledges support from the NIH (5T32AI007244-38). JEP acknowledges support from the NIH (T15LM011271). JOW acknowledges support from NIH (AI135992 and AI136056). DLR acknowledges support of the Medical Research Council (MC_UU_12014/12) and the Wellcome Trust (220977/Z/20/Z). MAS, PL and AR acknowledge the support of the Wellcome Trust (Collaborators Award 206298/Z/17/Z – ARTIC network), the European Research Council (grant agreement no. 725422 – ReservoirDOCS) and NIH grant R01AI153044. ALR is supported by the Canadian Institutes of Health Research as part of the Coronavirus Variants Rapid Response Network (CoVaRR-Net; CIHR FRN#175622) and acknowledges that VIDO receives operational funding from the Canada Foundation for Innovation – Major Science Initiatives Fund and from the Government of Saskatchewan through Innovation Saskatchewan and the Ministry of Agriculture. MK receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 874735 (VEO, Versatile Emerging infectious disease Observatory). RFG is supported by the NIH (R01AI132223, R01AI132244, U19AI142790, U54CA260581, U54HG007480, OT2HL158260), the Coalition for Epidemic Preparedness Innovation, the Wellcome Trust Foundation, Gilead Sciences, and the European and Developing Countries Clinical Trials Partnership Programme. ECH is supported by an Australian Research Council Laureate Fellowship (FL170100022). KGA is supported by the NIH (U19AI135995, U01AI151812, and UL1TR002550).

Authors' contributions:

Conceptualization: MW, KGA

Methodology: MW, JIL, AC-C, LM, JEP, MUGK, MAS, ALR, DLR, SAG, AR, JOW, RFG, PL, ECH, KGA

Software: LM, JIL, JEP, JOW, PL

Validation: MW, LM, JIL, JEP, PL, JOW, KGA

Formal analysis: MW, JIL, AC-C, LM, JEP, MUGK, MAS, ALR, DLR, SAG, AR, JOW, RFG, PL, ECH, KGA

Investigation: MW, JIL, AC-C, LM, JEP, MUGK, MAS, MK, ALR, DLR, CN, SAG, AR, JOW, RFG, PL, ECH, KGA

Resources: MW, JOW, KGA

Data Curation: MW, KGA

Writing - original draft preparation: MW, RFG

Writing - review and editing: MW, JIL, AC-C, LM, JEP, MUGK, MAS, MK, ALR, CN, DLR, SAG, AR, JOW, RFG, PL, ECH, KGA

Visualization: MW, JIL, LM, JEP, ALR, AR, JOW, RFG, PL, ECH, KGA

Supervision: MW, JOW, KGA

Project administration: MW, KGA

Funding acquisition: MW, JIL, AC-C, LM, JEP, MUGK, MAS, ALR, DLR, SAG, AR, JOW, RFG, PL, ECH, KGA

Competing Interests

JOW receives funding from the CDC via contracts to his institution unrelated to this research. MAS receives funding from Janssen Research & Development, US Food & Drug Administration, and the US Department of Veterans Affairs via contracts and grants unrelated to this research. RFG is a co-founder of Zalgen Labs, a biotechnology company developing countermeasures for emerging viruses. MW, ALR, AR, MAS, ECH, and KGA have received consulting fees and/or compensated expert testimony on SARS-CoV-2 and the COVID-19 pandemic. MK has participated in the second WHO mission to China to study the origins of the pandemic and has served as scientific advisor on emerging disease preparedness to the Guangdong CDC prior to 2020.

References

1. Sina Finance, 武汉不明原因肺炎已隔离 检测结果将第一时间公布 | Wuhan Pneumonia of Unknown Cause Cases Isolated, Test Results to be Announced ASAP (2019), (available at <https://finance.sina.cn/2019-12-31/detail-iihnzakh1074832.d.html?from=wap>).
2. Wuhan Health Commission, 武汉市卫生健康委员会 | Wuhan Municipal Health Commission's briefing on the current situation of pneumonia in our city (2019), (available at <https://web.archive.org/web/20200131202951/http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989>).
3. Pneumonia of unknown cause – China, (available at <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229>).
4. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) - China, 2020. *China CDC Wkly.* **2**, 113–122 (2020).
5. M. Worobey, Dissecting the early COVID-19 cases in Wuhan. *Science.* **374**, 1202–1204 (2021).
6. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* **395**, 497–506 (2020).
7. The-nCoV Outbreak Joint Field Epidemiology Investigation Team, Q. Li, An Outbreak of NCIP (2019-nCoV) Infection in China - Wuhan, Hubei Province, 2019-2020. *China CDC Wkly.* **2**, 79–80 (2020).
8. Wuhan Municipal Health Commission, 武汉市卫生健康委员会 | Wuhan Municipal Health Commission's briefing on unexplained viral pneumonia (2020), (available at <https://web.archive.org/web/20200105131338/http://wjw.wuhan.gov.cn/front/web/showDetail/2020010509020>).
9. WHO, WHO-convened global study of origins of SARS-CoV-2: China Part (2021), (available at <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>).
10. E. C. Holmes, S. A. Goldstein, A. L. Rasmussen, D. L. Robertson, A. Crits-Christoph, J. O. Wertheim, S. J. Anthony, W. S. Barclay, M. F. Boni, P. C. Doherty, J. Farrar, J. L. Geoghegan, X. Jiang, J. L. Leibowitz, S. J. D. Neil, T. Skern, S. R. Weiss, M. Worobey, K. G. Andersen, R. F. Garry, A. Rambaut, The origins of SARS-CoV-2: A critical review. *Cell.* **184**, 4848–4856 (2021).
11. X. Xiao, C. Newman, C. D. Buesching, D. W. Macdonald, Z.-M. Zhou, Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci. Rep.* **11**, 1–7 (2021).
12. C. M. Freuling, A. Breithaupt, T. Müller, J. Sehl, A. Balkema-Buschmann, M. Rissmann, A. Klein, C. Wylezich, D. Höper, K. Wernike, A. Aebischer, D. Hoffmann, V. Friedrichs, A. Dorhoi, M. H. Groschup, M. Beer, T. C.

- Mettenleiter, Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerg. Infect. Dis.* **26**, 2982–2985 (2020).
13. W. K. Jo, E. F. de Oliveira-Filho, A. Rasche, A. D. Greenwood, K. Osterrieder, J. F. Drexler, Potential zoonotic sources of SARS-CoV-2 infections. *Transbound. Emerg. Dis.* **68**, 1824–1834 (2021).
 14. I. R. Fischhoff, A. A. Castellanos, J. P. G. L. M. Rodrigues, A. Varsani, B. A. Han, Predicting the zoonotic capacity of mammals to transmit SARS-CoV-2. *Proc. Biol. Sci.* **288**, 20211651 (2021).
 15. W. Guizhen, “中疾控病疾控报 | Chinese CDC Disease Control Report” (53, Chinese CDC, 2020).
 16. Xinhua News, 好消息！新型冠状病毒溯源取得阶段性进展-新华网 | Good News! The Investigation into the Origin of the Coronavirus Makes Periodic Headway (2020), (available at http://www.xinhuanet.com/politics/2020-01/26/c_1125503792.htm).
 17. Beijing News, 新京报 - 好新闻, 无止境 | Beijing News –Boundless good news (2020), (available at <http://www.bjnews.com.cn/feature/2020/01/02/669054.html>).
 18. C. D. C. Chinese, 中国疾控中心在武汉华南海鲜市场检出大量新型冠状病毒 | Large Quantity of Novel Coronaviruses Detected by China CDC in Wuhan Huanan Seafood Market (2020), (available at https://www.chinacdc.cn/yw_9324/202001/t20200127_211469.html).
 19. Yicai, China detects large quantity of novel Coronavirus at Wuhan seafood market (2020), (available at <https://www.yicai.com/opinion/yicai.global/china-detects-large-quantity-of-novel-coronavirus-at-wuhan-seafood-market>).
 20. C. D. C. Chinese, 中国疾控中心呼吁公众做好自我防护 | China CDC Calls on the Public to Protect Themselves (2020), (available at https://www.chinacdc.cn/yw_9324/202001/t20200128_211498.html).
 21. C. D. C. Chinese, 在一线，疾控勇士与新型冠状病毒赛跑 | On theFront Line - CDC Warriors Racing against the Novel Coronavirus (2020), (available at https://www.chinacdc.cn/yw_9324/202002/t20200201_212137.html).
 22. Xinhua, China detects large quantity of novel coronavirus at Wuhan seafood market (2020), (available at https://web.archive.org/web/20200126230041/http://www.xinhuanet.com/english/2020-01/27/c_138735677.htm).
 23. D. P. Oran, E. J. Topol, Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann. Intern. Med.* **173**, 362–367 (2020).
 24. Report 2 - Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China. *Imperial College London*, (available at <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-2-update-case-estimates-covid-19/>).
 25. J. Giesecke, Primary and index cases. *Lancet.* **384**, 2024 (2014).
 26. Z. Shi, Z. Hu, A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res.* **133**, 74–87 (2008).
 27. S. Lytras, W. Xia, J. Hughes, X. Jiang, D. L. Robertson, The animal origin of SARS-CoV-2. *Science.* **373**, 968–970 (2021).
 28. A. S. Bernstein, A. W. Ando, T. Loch-Temzelides, M. M. Vale, B. V. Li, H. Li, J. Busch, C. A. Chapman, M. Kinnaird, K. Nowak, M. C. Castro, C. Zambrana-Torrel, J. A. Ahumada, L. Xiao, P. Roehrdanz, L. Kaufman, L. Hannah, P. Daszak, S. L. Pimm, A. P. Dobson, The costs and benefits of primary prevention of zoonotic pandemics. *Sci Adv.* **8**, eabl4183 (2022).
 29. W.-T. He, X. Hou, J. Zhao, J. Sun, H. He, W. Si, J. Wang, Z. Jiang, Z. Yan, G. Xing, M. Lu, M. A. Suchard, X. Ji, W. Gong, B. He, J. Li, P. Lemey, D. Guo, C. Tu, E. C. Holmes, M. Shi, S. Su, Virome characterization of game animals in China reveals a spectrum of emerging pathogens. *Cell* (2022), doi:10.1016/j.cell.2022.02.014.

30. Z. Peng, R. Wang, L. Liu, H. Wu, Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data. *ISPRS International Journal of Geo-Information*. **9**, 402 (2020/6).
31. D. Wang, J. Cai, T. Shi, Y. Xiao, X. Feng, M. Yang, W. Li, W. Liu, L. Yu, Z. Ye, T. Xu, J. Ma, M. Li, W. Chen, Epidemiological characteristics and the entire evolution of coronavirus disease 2019 in Wuhan, China. *Respir. Res.* **21**, 257 (2020).
32. J. Pekar, M. Worobey, N. Moshiri, K. Scheffler, J. O. Wertheim, Timing the SARS-CoV-2 index case in Hubei province. *Science*. **372**, 412–417 (2021).
33. Y. Jia, Z. Zheng, Q. Zhang, M. Li, X. Liu, Associations of Spatial Aggregation between Neighborhood Facilities and the Population of Age Groups Based on Points-of-Interest Data. *Sustain. Sci. Pract. Policy*. **12**, 1692 (2020).
34. A. Rambaut, E. C. Holmes, Á. O’Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. **5**, 1403–1407 (2020).
35. Outbreak.info Variant Reports. *outbreak.info*, (available at <https://outbreak.info/situation-reports>).
36. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. **395**, 565–574 (2020).
37. K. L. Stout, “Wuhan SARS”: Tracing the origin of the new virus to China’s wild animal markets (2020), (available at https://www.youtube.com/watch?v=Je0_U2ym_r0).
38. S. M. Porter, A. E. Hartwig, H. Bielefeldt-Ohmann, A. M. Bosco-Lauth, J. Jeffrey Root, Susceptibility of wild canids to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *bioRxiv* (2022), p. 2022.01.27.478082.
39. Y.-Z. Zhang, E. C. Holmes, A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell*. **181**, 223–227 (2020).
40. M. Standaert, E. Dou, In search for coronavirus origins, Hubei caves and wildlife farms draw new scrutiny. *The Washington Post* (2021), (available at https://www.washingtonpost.com/world/asia_pacific/china-covid-bats-caves-hubei/2021/10/10/082eb8b6-1c32-11ec-bea8-308ea134594f_story.html).
41. Epoch Times, 【独家】武汉华南海鲜市场检测的秘密 | [Exclusive] The Secrets Behind the Testing of Wuhan’s Huanan Seafood Market. *Epoch Times* (2020), (available at <https://www.epochtimes.com/gb/20/5/31/n12150755.htm>).
42. Xinhua, 中国疾控中心在武汉华南海鲜市场检出大量新型冠状病毒-新华网 | Large Quantity of Novel Coronaviruses Detected by China CDC in Wuhan Huanan Seafood Market (2020), (available at https://web.archive.org/web/20211217122036/http://www.xinhuanet.com/2020-01/27/c_1125504355.htm).
43. Wuhan Municipal Bureau of Landscape Architecture and Forestry, 武汉市园林和林业局 | Administrative Penalties in 2019 (2019), (available at https://web.archive.org/web/20211117124950/http://ylj.wuhan.gov.cn/zwgk/zwxgkzl_12298/cfqz/xzcf/202011/t20201110_1499879.shtml).
44. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
45. L. Lu, R. S. Sikkema, F. C. Velkers, D. F. Nieuwenhuijse, E. A. J. Fischer, P. A. Meijer, N. Bouwmeester-Vincken, A. Rietveld, M. C. A. Wegdam-Blans, P. Tolsma, M. Koppelman, L. A. M. Smit, R. W. Hakze-van der Honing, W. H. M. van der Poel, A. N. van der Spek, M. A. H. Spierenburg, R. J. Molenaar, J. de Rond, M. Augustijn, M.

- Woolhouse, J. A. Stegeman, S. Lycett, B. B. Oude Munnink, M. P. G. Koopmans, Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat. Commun.* **12**, 6802 (2021).
46. H.-L. Yen, T. H. C. Sit, C. J. Brackman, S. S. Y. Chuk, S. M. S. Cheng, H. Gu, L. D. J. Chang, P. Krishnan, D. Y. M. Ng, G. Y. Z. Liu, M. M. Y. Hui, S. Y. Ho, K. W. S. Tam, P. Y. T. Law, W. Su, S. F. Sia, K.-T. Choy, S. S. Y. Cheuk, S. P. N. Lau, A. W. Y. Tang, J. C. T. Koo, L. Yung, G. Leung, J. S. M. Peiris, L. L. M. Poon, Transmission of SARS-CoV-2 (Variant Delta) from Pet Hamsters to Humans and Onward Human Propagation of the Adapted Strain: A Case Study. *The Lancet* (2022), , doi:10.2139/ssrn.4017393.
 47. M. Worobey, J. Pekar, B. B. Larsen, M. I. Nelson, V. Hill, J. B. Joy, A. Rambaut, M. A. Suchard, J. O. Wertheim, P. Lemey, The emergence of SARS-CoV-2 in Europe and North America. *Science.* **370**, 564–570 (2020).
 48. S. Mahajan, C. Caraballo, S.-X. Li, Y. Dong, L. Chen, S. K. Huston, R. Srinivasan, C. A. Redlich, A. I. Ko, J. S. Faust, H. P. Forman, H. M. Krumholz, SARS-CoV-2 Infection Hospitalization Rate and Infection Fatality Rate Among the Non-Congregate Population in Connecticut. *Am. J. Med.* **134**, 812–816.e2 (2021).
 49. S. Temmam, K. Vongphayloth, E. B. Salazar, S. Munier, M. Bonomi, B. Regnault, B. Douangboubpha, Y. Karami, D. Chrétien, D. Sanamxay, V. Xayaphet, P. Paphaphanh, V. Lacoste, S. Somlor, K. Lakeomany, N. Phommavanh, P. Pérot, O. Dehan, F. Amara, F. Donati, T. Bigot, M. Nilges, F. A. Rey, S. van der Werf, P. T. Brey, M. Eloit, Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* (2022), doi:10.1038/s41586-022-04532-4.
 50. G. Gao, W. Liu, P. Liu, W. Lei, Z. Jia, X. He, L.-L. Liu, W. Shi, Y. Tan, S. Zou, X. Zhao, G. Wong, J. Wang, F. Wang, G. Wang, K. Qin, R. Gao, J. Zhang, M. Li, W. Xiao, Y. Guo, Z. Xu, Y. Zhao, J. Song, J. Zhang, W. Zhen, W. Zhou, B. Ye, J. Song, M. Yang, W. Zhou, Y. Bi, K. Cai, D. Wang, W. Tan, J. Han, W. Xu, G. Wu, Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market. *Research Square* (2022), doi:10.21203/rs.3.rs-1370392/v1.

Supplementary Materials for

The Huanan market was the epicenter of SARS-CoV-2 emergence

Authors: Michael Worobey*, Joshua I. Levy, Lorena M. Malpica Serrano, Alexander Crits-Christoph, Jonathan Pekar, Stephen A. Goldstein, Angela L. Rasmussen, Moritz U. G. Kraemer, Marc A. Suchard, Chris Newman, Marion P. G. Koopmans, Joel O. Wertheim, Philippe Lemey, David L. Robertson, Robert F. Garry, Edward C. Holmes, Andrew Rambaut, Kristian G. Andersen*

*Corresponding author. Email: worobey@arizona.edu (MW), andersen@scripps.edu (KGA)

Materials and Methods

Supplementary Text

File S1

Figs. S1-S17

Tables S1-S11

Supplementary References

Materials and Methods

Sources of data

COVID-19 case data from December of 2019. The COVID-19 case data we consider here is that reported in the WHO mission report (1). These are all individuals that were hospitalized and notified to China's national disease surveillance system by hospitals in Wuhan during the early phase of the epidemic. Importantly, they are included as "December" cases based on their date of COVID-19 symptom onset, not their date of hospitalization.

Population density data from Wuhan, China. The dataset used for population density was downloaded from worldpop.org: China population 2020: constrained 100 m resolution: projection WGS84: Geotiff format: constrained datasets. This estimates population densities only within areas mapped as containing built settlements (2). The Geotiff file was imported directly to R and cropped to the administrative boundaries of Wuhan, using a shapefile downloaded from <https://data.humdata.org/dataset/china-administrative-boundaries>. The worldpop.org data gives an estimate of the number of people per pixel. The resulting raster file for the Wuhan province has a total of 494,676 pixels. This raster file was transformed into a pixel image, from which the coordinates and values of each pixel were extracted.

COVID-19 case data from January and February of 2020. Geotagged Weibo COVID-19 case location data were shared with us by the authors (3) and consisted of 737 points defined by latitude and longitude. These were from users of a Weibo app designed for COVID-19 patients seeking medical information on the disease.

Extraction of December 2019 COVID-19 case locations

Geographical coordinates were not available for the residential locations of the 174 cases in Hubei province, China reported by the WHO mission to have had COVID-19 symptom onset in 2019 (all in December) (1). We therefore extracted locations from the 2-dimensional location data within the maps available in the SARS-COV-2 mission report and its annexes (1) using the following approach: We started with (1) Fig. 4 on page 148, Annex E of the WHO mission report (**Fig. S1**, this study), which depicts the spatial distribution of 164 cases living in Wuhan by residential location, along with information on whether each case was epidemiologically linked to the Huanan market, and (2) a map of Wuhan adapted from one generated by us during an earlier study (4) (**Figure S2**). We gratefully acknowledge that data used to create that map of Wuhan was acquired from OpenStreetMap®, which is licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF), ©OpenStreetMap contributors (<https://www.openstreetmap.org/copyright>).

We then overlaid these two maps in Adobe Illustrator v25.4.1 (**Fig. S3**) using the borders of Wuhan in each map to achieve a close overlap. While isolated case locations were easily discerned in WHO Figure 4 of Annex E, many were located in areas of high density. Moreover, the large circles used to indicate each residential location, with a diameter of approximately 2 km, plus the low resolution of the map, meant that many circles overlapped each other and could not be individually identified. Almost all other case maps in the WHO mission report had similarly large case markers and low resolution. Figure 23 on page 44 of the main WHO report (**Fig. S4**, this study), was the exception and appeared to resolve cases better than other figures in the report; however, it lacked borders of the city or other clear points of reference to use for reverse engineering the map's visual information into numerical coordinates. We therefore combined this source of information with that shown in **Fig. S3**, which involved non-uniform scaling of the vertical and horizontal dimensions of WHO Figure 23 in Illustrator until the points in that map were clearly placed in the center of the larger circles of WHO Annex E Figure 4 (**Fig. S5**). This revealed that the WHO Figure 23 circles had diameters of roughly 250 m.

Next, in Illustrator, we overlaid markers of similar size to the WHO Figure 23 ones over the WHO Figure 23 ones, using purple circles to represent cases that were depicted in WHO Annex E Figure 4 to be epidemiologically unlinked to the Huanan market and orange circles for linked cases (**Fig. S6**). All cases outside the region covered by WHO Figure 23 were isolated, and points could be easily placed in the center of each (**Fig. S7**). Removing the WHO Figure 23 and WHO Annex E Figure 4 layers revealed a map with points represented by circles with diameters of about 250 m (**Fig. S8**). This map was further annotated with information from (1) Figure 6 on page 78 of the main WHO mission report (1), which allowed us to discern the location of eleven lineage B cases and one lineage A case, and (2) Annex E Figure 17 (page

156), which allowed us to discern the locations of six family clusters. We assumed that each cluster involved two individuals (**Fig. S8**).

Cases were to some extent obscured in the region closest to the Huanan market because of a high density of plotted locations even in the highest-resolution map in the WHO mission report (WHO Figure 23) - circles marking individual cases in some instances overlapped one another. Nevertheless, we recovered 156 of the 164 cases depicted in WHO Annex E Figure 4. The eight cases for which we were not able to extract a location therefore likely lie in the high case density area near to the Huanan market.

Next we used WebPlotDigitizer v4.5 (Ankit Rohatgi; <https://automeris.io/WebPlotDigitizer>) to reverse engineer the cases marked on the map in **Fig. S8** and extract the underlying numerical data (latitude and longitude for each point). We used the map locations and latitude and longitude Huanan and Qiyimen markets to generate two X-axis and two Y-axis reference points (**Fig. S9**). We then marked each case in WebPlotDigitizer as shown by the example in **Fig. S10**. Once all COVID-19 cases locations were marked, we converted the points into latitudes and longitudes in geojson format and subsequently annotated the file with information on which were family cases and which were lineage A and lineage B cases (File name '156_Wuhan_cases.geojson').

Kernel density estimates

Densities were calculated in ESRI ArcGIS Online using a bandwidth of 3 km, an equal interval classification, and 7 classes. Sensitivity analyses using bandwidths of 2 km and 4 km were also conducted.

Analyses of distances to the Huanan market

Haversine distances to the Huanan market were calculated for each of the 156 December 2019 cases. The median distance from cases to the Huanan market was calculated separately for (1) all 156 case, (2) the 33 cases epidemiologically linked to the Huanan market, (3) for the 123 cases not epidemiologically linked to the market, (4) the nine lineage B cases linked to the market, (5) the two lineage B cases unlinked to the market, and (6) the two lineage A cases unlinked to the market.

For the Wuhan population density data, the weighted median haversine distance to the Huanan market was calculated, with the weights assigned to each point being the pixel value of each point, which represents the population density for the 100 m by 100 m square with that point at its center. We also calculated the median haversine distance of the 737 January-February 2020 Weibo COVID-19 cases (3).

To test whether the December cases were closer to the Huanan market than expected, null distributions were generated from both the population density data and the Weibo data, with sample sizes matching the six listed above for the December cases and sampling with replacement. 1000 pseudoreplicates were generated in each case and the median distance to the Huanan market of each set of locations in each pseudoreplicate was calculated. For the population density data sampling weights were assigned to each point, the weight being the pixel value, such that points representing higher population densities had a correspondingly higher probability of being sampled. The median distances between the Early December cases (epidemiologically linked and unlinked to Huanan) were compared to these null distributions.

Analyses of centering on the Huanan market

Centroids were determined by calculating the median latitude and median longitude of groups of locations (*e.g.*, the 156 COVID-19 cases locations from December 2019). Centroids were determined for the six groupings of December COVID-19 cases listed in the section above, as well as for the population density data (worldpop.org) and the Weibo data (3). The haversine distance between each resulting centroid and Huanan market was obtained. The population density centroid was estimated by taking the weighted median latitude and weighted median longitude of the population density dataset (worldpop), assigning the pixel values as the weights, while the centroid of the Weibo data (3) was defined by the median of the 737 latitudes and the median of the 737 longitudes.

Significance testing of the distance between the December centroids and the Huanan market compared to the null distributions was based on random samples (with replacement) of 1,000 points from the population density data (weighted) or the Weibo data (3), to represent plausible starting point of the COVID-19 epidemic in Wuhan, insofar as the centroid of early cases might reflect the starting point of the epidemic.

December COVID-19 cases sensitivity analyses

We implemented an approach to investigate how robust the results with our geocoding analyses were to possible error in locations extracted from the WHO mission report maps. As shown in **Figs. S1-S11**, we were able to reliably ascertain coordinates of points marked on maps in the WHO mission report, likely to within approximately 50 m. To explore how robust the results of our statistical analyses were to uncertainty in geocoded locations, we resampled each point randomly from anywhere within a circle of radius = 500 m or 1000 m, centered on our geocoded location. In other words, for each point of the 156 in our December COVID-19 case location data set, we introduced noise such that its location in sensitivity analyses could be up to 1 km from where we had placed it. We ran the same sensitivity analyses for just the Huanan market-linked cases (33) and for the Huanan-market-unlinked case (123) (**Table S1**).

We further examined the sensitivity of our analyses to both location “noise” and, simultaneously, to missing data. There were eight cases for which we were not able to extract a location, and these were highly likely from the region closest to the Huanan market, with a high density of plotted (and therefore likely overlapping) locations. We performed 1,000 iterations where each location was uniformly sampled from a circle polygon centered on our geocoded location with a radius of 500 m, or 1000 m, which exceeded by far the possible imprecision in the extracted locations. To assess the impact of the missing data, we also uniformly sampled eight additional locations from the 25% kernel density estimate of the 156 December cases at each iteration (reflecting the fact that our location extraction procedure likely missed cases in the high-density area).

Finally, we generated 1000 replicates sampling 102 cases from the set of 123 cases not linked to the Huanan market, to assess whether our results were robust to mis-assignment of some ‘linked’ cases as ‘unlinked’.

As described above, we computed the median distance and the centroid distance to the Huanan market for each iteration and compared their average across all iterations to the null distributions drawn from the Weibo cases and worldpop.org samples. In addition, we computed the range of P-values for comparing the metrics at each iteration to the null distributions. These analyses were performed for the resampled locations as well as for the resampled locations with the additional samples representing missing cases, in both cases for all locations, for the locations of the cases linked to the Huanan market, and for the cases not linked to the Huanan market. The missing cases were considered to be linked to the Huanan market as our extraction misses at least that many linked cases. All results were robust to error in location estimation and missing data (**Table S1**).

Tolerance contour analysis

We studied variations in relative risk, $r(z) = f(z)/g(z)$, at each position z , where $f(z)$ is the density of the test distribution, and $g(z)$ is the density of the control distribution. We explicitly tested the null hypothesis $H_0: r(z) = 1$, against an alternative hypothesis of increased relative risk, $H_1: r(z) > 1$. Using an asymptotic p-value computation (shown to be more robust than a Monte Carlo approach) (5), we calculated and plotted contours corresponding to significant values of $P(z)$, a pointwise estimate of statistical significance. For bandwidth estimation, we performed least squares cross validation on points within the same rectangular region as in **Fig. 1C** (Latitude:[30.4433,30.7466], Longitude:[114.0875,114.4741]), which contains more than 92% (144 out of 156) of all early cases.

Lineage ‘A’ and ‘B’ case locations

We linked SARS-CoV-2 genome lineage information (‘A’ or ‘B’) (6) and residential location for twelve COVID-19 cases among the 156 from December 2019 for which we extracted locations - 11 lineage B (nine linked to the Huanan market and 2 unlinked) and one lineage A (**Figs. S1-S11**). Because we had earlier determined the location of the single lineage A virus (7) among the thirteen December cases for which genomic information was reported by the WHO mission report (1)

of which we were able to link twelve to particular cases – we were certain that the remaining eleven cases had been infected by lineage B. Location information was also available for one additional lineage A case with COVID-19 onset in 2019: virus ‘Wuhan/WH04/2020’. This individual had stayed at a hotel near the Huanan market for the five days before fever onset (8). These are the two earliest-onset lineage A cases currently reported, and the eleven lineage B genomes are among the twelve earliest lineage B genomes. Lacking residential location information for the case that yielded WH04, we assigned a 383 m distance to the Huanan market for it: the nearest hotel is 27 m away from the edge of the Huanan market, and the median distance from the twenty nearest hotels, for which we could obtain latitude and longitude using Bing Maps, to the center point of the market was 383 m (**Table S10**). The GISAID acknowledgments table can be found in **Table S11**.

Floorplan of the Huanan market.

Two maps of the internal arrangement of stalls in the Huanan Market are available: The CCDC map (9) (**File S1**) and the WHO map (1). While the two maps agree on key features, including the general arrangement of the stalls, neither map is drawn to scale. The internal location of stalls in the Huanan Market for our study was discerned through detailed analysis of satellite photographs (Google Maps, Google Earth, Baidu Maps), aerial photographs, and other images of the market. Use of Baidu Total View, which displays interactive panoramic images at street level, allowed for construction of an accurate detailed internal map of the locations of individual stalls. The outlines of the eastern and western sections of the market were traced from Google Maps using Adobe Illustrator. Locations of the key internal structural elements of the main areas of the Market, the vertical pillars, were determined from Total View images and other photographs then mapped onto the external walls of the eastern and western sections of the market. The western part of the market is approximately 70 meters wide, with pillars spaced an average of 8.75 meters apart (9 pillars per street), and the pillars along the main market building span about 126 meters, with pillars spaced about 9 meters apart (15 pillars total). Stalls on the north side of the canopy are supported by smaller metal pillars and are spread over the full market width of 70 meters and are each 6 meters deep. Additional stalls and storage units in the West market parking lot were not included, as they were not studied by either the CCDC or WHO report. The eastern part of the market is approximately 92 meters at its widest, and the main building is approximately 90 meters long (supported by a series of 15 pillars). Distances were all confirmed using Google Earth’s measurement tool. The panoramic images of Baidu Total View were then used to map the location of individual numbered walkways (“streets”) throughout the market; major vehicle passages; and the dimensions of other major structures, such as the canopy in the eastern section of the market. We also cross-checked our conclusions with an independent researcher’s collation covering much of the same information (<http://babarlephant.free-host.net/visiting-the-wuhan-seafood-market> courtesy of @babarlephant). The map was then converted into geojson format for spatial analyses.

Analysis of environmental samples in the Huanan market

The location and quantity of positive environmental samples were taken directly from the CCDC map (9) (**File S1**), with the exception of two positive businesses in the southwest corner of the market that were only noted on the WHO map (1). Locations of known live animal vendors, cases, as well as businesses with no positive environmental samples were obtained from the WHO map (1). Testing of the significance of live animal vendors and/or human SARS-CoV-2 cases on the number of positive environmental samples found in a business was performed using a binomial GLM available in the ‘stats’ package in R. Distances between businesses were defined as the distance between their respective centroids. Spatial relative risk analysis was performed using the ‘sparr’ package in R, using linear boundary kernels for edge correction (10), with bandwidth selection performed using least squares cross-validation. Tolerance contours were calculated using the same robust asymptotic approximation used in **Fig. 3 (5)**. Kernel density estimates of positive environmental sample density as well as sampled businesses were performed the same bandwidth identified from cross validation in the relative risk analysis.

Market stalls were assigned by categories of the types of goods sold through integration of several different sources. The WHO mission report (1) identified 10 stalls involved in the trade of live animals (Appendix F, Table 3). We further obtained names and descriptions of stalls from the TianYanCha.com business directory (**Table S5**) based on stall addresses

within the market. In many cases, business names and meat products sold were further confirmed through photographic evidence of stalls. Three stalls were identified as involved in domesticated wildlife sales from an official local forestry bureau fine for their registered owners for illegal hedgehog sales in summer 2019 (11). One of these two stalls (street 8, 25) yielded an environmental positive from the interior of a freezer, but was not noted as selling domesticated wildlife in Figure 2 from the “ANIMAL AND ENVIRONMENT STUDIES” section of the WHO mission report (1).

Analysis of human cases in Huanan Seafood Market

The location and timing of human cases was taken from the WHO report, with the exception of one case identified by media as well as in recent work. Analysis was performed using the ‘sparr’ package in R (10). Bandwidth selection for each kernel density estimate was determined by likelihood cross-validation.

Translations

All translations in **File S1** were done using the “general” and/or “academic” premium service at TheWordPoint (<https://thewordpoint.com/>). Originals can be accessed via the original URLs or WayBackMachine URLs provided in the “documents_index.csv” file in **File S1**.

Data and code availability

Data and code for this manuscript is on GitHub: https://github.com/sars-cov-2-origins/paper_2022_huanan-market.

Supplementary Text

Animal screening from SARS-CoV-2 in China

While >80,000 animal samples were tested and no evidence was found for presence of, or exposure to, SARS-CoV-2 (1), this is far from conclusive evidence for most animals. The negative predictive value of the surveys depends on the assumed prevalence. For serology, for instance, if one assumes a reservoir host in which the virus is endemic, seroprevalence likely is high, but will depend on the age of the animals and infection dynamics in that particular species. In addition, China is a vast country, and the possible sources of animals could be over a long distance. Therefore, the catchment area for surveillance studies is huge (12). Therefore, tracking back along the market supply chain is crucial for a more targeted effort, as recommended in the WHO mission report (1). In addition, it is possible that a one-off spillover event occurred in a farm that supplied the market. Such events would be needles in a haystack that are virtually unresolvable.

Information shared with us regarding the live mammal photographs purportedly from 3 December, 2019 at the Huanan market

The photos in **Fig. 3A-C** were posted on Sina Weibo purportedly by an individual concerned about the conditions under which live animals were being traded at the market, and they have been reported on previously (13). A CNN journalist, who has contacted the source directly, shared with us that the source stated that he had taken the photographs at the Huanan market on 3 December, 2019. The source stated that he had lost his phone with the original image files but had retrieved files from a friend's WeChat chatting history; the chat happened between this source and his friend on the day that these photographs were taken, the source said. The CNN journalist stated that WeChat would indeed compress the images and strip them of metadata, consistent with the files provided by the source to the CNN journalist, then provided by the CNN reporter to us, and now provided by us as part of this study. Consistent with the source's account of when and where he had taken the photographs, he had posted them on Weibo on 3 January, 2020, within just a few days of the first public announcement of unexplained pneumonia cases linked to the Huanan market. By 4 or 5 January the photos had been removed from Weibo.

Supplementary File

File S1. Translations and URLs for relevant articles and reports

File is available from [DOI: 10.5281/zenodo.6291868](https://doi.org/10.5281/zenodo.6291868)

Supplementary Figures

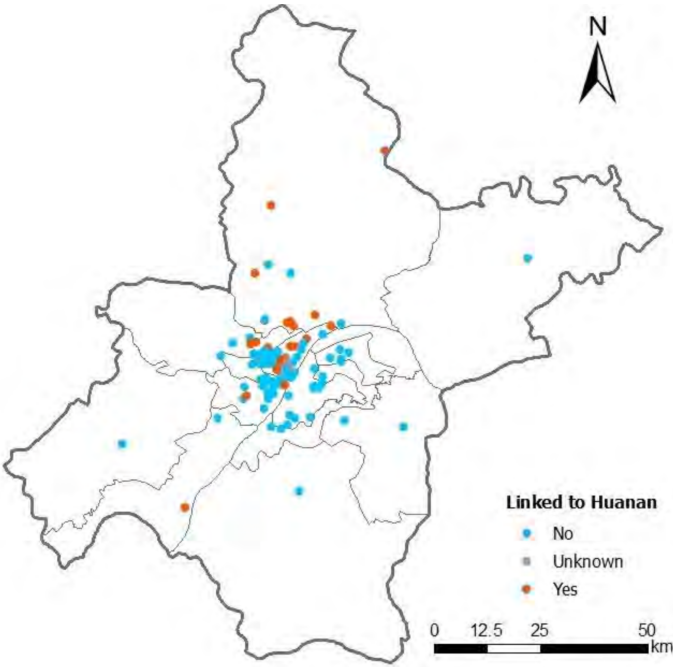


Figure S1. COVID-19 cases in Wuhan in December. This is Fig. 4 on page 148, Annex E of the WHO mission report.

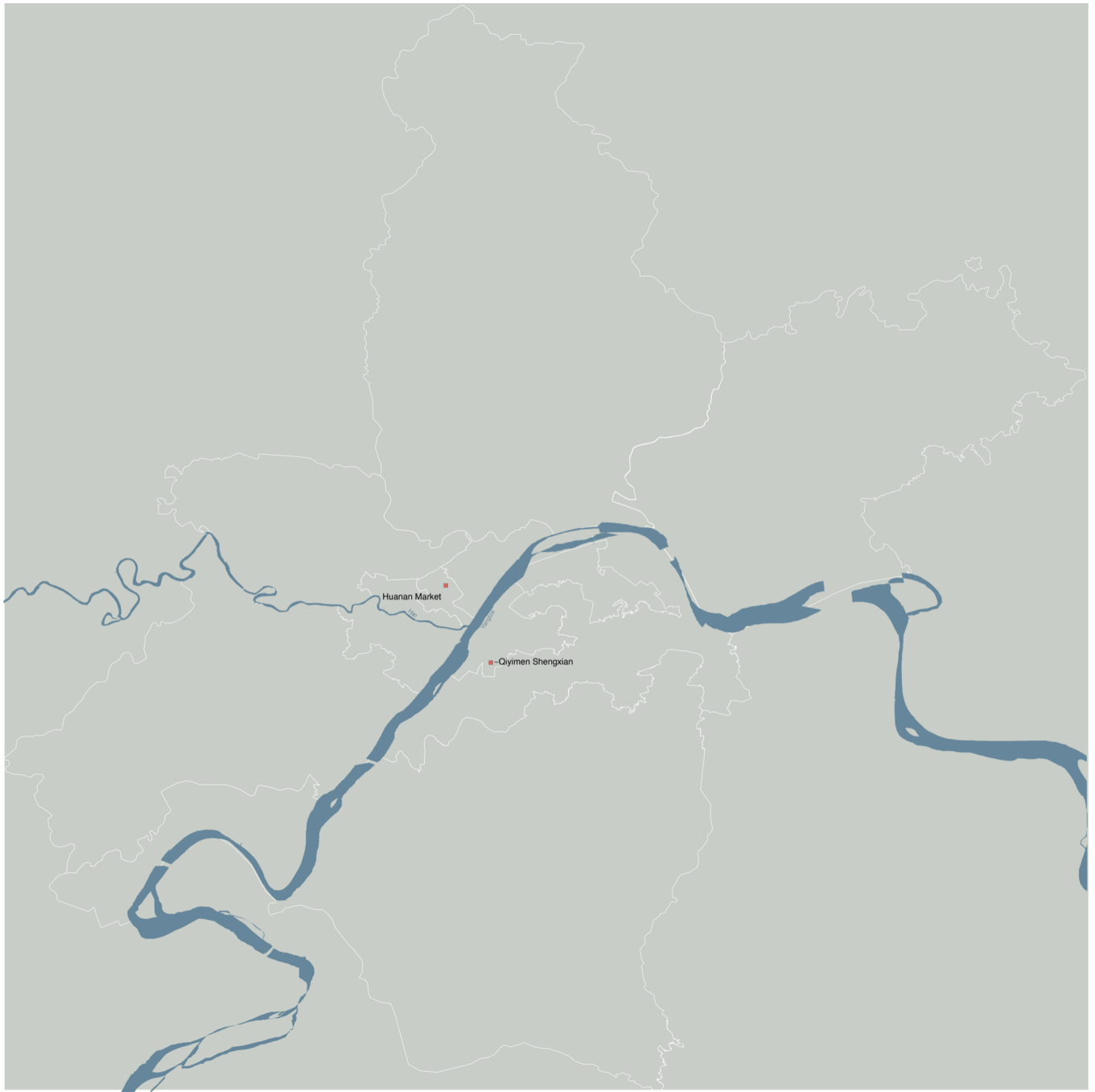


Figure S2. Map of Wuhan. Credit: © OpenStreetMap contributors.

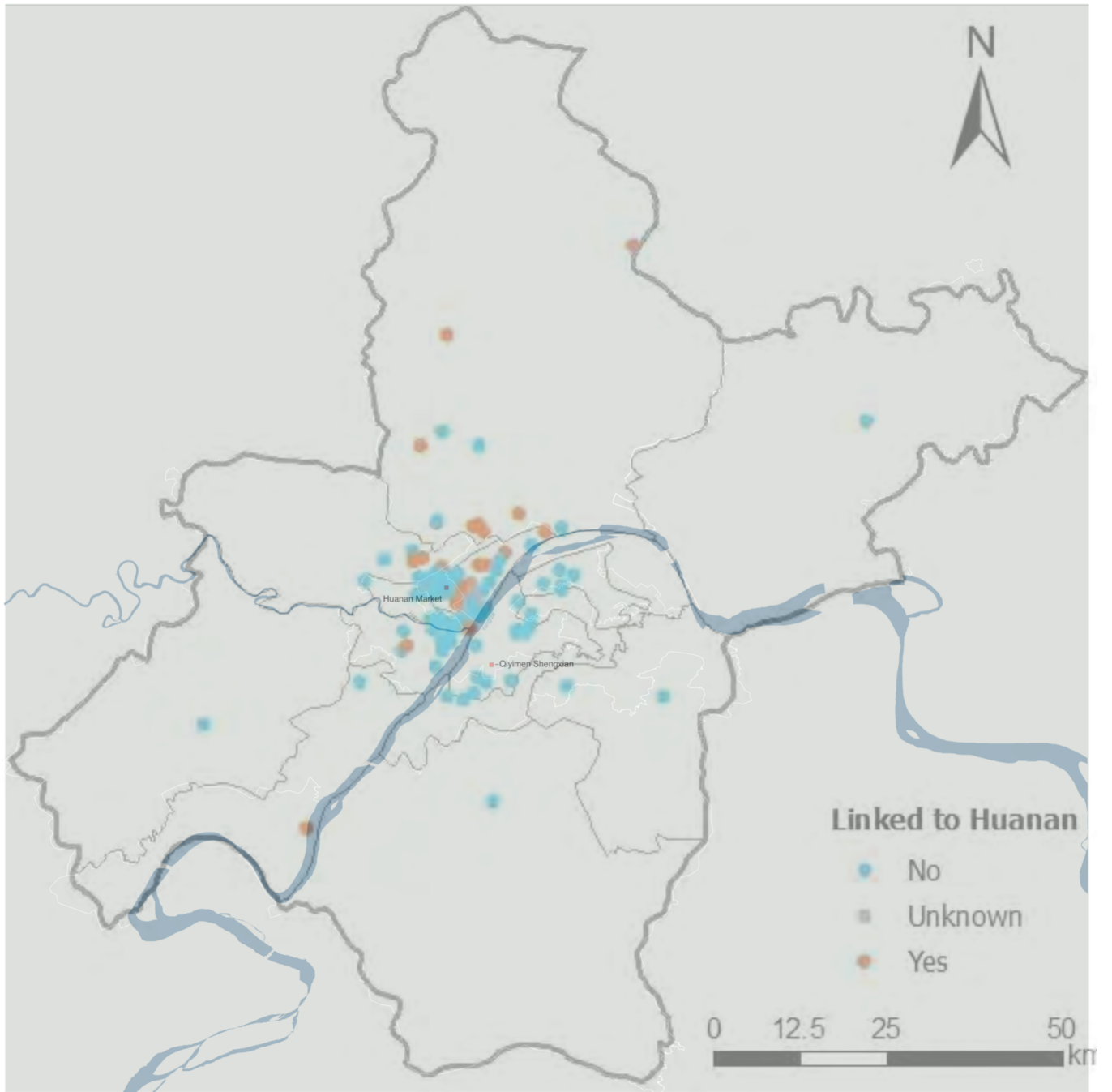


Figure S3. Overlay of Figs. S1 and S2.

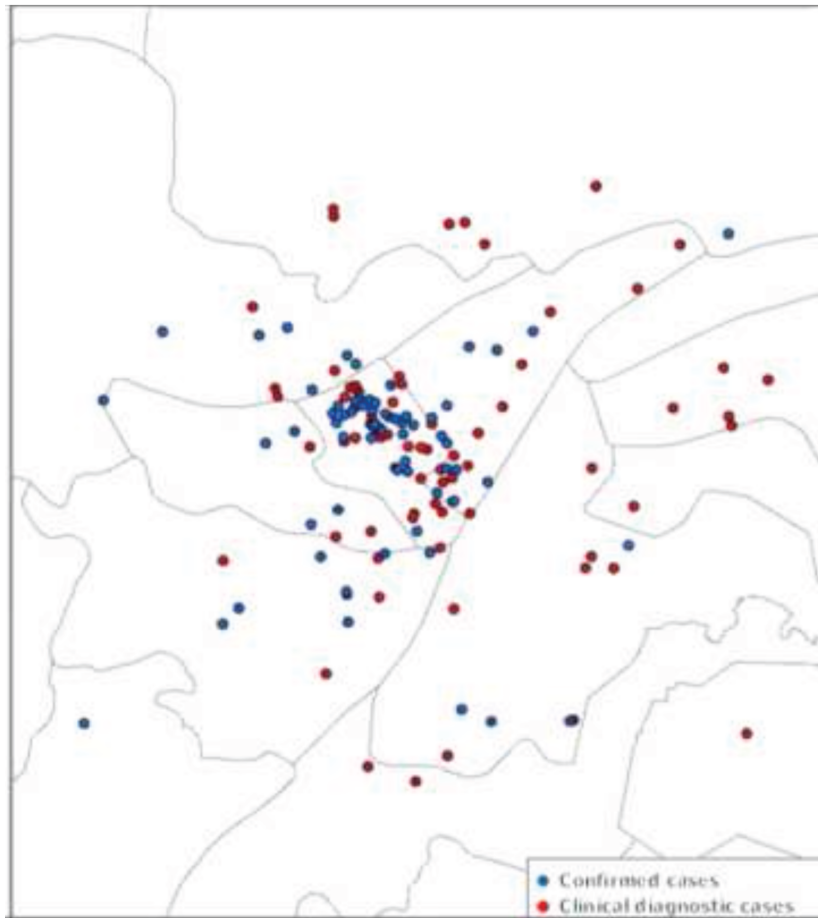


Figure S4. Fig. 23 page 44 from the WHO mission report. The radius of each circle marker each case location is about 125 m.

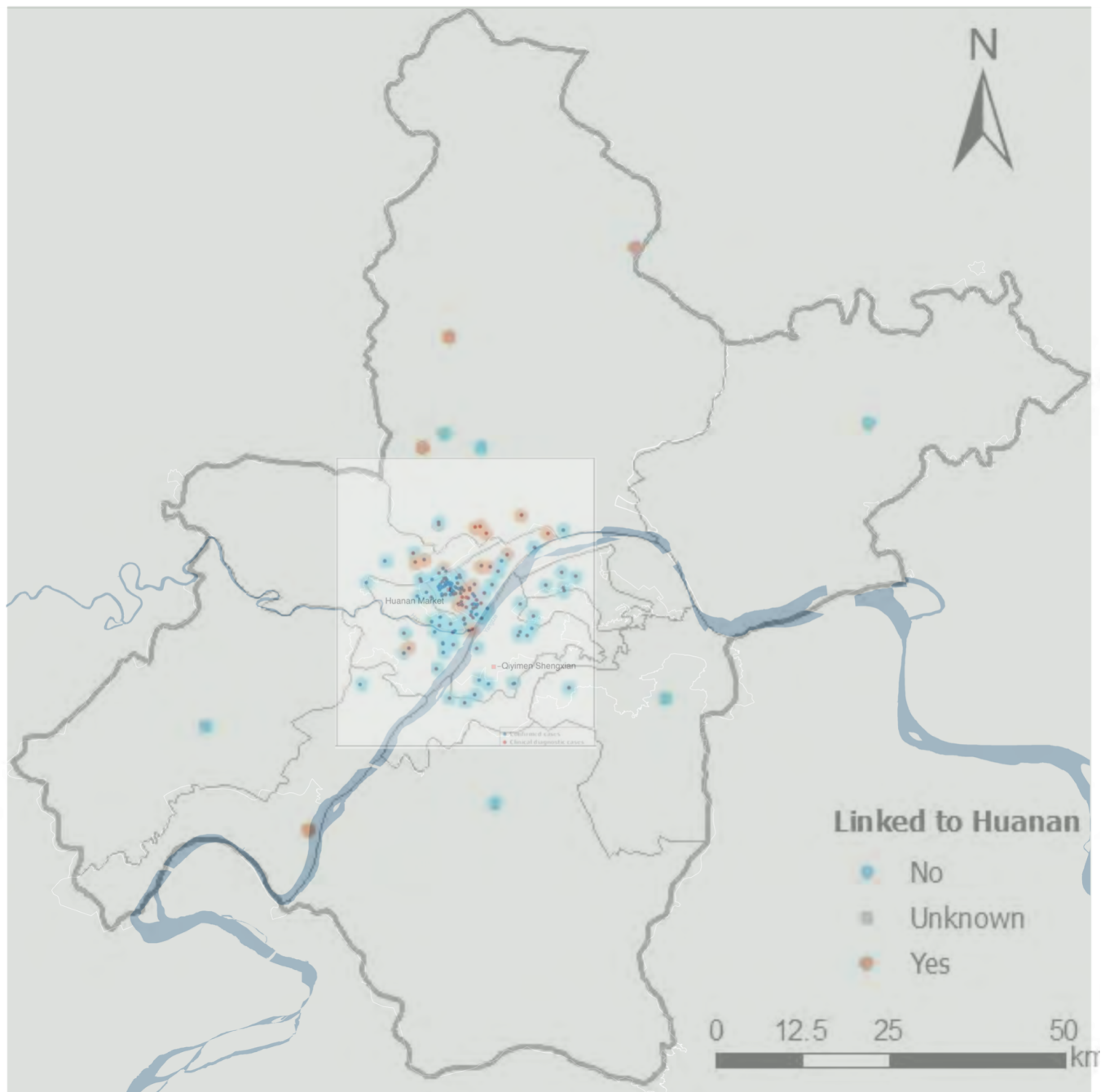


Figure S5. Overlay of Fig. S3 and Fig. 23 from page 44 of the main WHO mission report.

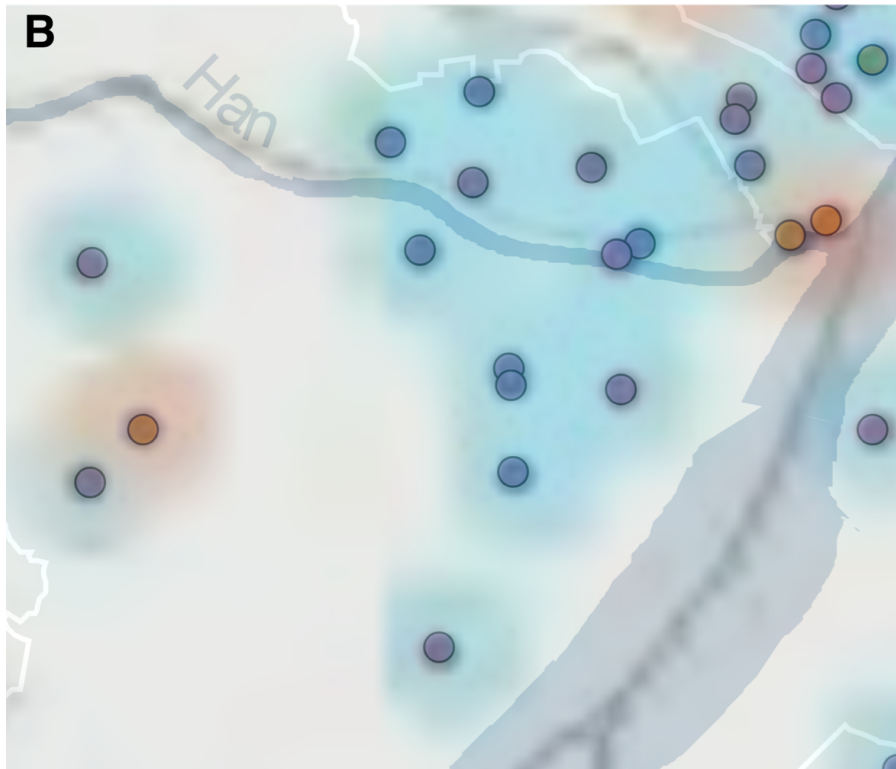
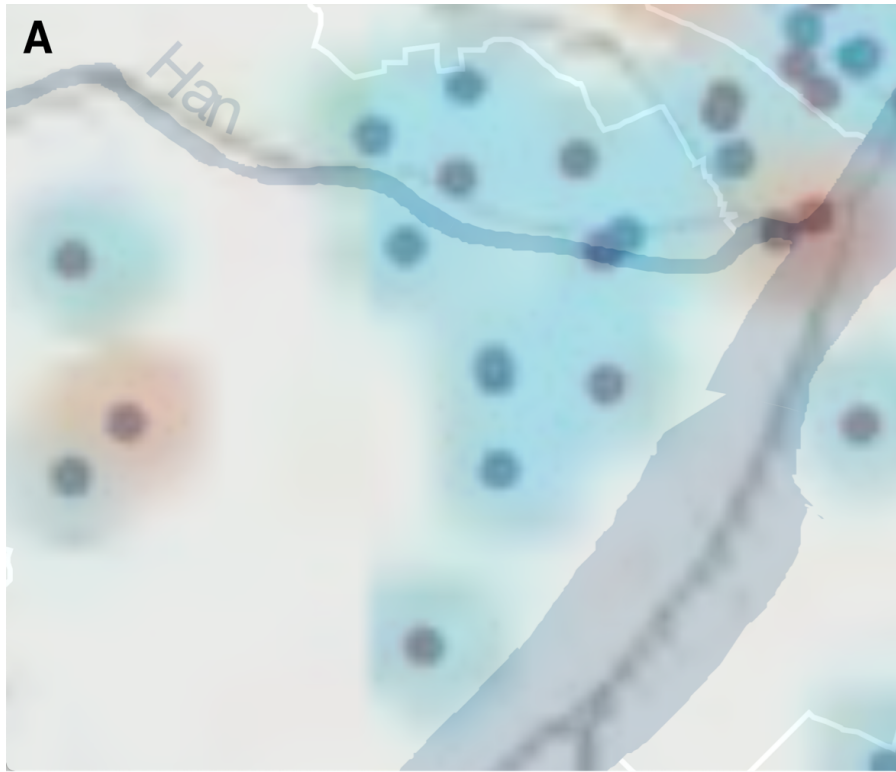


Figure S6. Close-up before and after adding points to the OpenStreetMap map.

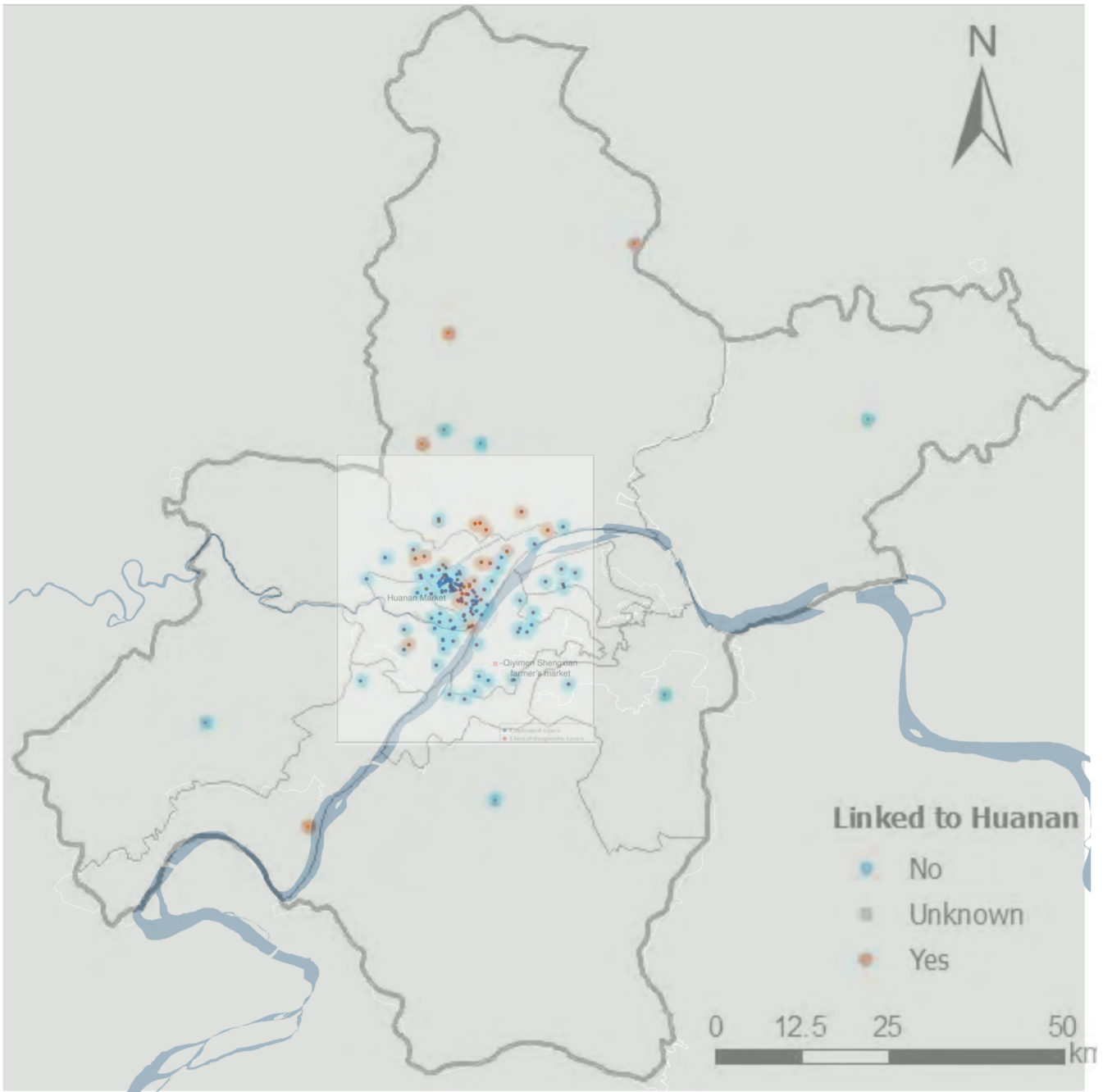


Figure S7. Fig. S4 with all cases added.



Figure S8 - Case map showing family cases and lineage A and lineage B genome sequences.

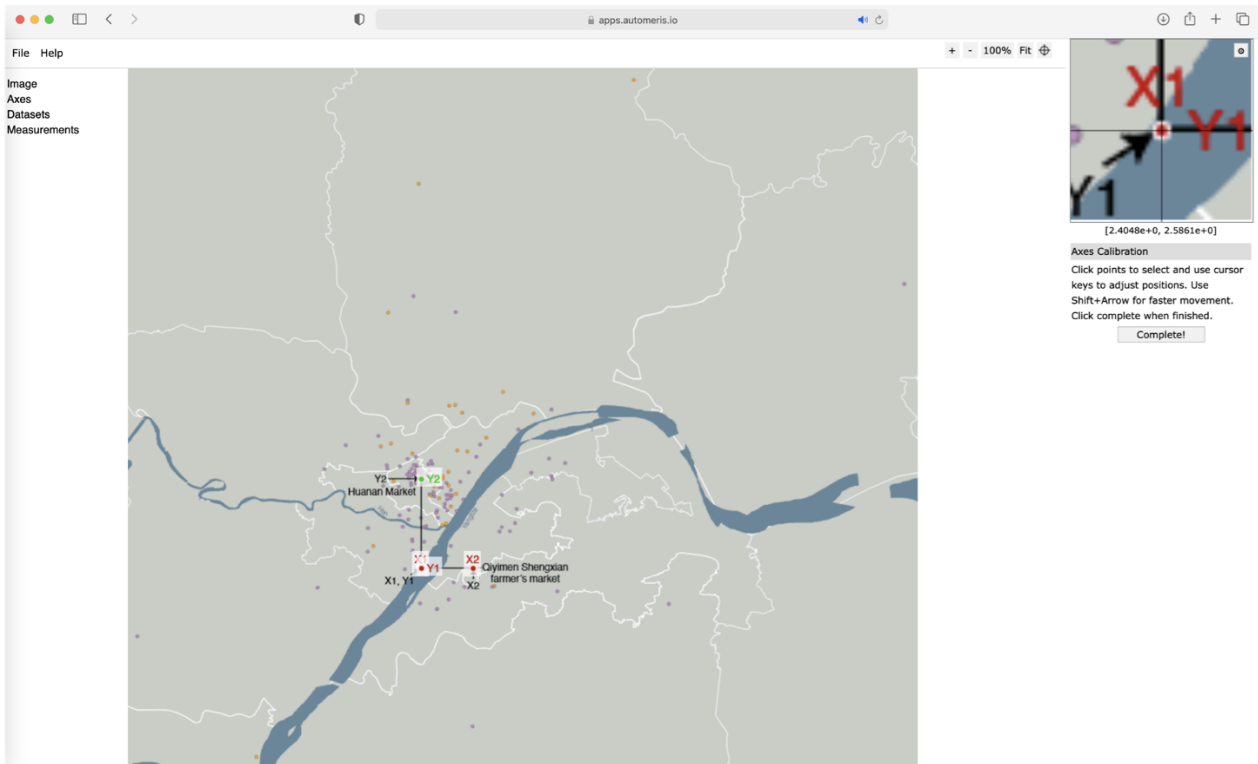
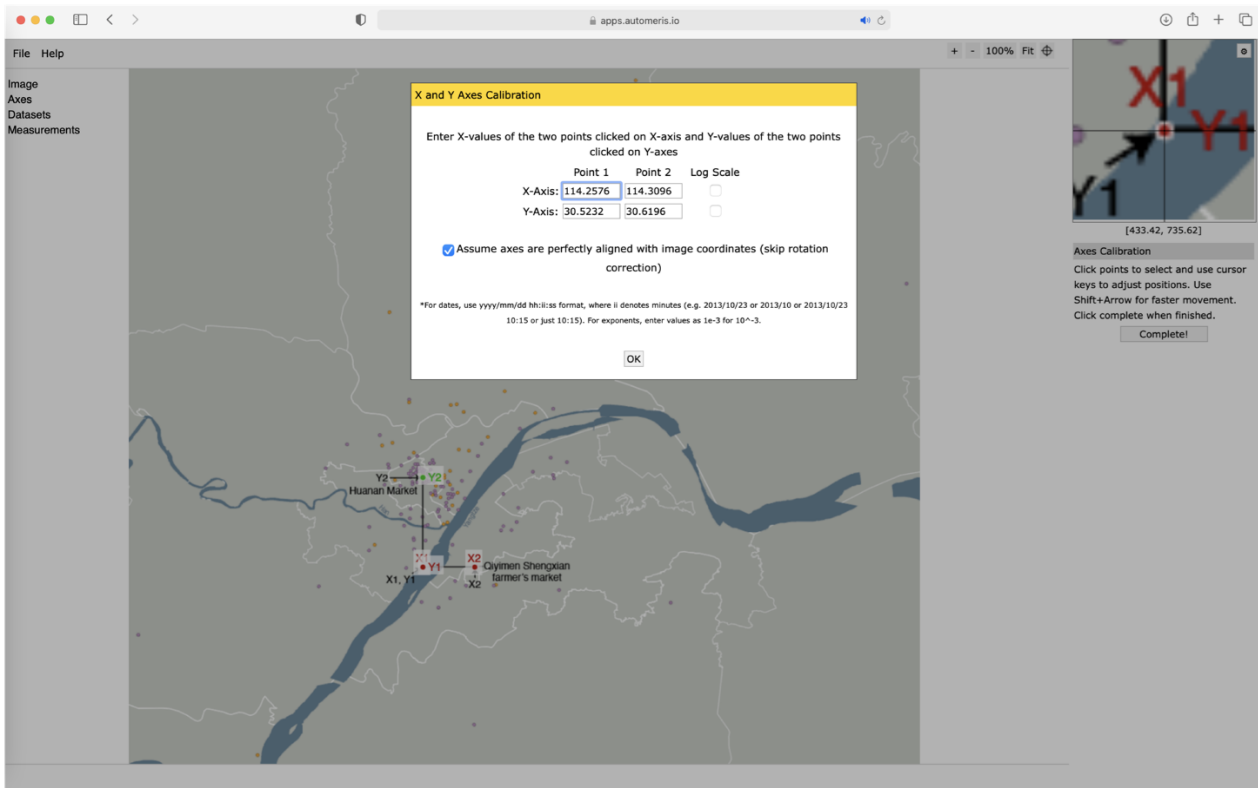
A**B**

Figure S9. WebPlotDigitizer, showing how reference points are input.

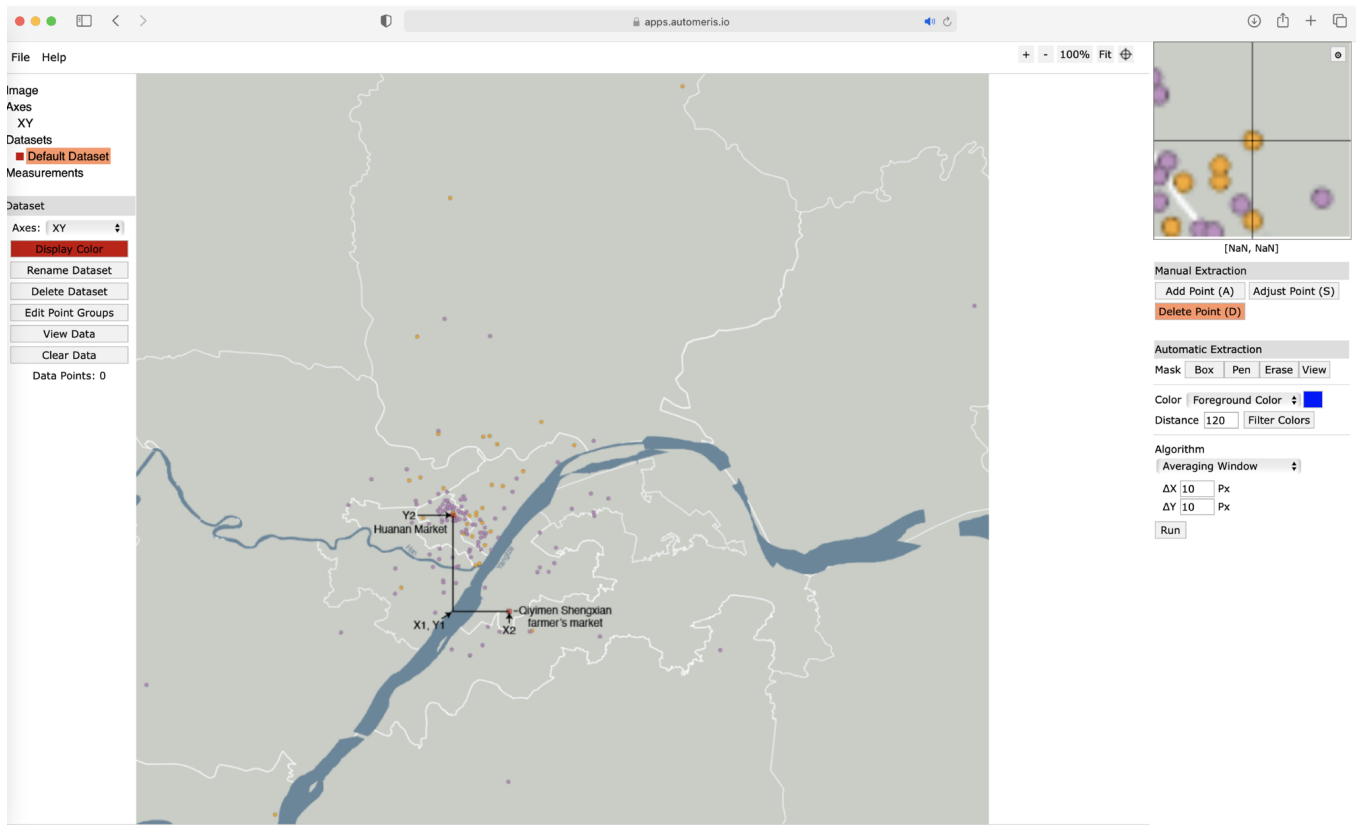


Figure S10. WebPlotDigitizer image, showing how points are added.

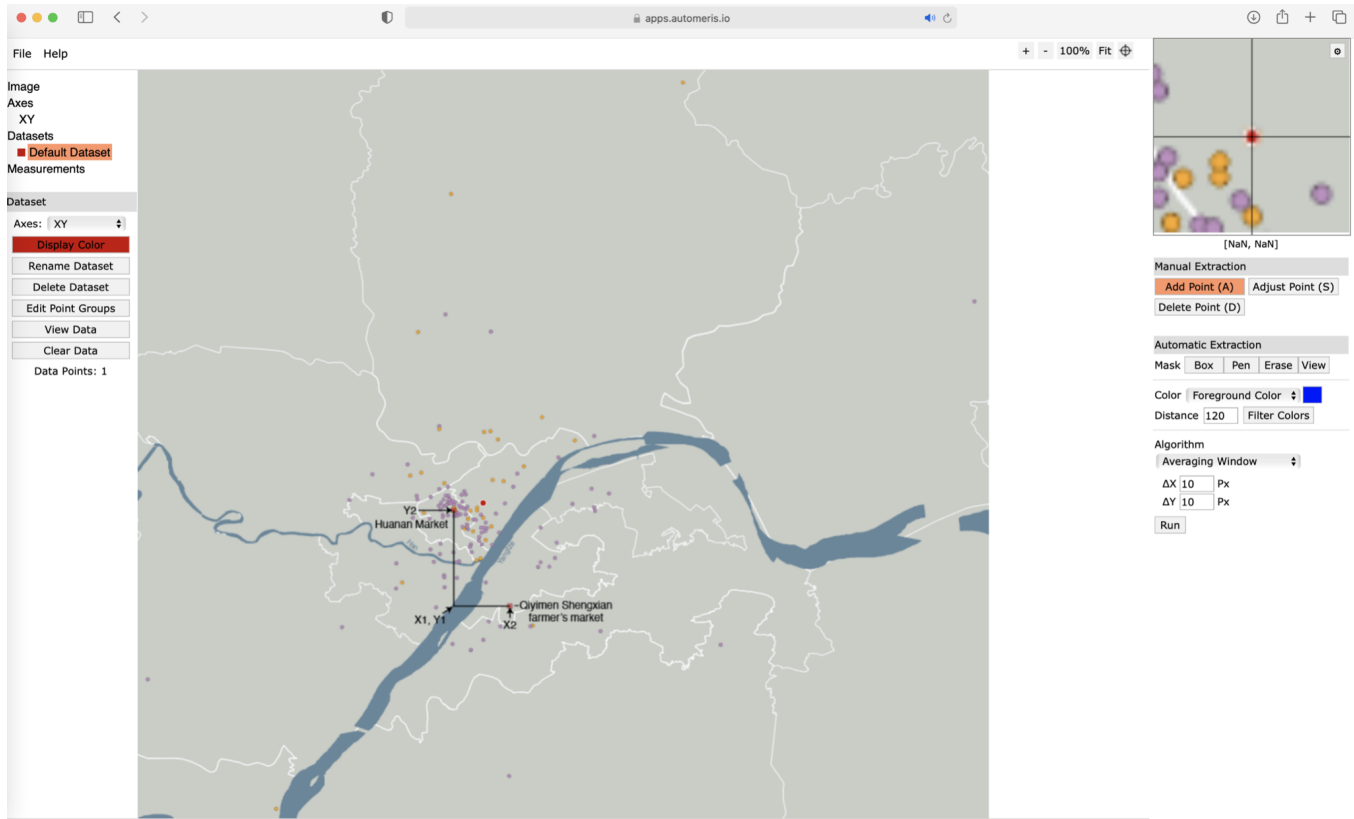


Figure S11. WebPlotDigitizer image, showing how points are added.

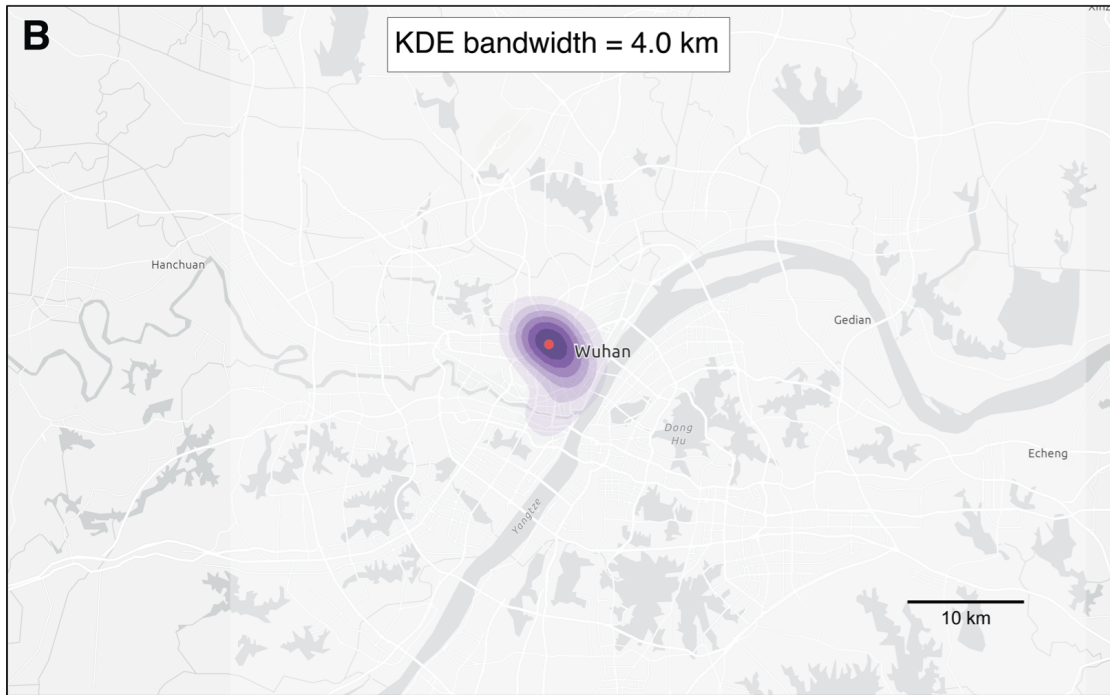
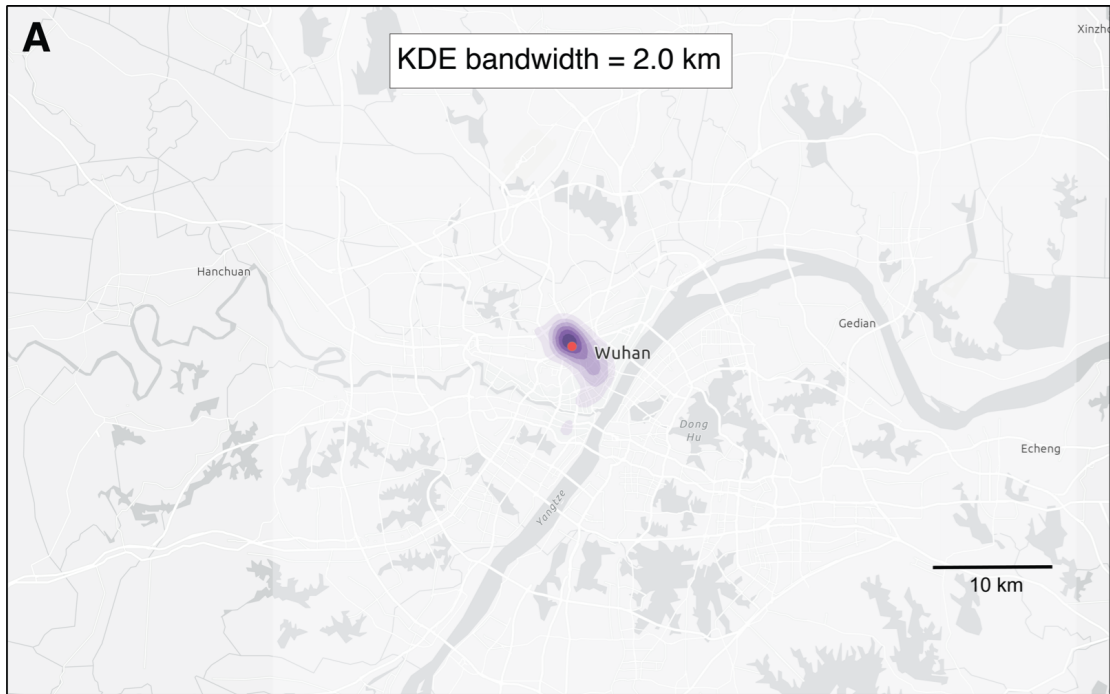


Figure S12. Different bandwidths of KDE.

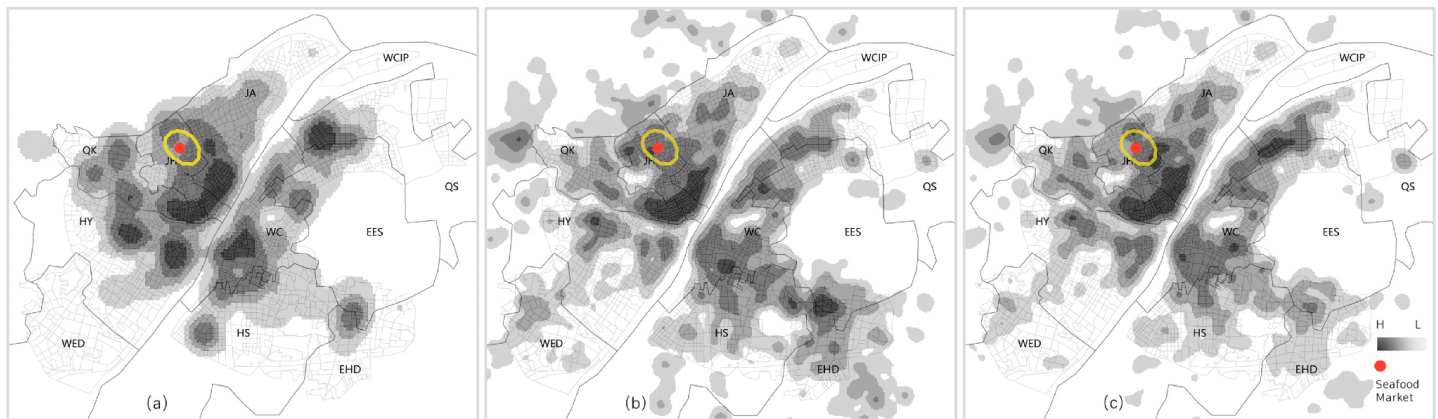


Figure S13. Comparison of highest density area of COVID-19 cases in December 2019 to density of cases, overall population and elderly population in Wuhan in January and February 2020. This figure has been adapted slightly from Peng *et al.* (3) by the addition, in panels a through c of a yellow polygon representing the highest KDE of December 2019 COVID-19 cases from the current study's Fig. 1C. The red circle indicates the location of the Huanan market and is part of the original figure. As described in Peng *et al.* (3), panel a represents COVID-10 cases by Weibo data, b population density, and c elderly population density. Peng *et al.* retain copyright and the figure is made available under the CC BY 4.0 International License.

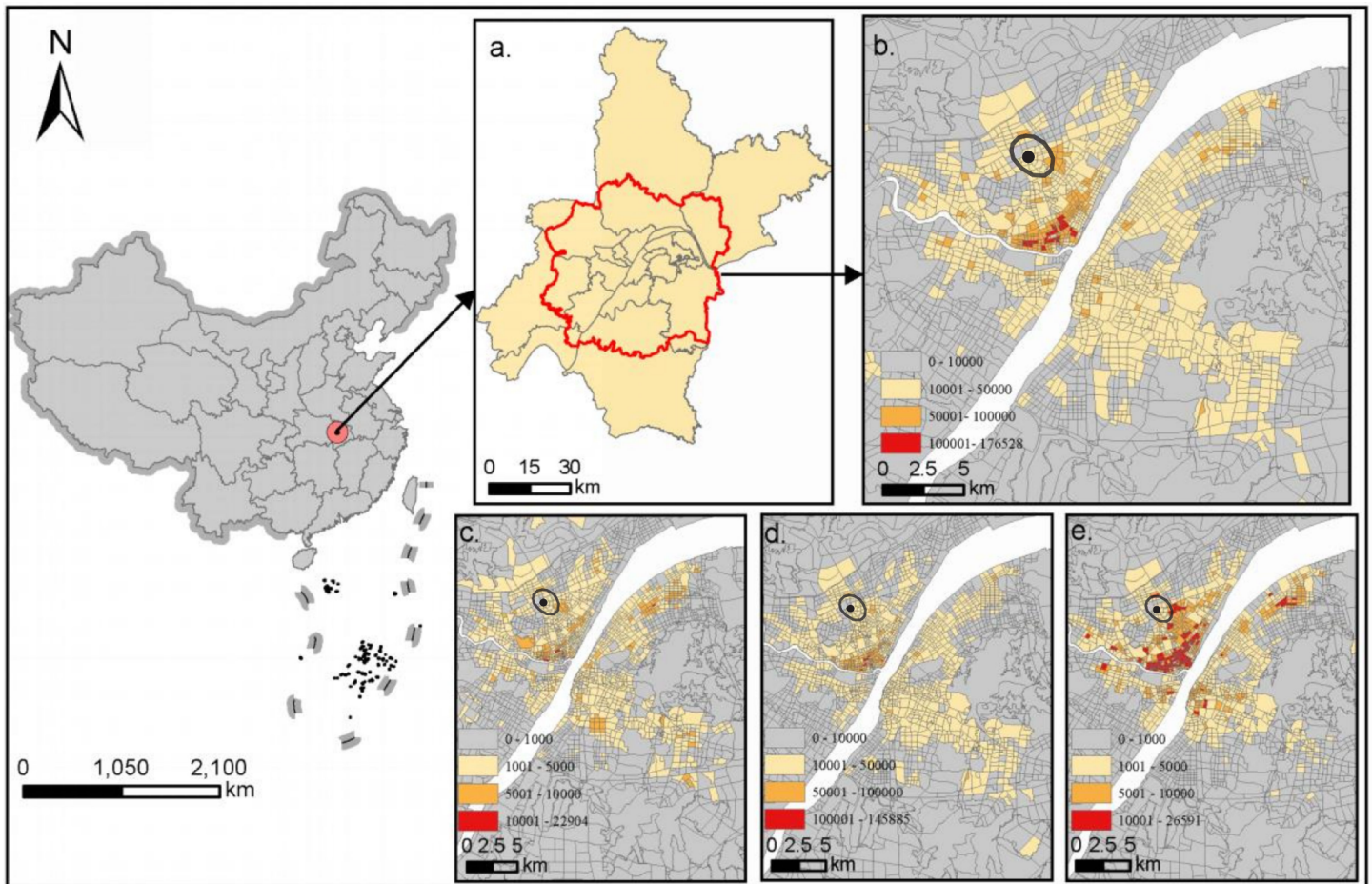


Figure S14. Comparison of highest density area of COVID-19 cases in December 2019 to density of overall population, child population, adult population and elderly population in Wuhan. This figure has been adapted slightly from Jia *et al.* (14) by the addition, in panels **b** through **e** of a polygon representing the highest KDE of December 2019 COVID-19 cases from the current study's Fig. 1C, plus the location of the Huanan market (black circle). As described in Jia *et al.* (14), panel **b** represents overall population density in central Wuhan, **c** children, **d** adult, and **e** elderly population density. Jia *et al.* retain copyright and the figure is made available under the CC BY 4.0 International License.

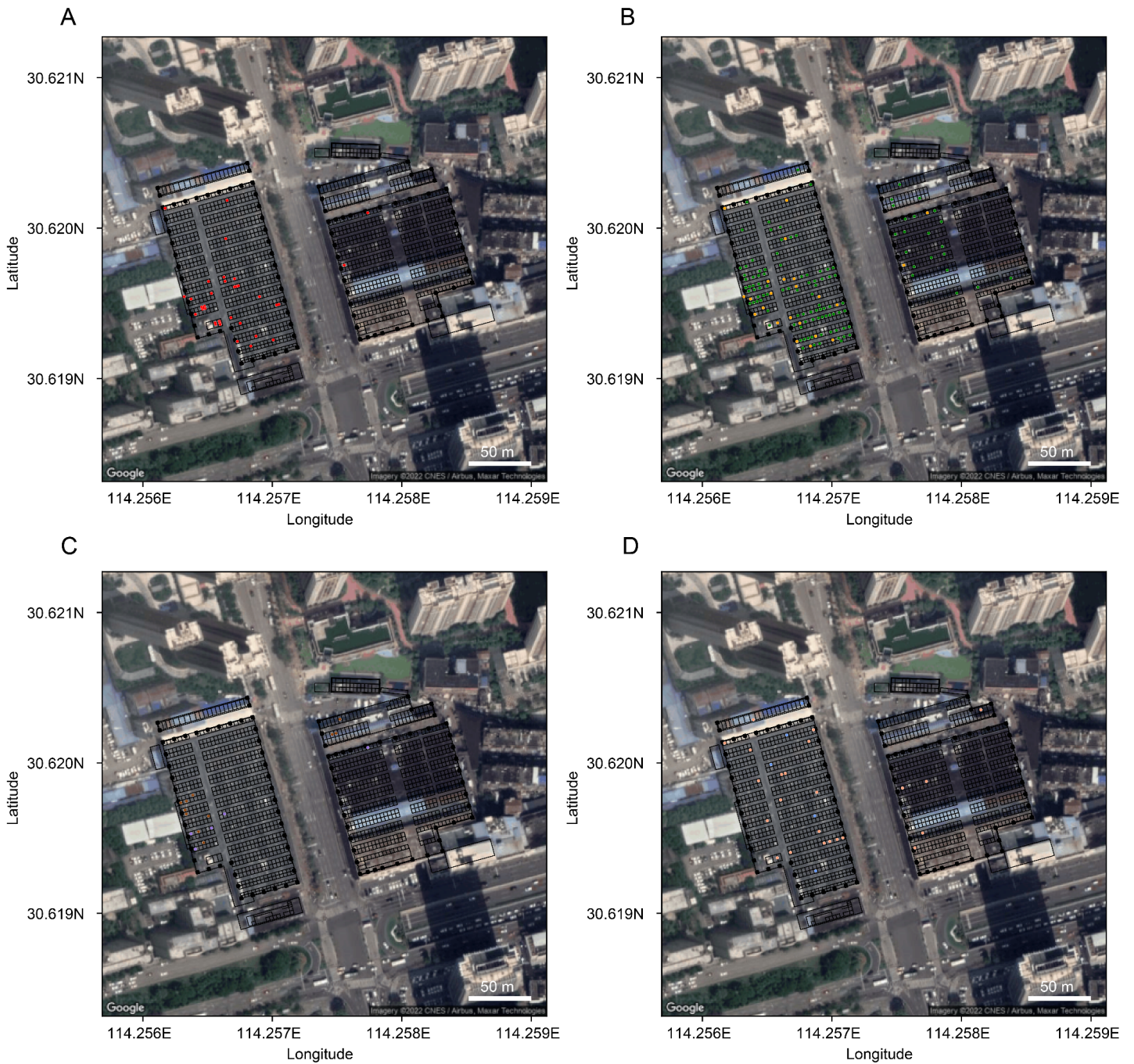
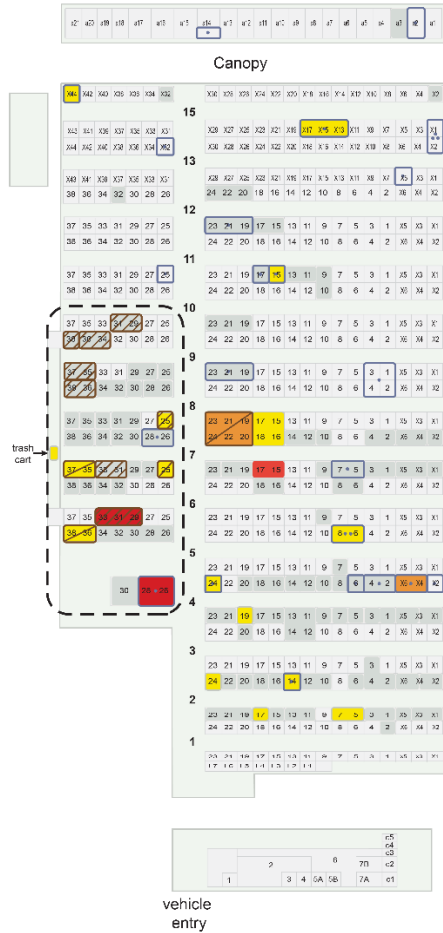


Figure S15. Geospatial mapping of Huanan Seafood Market. A. Geopoints marking the locations of positive environmental samples (red), with jitter added to samples from stalls with multiple positive samples for visualization. B. Centroids of stalls with positive (orange) and negative (green) environmental samples. C. Centroids of known live mammal vendors (dark brown) and unknown meat vendors (lavender). D. Centroids of businesses with human cases up to and including December 20th (light blue), and after December 20th (salmon). Satellite imagery is obtained via the Google Maps Static API using Salem (15).

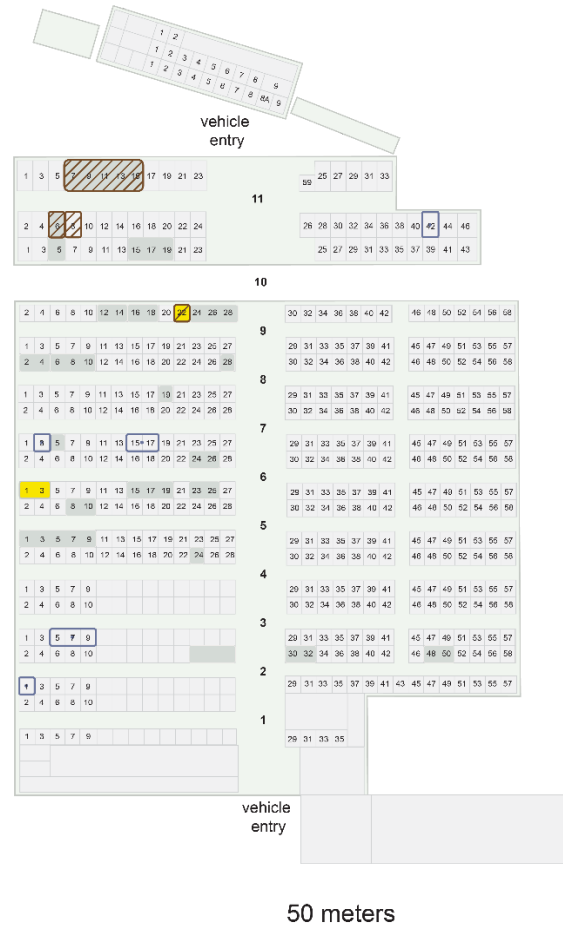
West Side



West Market Environmental Positives

- Live mammals**
 - Street 6, Stalls 29,31,33
 - Street 8, Stall 25
- Seafood**
 - Street 15, Stalls X44
 - Street 2, Stall 14
 - Street 2, Stall 17
 - Street 2, Stall 24
 - Street 2, Stall 14
 - Street 2, Stall 5
 - Street 7, Stalls 15,17
 - Street 7, Stalls 16,18, Street 8, Stalls 15,17
 - Street 11, Stall 19
 - Street 11, Stall 24
 - Street 4, Stalls 26,28
 - Street 4, Stalls X4,X6
 - Street 5, Stalls 6,8
- Unknown meat**
 - Street 8, Stalls 19,21,23, Street 7, Stalls 20,22,24
 - Street 7, Stall 25 (only in CCDC report)
 - Street 5, Stall 36,38 (only in WHO report)
 - Street 7, Stall 35,37 (only in WHO report)
- Meat associated**
 - Trash cart (only in CCDC report)
- Vegetables**
 - Street 11, Stalls 15
- Unknown**
 - Street 15, Stalls X13,X15,X17

East Side



East Market Environmental Positives

- Seafood**
 - Street 6, Stalls 1,3
- Unknown meat**
 - Street 9, Stall 22

Figure S16. Huanan market environmental positives and stall addresses. Map of Huanan market including stall addresses and the type of products sold in each stall with a SARS-CoV-2 positive environmental sample.

Supplementary Tables

Table S1. Significance tests for median distance and centroid distance to the Huanan market and robustness to location imprecision and missing cases.

Data	Null	Statistic	
		Median distance to Huanan market	Centroid distance to Huanan market
156 cases	Weibo	$p < 0.001$	$p = 0.002$
	worldpop.org	$p < 0.001$	$p = 0.002$
123 cases, not linked	Weibo	$p < 0.001$	$p = 0.002$
	worldpop.org	$p < 0.001$	$p = 0.002$
33 cases, linked	Weibo	$p < 0.007$	$p = 0.064$
	worldpop.org	$p < 0.001$	$p = 0.02$
156 cases resampled with location uncertainty ($r = 0.5$ km)*	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.015)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.004)
156 cases resampled with location uncertainty ($r = 1.0$ km)	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.015)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.005)
123 cases, not linked, resampled with location uncertainty ($r = 0.5$ km)	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.015)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.001 - 0.004)
123 cases, not linked, resampled with location uncertainty ($r = 1.0$ km)	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.022)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (>0.001 - 0.005)
33 cases, linked, resampled with location uncertainty ($r = 0.5$ km)	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.064$ (0.039 - 0.081)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.021$ (0.012 - 0.031)
33 cases, linked, resampled with location uncertainty ($r = 1.0$ km)	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.065$ (0.034 - 0.094)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.021$ (0.010 - 0.038)
156 cases resampled with location uncertainty ($r = 0.5$ km), and 8 added “missing” cases**	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.015)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.001 - 0.003)
156 cases resampled with location uncertainty ($r = 1.0$ km), and 8 added “missing” cases	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.015)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.001 - 0.003)
33 linked cases resampled with location uncertainty ($r = 0.5$ km), and 8 added “missing” cases	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.034$ (0.002 - 0.051)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.009$ (0.002 - 0.016)

33 linked cases resampled with location uncertainty ($r = 1.0$ km), and 8 added “missing” cases	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.034$ (0.002 - 0.051)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.009$ (0.002 - 0.016)
102 subsampled cases, not linked	Weibo	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (0.002 - 0.022)
	worldpop.org	$p < 0.001$ ($p < 0.001$)	$p = 0.002$ (<0.001 - 0.006)

*For the analysis with resampled cases with location uncertainty, we report the P-value for the median distance and centroid distance averaged over 1000 resamplings (cfr. Methods in Supplementary Material) as well as the range of P-values (in between brackets) for the 1000 individual resamplings. The r value represents the radius of the circle polygon with the extracted locations as the center from which new locations are resampled.

**See Supplementary Text.

Table S2. Environmental samples from the Huanan market.

Code	Area	Street	Stall	Store type	Known cases in stall	Wildlife area?	Illegal Hedgehog sales 2019	Type	RT-PCR	N gene seq.	S gene seq.	NGS	GISAIID Sequence	Length	Coverage	GISAIID Culture sequence
E61	East	6	1,3	Seafood				Ground surface	POS							
E48	East	9	22	Meat, unknown				Blood from ground surface, in front of door	POS	>99%						
D32	West	15	X15		Yes			Wagon surface	POS							
B17	West		X44	Seafood	Yes			Scale	POS							
G93	West	8	19, 21, 23	Meat, unknown				Dirty water	POS							
F54	West	2	14	Seafood	Yes			Ground surface	POS	>99%		>99%	EPI_ISL_4_08512	25,342	2.4	
F33	West		17	Seafood				Sluice gate	POS							
F46	West		24	Seafood				Ground surface	POS	>99%	>99%					
E7	West		5,7	Seafood				Foam dispenser	POS	>99%						
A20	West	7		Seafood				Gloves	POS	>99%	>99%					
A18	West		15, 17	Seafood				Shoe bottoms	POS	>99%						
A2	West			Seafood				Ground surface	POS	>99%						
A63	West		16, 18	Seafood				Ground surface	POS		>99%					
A61	West		20, 22, 24	Meat, unknown	Adjacent			Ground surface	POS							
A55	West		25	Meat, unknown†††	Adjacent	Yes		Ground surface	POS	>99%						
A33	West		West end	Meat associated		Yes		Trash cart	POS							
F13	West	11	15	Vegetable	Yes			Wall surface	POS	>99%	>99%	>99%	EPI_ISL_4_08511	28,557	25.6	EPI_ISL_408514 EPI_ISL_408515
A101	West	4	19	Seafood				Door surface	POS							
A96	West		24	Seafood	Yes			Ground surface	POS							
A14	West		26	Seafood	Yes	Yes		Fish packaging surface	POS							
A15	West			Seafood	Yes	Yes		Door surface	POS							
A90	West			Seafood	Yes	Yes		Ground surface	POS							
A87	West		28	Seafood	Yes	Yes		Door surface	POS							
A88	West			Seafood	Yes	Yes		Ground surface	POS							
F98	West		X6, X4	Seafood	Yes			Ground surface	POS	>99%						
F100	West	Seafood		Yes			Ground surface	POS	>99%							
B5	West	5	6, 8	Seafood	Yes			Ground surface	POS	>99%		>99%	EPI_ISL_4_08513	1,065	2.4	
Q37	West	8	25	Domesticated wildlife††		Yes	Yes†	Interior surface of freezer	NEG			>99%				

Q61	West	6	29	Domesticated wildlife	Yes	Yes†	Cart 1	NEG	>99%						
Q64	West		29	Domesticated wildlife	Yes	Yes†	Cart 2	POS	>99%						
Q68	West		29	Domesticated wildlife	Yes	Yes†	Ground surface	POS	>99%						
Q69	West		29	Domesticated wildlife	Yes	Yes†	Hair/feather removal machine (epilator)	POS	>99%						
Q70	West		29	Domesticated wildlife	Yes	Yes†	Metal cage in inner room	NEG	>99%						
WHO	West	5	36,38	Meat, unknown†††	Yes		Unknown	POS							
WHO	West	7	35,37	Meat, unknown	Yes		Unknown	POS							

Methodology: Business names were obtained from the TianYanCha.com business directory (see Table S5) based on addresses. In many cases, business names were further confirmed through photographic evidence of stalls. Two stalls were identified as involved in domesticated wildlife sales from an official local forestry bureau fine for their registered owners for illegal hedgehog sales in summer 2019 (11).

† The "Operators" of these businesses listed in TianYanCha.com business directory were fined for illegal hedgehog sales in summer 2019, in an official government report (11)

†† Reported in WHO mission report Appendix F, Table 3, contains "game" in the business name, lists game retail as a service in the business directory, and was fined for illegal hedgehog sales in 2019 (see †) (1).

††† Store selling meat products inferred from photographic evidence.

Table S3 . Live mammals traded at the Huanan market, October - December, 2019

Species	Family/Order	Susceptibility	Observed at Huanan market, October 2019	Observed at Huanan market, November 2019	Observed at Huanan market, December 2019*
Raccoon dog (<i>Nyctereutes procyonoides</i>)**	Canidae/Carnivora	Live*** (16)(17)(18)	Yes	Yes	Yes
Amur hedgehog (<i>Erinaceus amurensis</i>)	Erinaceidae/Eulipotyphla		Yes	Yes	
Hog badger (<i>Arctonyx albogularis</i>)	Mustelidae/Carnivora	ACE2**** (18)	Yes	Yes	
Asian badger (<i>Meles leucurus</i>)	Mustelidae/Carnivora	See raccoon dog, mink, etc.	Yes	Yes	
Chinese hare (<i>Lepus sinensis</i>)	Leporidae/Lagomorpha	Rabbit: (19)	Yes	Yes	
Chinese bamboo rat (<i>Rhizomys sinensis</i>)	Spalacidae/Rodentia	ACE2 (20)	Yes	Yes	
Malayan porcupine (<i>Hystrix brachyura</i>)	Hystriidae/Rodentia	See bamboo rat	Yes	Yes	Yes
Chinese muntjac (<i>Muntiacus reevesi</i>)	Cervidae/Artiodactyla	White-tailed deer: Live (21)	Yes	Yes	Yes
Marmot (<i>Marmota himalayana</i>)	Sciuridae/Rodentia	See bamboo rat	Yes	Yes	Yes
Red fox (<i>Vulpes vulpes</i>)	Canidae/Carnivora	Live (22), Serology***** (23)	Yes	Yes	Yes
Siberian weasel (<i>Mustela sibirica</i>)	Mustelidae/Carnivora	See raccoon dog, mink, etc.	No	No	
Pallas's squirrel (<i>Callosciurus erythraeus</i>)	Sciuridae/Rodentia	See bamboo rat	No	No	
Masked palm civet (<i>Paguma larvata</i>)	Viverridae/Carnivora	ACE2 (18)(20)(24)	No	No	
Coypu (<i>Myocastor coypus</i>)	Echimyidae/Rodentia	See bamboo rat	No	No	
Mink (<i>Neovison vison</i>)**	Mustelidae/Carnivora	Live (25)(26)(27)(28)(29)	No	No	
Red squirrel (<i>Sciurus vulgaris</i>)	Sciuridae/Rodentia	See bamboo rat	No	No	
Wild boar (<i>Sus scrofa</i>)	Suidae/Artiodactyla	Serology (23), ACE2 (24)(30)	No	No	
Complex-toothed flying squirrel (<i>Trogopterus xanthipes</i>)	Sciuridae/Rodentia	See bamboo rat	No	No	

*From photographic evidence from 3 December, 2019 (13).

**Established transmissibility, not just susceptibility.

***Demonstrated susceptibility in live animals

****Based on serological findings in wild populations

*****Based on predicted or demonstrated ACE2 binding.

Table S4. Key numbers regarding environmental sampling at the Huanan market

Key numbers		Source
585	Total number of Environmental Samples, CCDC	CCDC page 1
26	Number of positive stalls, CCDC	From CCDC table
33	Total number of Positive samples, CCDC	CCDC page 1
31	Number of positive samples from western section, CCDC	CCDC page 2
718	Total number of Environmental Samples, WHO	WHO page 95
678	Total number of stalls in the market, WHO	WHO page 96
134	Number of stalls sampled in the market, WHO	WHO page 96
21	Number of positive stalls, WHO	WHO Annex E4, 4.5 Market environment
113	Number of negative stalls, WHO	WHO Annex E4, 4.5 Market environment
40	Number of positive samples, WHO	WHO page 95
69	Number of positive samples including sewage and warehouses, WHO	WHO page 95
174	All December 2019 cases	WHO page 43 (Fig 22)
164	All Wuhan December 2019 cases	WHO page 43
55	Cases connected to Huanan Market	WHO, Annex E4 "Table 3. General information of cases"
30	Huanan Market Vendor cases	WHO, Annex E4 "Table 3. General information of cases"

Table S5. Public business registry of specific stalls with SARS-CoV-2 positive samples at the Huanan market.

Street	Address	Tianyancha link
8	25	https://www.tianyancha.com/company/2336478151
5	"Corridor" (36,38)	https://www.tianyancha.com/company/2334674880
7	25,27	https://www.tianyancha.com/company/2339714863
7,8	19,20,21,22,23,24	https://www.tianyancha.com/company/2336843423
6	29,31,32,33	https://www.tianyancha.com/company/2336320983
2	14	https://www.tianyancha.com/company/1434754801

Table S6. SARS-CoV-2-negative environmental samples from the Huanan market

Section of market	Street	Stall
West		A3
West	15	X2
West	15	X32
West	12	20-24
West	12	32
West	12	15-17
West	12	19-23
West	11	17
West	11	11-13
West	11	9
West	10	19-23
West	10	29-31
West	9	34-36
West	9	38
West	9	19-23
West	9	27-29
West	9	25
West	9	35-37
West	8	26-28
West	8	30
West	8	32-34
West	8	36-38
West	8	29-31
West	8	33
West	8	35
West	8	37
West	7	X2
West	7	X4-X6
West	7	2-4
West	7	10
West	7	12-14
West	7	26-28
West	7	30-32
West	7	34
West	7	36-38
West	7	31-33
West	7	29

West	7	27
West	7	25
West	7	23
West	7	19-21
West	7	9
West	7	5-7
West	7	1-3
West	7	X3-X5
West	7	X1
West	6	38
West	6	36
West	6	34
West	6	32
West	6	16-18
West	6	6-8
West	6	9
West	5	26-34
West	5	7
West	4	30
West	4	20
West	4	16-18
West	4	12-14
West	4	10
West	4	8
West	4	6
West	4	2-4
West	4	X4-X6
West	4	X2
West	4	21-23
West	4	15-17
West	4	11-13
West	4	9
West	4	7
West	4	5
West	4	3
West	4	1
West	4	X3-X5
West	4	X1

West	3	20
West	2	22
West	2	20
West	2	18
West	2	16
West	2	10-12
West	2	6
West	2	2-4
West	2	X4-X6
West	2	X2
West	2	19-23
West	2	15
West	2	11-13
West	2	1-3
West	2	X5
West	2	X1-X3
West	1	2
East	11	7-15
East	11	6
East	10	17-21
East	10	5
East	9	24-26
East	9	14-16
East	9	8-12
East	8	2-4
East	8	6-10
East	8	28
East	8	19
East	7	5
East	6	24-26
East	6	23-25
East	6	15-19
East	5	8-10
East	5	1-9
East	4	24
East	2	24-28
East	2	30-32
East	2	48-50

Table S7. Human SARS-CoV-2-positive cases at the Huanan market

Source	Location	Street	Stalls	Number	Notes	Date (before/by)
Media	West	2	14	1	10 December 2019 onset; not in WHO report (7)	12-10
WHO report	West	12	19,21,23	1		12-13
WHO report	West	5	6,8	1		12-20
WHO report	West	5	6,8	1		12-27
WHO report	West	7	5,7	1		12-20
WHO report	West	13	X2	2	also stall 15 X1	12-27
WHO report	West	13	X2	1	also stall 15 X1	12-31
WHO report	West	15	X13,X15,X17	1		12-20
WHO report	West		A2	1		12-20
WHO report	West	4	26,28	1		12-31
WHO report	West	4	24	1		12-31
WHO report	West	4	X2	1		12-31
WHO report	West	4	X4,X6	1		12-31
WHO report	West	4	2,4	1		12-31
WHO report	West	4	6	1		12-31
WHO report	West	7	26,28	1		12-31
WHO report	West	9	12,21,23	1		12-31
WHO report	West	8	2,4	1	also stall 9 1,3	12-31
WHO report	West	11	25	1		12-31
WHO report	West	11	15,17	1		12-31
WHO report	West	13	X32	1		12-31
WHO report	West	13	X5	1		12-31
WHO report	West		A15	1		12-31
WHO report	West	15	X44	1		12-31
WHO report	East	2	1	1		12-31
WHO report	East	3	5,7,9	1		12-31
WHO report	East	7	3	1		12-31
WHO report	East	7	19,21	1		12-31
WHO report	East	11	42	1		12-31

Table S8 : Binomial GLM results for possible values of total samples (N) per stall

Samples per stall	N=5	N=6	N=7	N=8	N=9	N=10
Estimated coefficient (wildlife stall distance)	-0.0290	-0.0287	-0.0285	-0.0283	-0.0282	-0.0281
Detection odds % decrease/m (wildlife stall distance)	2.86	2.83	2.81	2.79	2.78	2.77
P-value (wildlife stall distance)	0.00419	0.00434	0.00445	0.00454	0.00461	0.00466
Estimated coefficient (case distance)	-0.0692	-0.0684	-0.0679	-0.0675	-0.0672	-0.0670
Detection odds % decrease/m (case distance)	6.68	6.62	6.57	6.53	6.50	6.48
P-value (case distance)	0.01323	0.01355	0.01379	0.01396	0.01410	0.01421

Table S9. Binomial GLM results for possible values of total samples (N) per stall, including unknown meat vendors and wildlife vendors

Samples per stall	N=5	N=6	N=7	N=8	N=9	N=10
Estimated coefficient (wildlife stall distance)	-0.0354	-0.0350	-0.0347	-0.0345	-0.0343	-0.0342
Detection odds % decrease/m (wildlife stall distance)	3.48	3.44	3.41	3.39	3.38	3.36
P-value (wildlife stall distance)	0.00242	0.00252	0.00258	0.00264	0.00268	0.00271
Estimated coefficient (case distance)	-0.0764	-0.0756	-0.0750	-0.0745	-0.0742	-0.0739
Detection odds % decrease/m (case distance)	7.36	7.28	7.22	7.18	7.15	7.12
P-value (case distance)	0.00752	0.00772	0.00787	0.00798	0.00806	0.00813

Table S10. Distances from the Huanan market to the twenty nearest hotels

Hotel	Distance to Huanan market (m)*	Coordinates
Jinjiang Inn Wuhan Hankou Railway Station Branch	98 m**	30.6194 N, 114.2586 W
Non-Fan Premium Hotel (Wuhan Hankou Railway Station)	117 m	30.6194 N, 114.2588 W
100 Inns & Hotel (Wuhan Hankou Railway Station Branch)	126 m	30.6194 N, 114.2589 W
Dongfang Jianguo Hotel	402 m	30.6204 N, 114.2535 W
Lavande Hotels (Hankou Railway Station Grand Wuhan 1911 Branch)	183 m	30.6187 N, 114.2592 W
Sucha Hotel (Wuhan Hankou Railway Station)	308 m	30.6199 N, 114.2608 W
Tianya 1911 Hotel	279 m	30.6171 N, 114.2573 W
Building Lan Business Hotel	293 m	30.621 N, 114.255 W
Yibai Liangpin Hotel	422	30.6193 N, 114.2532 W
Lingbo Fashion Hotel	339	30.6207 N, 114.2543 W
Super 8 Hotel (Hankou Railway Station Branch)	422	30.6193 N, 114.2532 W
Huayi Selected Hotel	481	30.6192 N, 114.2526 W
Caiguang Hostel	491	30.6191 N, 114.2525 W
Kaixin Express Hotel	422	30.6193 N, 114.2532 W
Xana Hotel Hankou Railway Station	404	30.6182 N, 114.2537 W
Meirujia Hotel (Hankou Railway Station Branch)	336	30.6193 N, 114.2541 W
Hanting Hotel Hankou Railway Station Branch	364	30.6198 N, 114.2538 W
City Comfort Inn Chain Hankou Railway Station	407	30.619 N, 114.2534 W
Huangting Business Motel Hankou Railway Station	409	30.6189 N, 114.2534 W
7 days Inn Hankou Railway Station	402	30.6195 N, 114.2534 W

*Distance from hotel to central market point (30.6196 N, 114.2576 W); median = 383 m.

**Distance to the closest edge of the market is closer for each hotel than to the central market point. For example, the Jinjiang Inn was 27 m from the edge of the east section of the market, but 98 m from the center point. To be conservative, we used distances from each hotel to the market's center point.

Table S11. Acknowledgement of GISAID and originating laboratories

Virus Name	Accession No.	Collection Date	Originating Laboratory	Submitting Laboratory	Authors
hCoV-19/Wuhan/IPBCAMS-WH-01/2019	EPI_ISL_402123	2019-12-24	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu
hCoV-19/Wuhan/Hu-1/2019	EPI_ISL_402125	2019-12-31	National Institute for Communicable Disease Control and Prevention (ICDC) Chinese Center for Disease Control and Prevention (China CDC)	National Institute for Communicable Disease Control and Prevention (ICDC) Chinese Center for Disease Control and Prevention (China CDC)	Zhang, Y.-Z., Wu, F., Chen, Y.-M., Pei, Y.-Y., Xu, L., Wang, W., Zhao, S., Yu, B., Hu, Y., Tao, Z.-W., Song, Z.-G., Tian, J.-H., Zhang, Y.-L., Liu, Y., Zheng, J.-J., Dai, F.-H., Wang, Q.-M., She, J.-L. and Zhu, T.-Y.
hCoV-19/Wuhan/IPBCAMS-WH-05/2020	EPI_ISL_403928	2020-01-01	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu
hCoV-19/Wuhan/IPBCAMS-WH-04/2019	EPI_ISL_403929	2019-12-30	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu
hCoV-19/Wuhan/IPBCAMS-WH-03/2019	EPI_ISL_403930	2019-12-30	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu
hCoV-19/Wuhan/IPBCAMS-WH-02/2019	EPI_ISL_403931	2019-12-30	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College	Lili Ren, Jianwei Wang, Qi Jin, Zichun Xiang, Zhiqiang Wu, Chao Wu, Yiwei Liu
hCoV-19/Wuhan/WHU01/2020	EPI_ISL_406716	2020-01-02	State Key Laboratory of Virology, Wuhan University	State Key Laboratory of Virology, Wuhan University	Chen, L., Liu, W., Zhang, Q., Xu, K., Ye, G., Wu, W., Sun, Z., Liu, F., Wu, K., Mei, Y., Zhang, W., Chen, Y., Li, Y., Shi, M., Lan, K. and Liu, Y.
hCoV-19/Wuhan/WHU02/2020	EPI_ISL_406717	2020-01-02	State Key Laboratory of Virology, Wuhan University	State Key Laboratory of Virology, Wuhan University	Chen, L., Liu, W., Zhang, Q., Xu, K., Ye, G., Wu, W., Sun, Z., Liu, F., Wu, K., Mei, Y., Zhang, W., Chen, Y., Li, Y., Shi, M., Lan, K. and Liu, Y.
hCoV-19/Wuhan/WH01/2019	EPI_ISL_406798	2019-12-26	General Hospital of Central Theater Command of People's Liberation Army of China	BGI & Institute of Microbiology, Chinese Academy of Sciences & Shandong First Medical University & Shandong Academy of Medical Sciences & General Hospital of Central Theater Command of People's Liberation Army of China	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu
hCoV-19/Wuhan/HBCDC-HB-02/2019	EPI_ISL_412898	2019-12-30	Wuhan Jinyintan Hospital	Hubei Provincial Center for Disease Control and Prevention	Bin Fang, Xiang Li, Xiao Yu, Linlin Liu, Bo Yang, Faxian Zhan, Guojun Ye, Xixiang Huo, Junqiang Xu, Bo Yu, Kun Cai, Jing Li, Yongzhong Jiang.

hCoV-19/Wuhan/ IME-WH01/2019	EPI_ISL_529213	2019-12-30	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Fan, Hang; Qin, E.; Wu, Y.; Guo, Y.; Zhang, X.; Yong, Y.; Hou, J.; Xu, Z.; Mu, J.; Teng, Yue; Mi, Z.; Yang, R.; Song, Yajun.; Li, B.; Cui, Y.
hCoV-19/Wuhan/ IME-WH03/2019	EPI_ISL_529215	2019-12-30	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Fan, Hang; Qin, E.; Wu, Y.; Guo, Y.; Zhang, X.; Yong, Y.; Hou, J.; Xu, Z.; Mu, J.; Teng, Yue; Mi, Z.; Yang, R.; Song, Yajun.; Li, B.; Cui, Y.
hCoV-19/Wuhan/ IME-WH04/2019	EPI_ISL_529216	2019-12-30	Beijing Institute of Microbiology and Epidemiology	Beijing Institute of Microbiology and Epidemiology	Fan, Hang; Qin, E.; Wu, Y.; Guo, Y.; Zhang, X.; Yong, Y.; Hou, J.; Xu, Z.; Mu, J.; Teng, Yue; Mi, Z.; Yang, R.; Song, Yajun.; Li, B.; Cui, Y.
hCoV-19/Wuhan/ WH04/2020	EPI_ISL_406801	2020-01-05	General Hospital of Central Theater Command of People's Liberation Army of China	BGI & Institute of Microbiology, Chinese Academy of Sciences & Shandong First Medical University & Shandong Academy of Medical Sciences & General Hospital of Central Theater Command of People's Liberation Army of China	Weijun Chen, Yuhai Bi, Weifeng Shi and Zhenhong Hu

Supplementary References

1. WHO, WHO-convened global study of origins of SARS-CoV-2: China Part (2021), (available at <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>).
2. M. Bondarenko, D. Kerr, A. Sorichetta, A. Tatem, Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs (2020), (available at <https://www.worldpop.org/doi/10.5258/SOTON/WP00684>).
3. Z. Peng, R. Wang, L. Liu, H. Wu, Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data. *ISPRS International Journal of Geo-Information*. **9**, 402 (2020/6).
4. E. C. Holmes, S. A. Goldstein, A. L. Rasmussen, D. L. Robertson, A. Crits-Christoph, J. O. Wertheim, S. J. Anthony, W. S. Barclay, M. F. Boni, P. C. Doherty, J. Farrar, J. L. Geoghegan, X. Jiang, J. L. Leibowitz, S. J. D. Neil, T. Skern, S. R. Weiss, M. Worobey, K. G. Andersen, R. F. Garry, A. Rambaut, The origins of SARS-CoV-2: A critical review. *Cell*. **184**, 4848–4856 (2021).
5. M. L. Hazelton, T. M. Davies, Inference based on kernel estimates of the relative risk function in geographical epidemiology. *Biom. J.* **51**, 98–109 (2009).
6. A. Rambaut, E. C. Holmes, Á. O’Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, O. G. Pybus, A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* **5**, 1403–1407 (2020).
7. M. Worobey, Dissecting the early COVID-19 cases in Wuhan. *Science*. **0**, eabm4454 (2021).
8. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. **395**, 565–574 (2020).
9. W. Guizhen, “中疾控病疾控报 | Chinese CDC Disease Control Report” (53, Chinese CDC, 2020).
10. T. M. Davies, J. C. Marshall, M. L. Hazelton, Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. *Stat. Med.* **37**, 1191–1221 (2018).
11. Wuhan Municipal Bureau of Landscape Architecture and Forestry, 武汉市园林和林业局 | Administrative Penalties in 2019 (2019), (available at https://web.archive.org/web/20211117124950/http://ylj.wuhan.gov.cn/zwgk/zwxxgkzl_12298/cfqz/xzcf/202011/t20201110_1499879.shtml).
12. E. C. Holmes, A. Rambaut, K. G. Andersen, Pandemics: spend on surveillance, not prediction. *Nature*. **558**, 180–182 (2018).
13. K. L. Stout, “Wuhan SARS”: Tracing the origin of the new virus to China’s wild animal markets (2020), (available at https://www.youtube.com/watch?v=Je0_U2ym_r0).
14. Y. Jia, Z. Zheng, Q. Zhang, M. Li, X. Liu, Associations of Spatial Aggregation between Neighborhood Facilities and the Population of Age Groups Based on Points-of-Interest Data. *Sustain. Sci. Pract. Policy*. **12**, 1692 (2020).
15. F. Maussion, TimoRoth, R. Bell, F. Li, J. Landmann, M. Dusch, tbridel, *fmaussion/salem: v0.3.7* (Zenodo, 2021; <https://zenodo.org/record/596573>).
16. C. M. Freuling, A. Breithaupt, T. Müller, J. Sehl, A. Balkema-Buschmann, M. Rissmann, A. Klein, C. Wylezich, D. Höper, K. Wernike, A. Aebischer, D. Hoffmann, V. Friedrichs, A. Dorhoi, M. H. Groschup, M. Beer, T. C. Mettenleiter, Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerg. Infect. Dis.* **26**, 2982–2985 (2020).

17. K. Wernike, A. Aebischer, A. Michelitsch, D. Hoffmann, C. Freuling, A. Balkema-Buschmann, A. Graaf, T. Müller, N. Osterrieder, M. Rissmann, D. Rubbenstroth, J. Schön, C. Schulz, J. Trimpert, L. Ulrich, A. Volz, T. Mettenleiter, M. Beer, Multi-species ELISA for the detection of antibodies against SARS-CoV-2 in animals. *Transbound. Emerg. Dis.* **68**, 1779–1785 (2021).
18. X. Zhao, D. Chen, R. Szabla, M. Zheng, G. Li, P. Du, S. Zheng, X. Li, C. Song, R. Li, J.-T. Guo, M. Junop, H. Zeng, H. Lin, Broad and Differential Animal Angiotensin-Converting Enzyme 2 Receptor Usage by SARS-CoV-2. *J. Virol.* **94** (2020), doi:10.1128/JVI.00940-20.
19. A. Z. Mykytyn, M. M. Lamers, N. M. A. Okba, T. I. Breugem, D. Schipper, P. B. van den Doel, P. van Run, G. van Amerongen, L. de Waal, M. P. G. Koopmans, K. J. Stittelaar, J. M. A. van den Brand, B. L. Haagmans, Susceptibility of rabbits to SARS-CoV-2. *Emerg. Microbes Infect.* **10**, 1–7 (2021).
20. P. Chen, J. Wang, X. Xu, Y. Li, Y. Zhu, X. Li, M. Li, P. Hao, Molecular dynamic simulation analysis of SARS-CoV-2 spike mutations and evaluation of ACE2 from pets and wild animals for infection risk. *Comput. Biol. Chem.* **96**, 107613 (2022).
21. V. L. Hale, P. M. Dennis, D. S. McBride, J. M. Nolting, C. Madden, D. Huey, M. Ehrlich, J. Grieser, J. Winston, D. Lombardi, S. Gibson, L. Saif, M. L. Killian, K. Lantz, R. M. Tell, M. Torchetti, S. Robbe-Austerman, M. I. Nelson, S. A. Faith, A. S. Bowman, SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature.* **602**, 481–486 (2022).
22. S. M. Porter, A. E. Hartwig, H. Bielefeldt-Ohmann, A. M. Bosco-Lauth, J. Jeffrey Root, Susceptibility of wild canids to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *bioRxiv* (2022), p. 2022.01.27.478082.
23. L. Jemeršić, I. Lojkić, N. Krešić, T. Keros, T. A. Zelenika, L. Jurinović, D. Skok, I. Bata, J. Boras, B. Habrun, D. Brnić, Investigating the Presence of SARS CoV-2 in Free-Living and Captive Animals. *Pathogens.* **10** (2021), doi:10.3390/pathogens10060635.
24. C. S. Lupala, V. Kumar, X.-D. Su, C. Wu, H. Liu, Computational insights into differential interaction of mammalian angiotensin-converting enzyme 2 with the SARS-CoV-2 spike receptor binding domain. *Comput. Biol. Med.* **141**, 105017 (2022).
25. C. D. Eckstrand, T. J. Baldwin, K. A. Rood, M. J. Clayton, J. K. Lott, R. M. Wolking, D. S. Bradway, T. Baszler, An outbreak of SARS-CoV-2 with high mortality in mink (*Neovison vison*) on multiple Utah farms. *PLoS Pathog.* **17**, e1009952 (2021).
26. L. Lu, R. S. Sikkema, F. C. Velkers, D. F. Nieuwenhuijse, E. A. J. Fischer, P. A. Meijer, N. Bouwmeester-Vincken, A. Rietveld, M. C. A. Wegdam-Blans, P. Tolsma, M. Koppelman, L. A. M. Smit, R. W. Hakze-van der Honing, W. H. M. van der Poel, A. N. van der Spek, M. A. H. Spierenburg, R. J. Molenaar, J. de Rond, M. Augustijn, M. Woolhouse, J. A. Stegeman, S. Lycett, B. B. Oude Munnink, M. P. G. Koopmans, Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat. Commun.* **12**, 6802 (2021).
27. N. Oreshkova, R. J. Molenaar, S. Vreman, F. Harders, B. B. Oude Munnink, R. W. Hakze-van der Honing, N. Gerhards, P. Tolsma, R. Bouwstra, R. S. Sikkema, M. G. Tacken, M. M. de Rooij, E. Weesendorp, M. Y. Engelsma, C. J. Brusckke, L. A. Smit, M. Koopmans, W. H. van der Poel, A. Stegeman, SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill.* **25** (2020), doi:10.2807/1560-7917.ES.2020.25.23.2001005.
28. A. S. Hammer, M. L. Quaade, T. B. Rasmussen, J. Fonager, M. Rasmussen, K. Mundbjerg, L. Lohse, B. Strandbygaard, C. S. Jørgensen, A. Alfaro-Núñez, M. W. Rosenstjerne, A. Boklund, T. Halasa, A. Fomsgaard, G. J. Belsham, A. Bøtner, SARS-CoV-2 Transmission between Mink (*Neovison vison*) and Humans, Denmark. *Emerg. Infect. Dis.* **27**, 547–551 (2021).
29. Z. Song, L. Bao, W. Deng, J. Liu, E. Ren, Q. Lv, M. Liu, F. Qi, T. Chen, R. Deng, F. Li, Y. Liu, Q. Wei, H. Gao, P. Yu, Y. Han, W. Zhao, J. Zheng, X. Liang, F. Yang, C. Qin, Integrated histopathological, lipidomic, and metabolomic profiles reveal mink is a useful animal model to mimic the pathogenicity of severe COVID-19 patients. *Signal Transduct Target Ther.* **7**, 29 (2022).

30. H.-L. Zhang, Y.-M. Li, J. Sun, Y.-Y. Zhang, T.-Y. Wang, M.-X. Sun, M.-H. Wang, Y.-L. Yang, X.-L. Hu, Y.-D. Tang, J. Zhao, X. Cai, Evaluating angiotensin-converting enzyme 2-mediated SARS-CoV-2 entry across species. *J. Biol. Chem.* **296**, 100435 (2021).