nature

https://doi.org/10.1038/s41586-023-06043-2

Accelerated Article Preview

Surveillance of SARS-CoV-2 at the Huanan Seafood Market

Received: 17 February 2022

Accepted: 3 April 2023

Accelerated Article Preview

Cite this article as: Liu, W. J. et al. Surveillance of SARS-CoV-2 at the Huanan Seafood Market. *Nature* https://doi.org/10.1038/s41586-023-06043-2 (2023)

William J. Liu, Peipei Liu, Wenwen Lei, Zhiyuan Jia, Xiaozhou He, Weifeng Shi, Yun Tan, Shumei Zou, Gary Wong, Ji Wang, Feng Wang, Gang Wang, Kun Qin, Rongbao Gao, Jie Zhang, Min Li, Wenling Xiao, Yuanyuan Guo, Ziqian Xu, Yingze Zhao, Jingdong Song, Jing Zhang, Wei Zhen, Wenting Zhou, Beiwei Ye, Juan Song, Mengjie Yang, Weimin Zhou, Yuting Dai, Gang Lu, Yuhai Bi, Wenjie Tan, Jun Han, George F. Gao & Guizhen Wu

Surveillance of SARS-CoV-2 at the Huanan Seafood Market

William J. Liu,[1]# Peipei Liu,[1]# Wenwen Lei,[1]# Zhiyuan Jia,[1]# Xiaozhou He,[1]# Weifeng Shi,[2]# Yun Tan,[3]# Shumei Zou,[1] Gary Wong,[4] Ji Wang,[1] Feng Wang,[1] Gang Wang,[1] Kun Qin,[1] Rongbao Gao,[1] Jie Zhang,[1] Min Li,[1] Wenling Xiao,[1,5] Yuanyuan Guo,[1] Ziqian Xu,[1] Yingze Zhao,[1] Jingdong Song,[1] Jing Zhang,[1] Wei Zhen,[1] Wenting Zhou,[1] Beiwei Ye,[1] Juan Song,[1] Mengjie Yang,[1] Weimin Zhou,[1] Yuting Dai,[3] Gang Lu,[3] Yuhai Bi,[6] Wenjie Tan[1], Jun Han[1], George F. Gao,[1,6] Guizhen Wu[1]

[1] NHC Key Laboratory of Biosafety, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing 102206, China;

[2] Key Laboratory of Emerging Infectious Diseases in Universities of Shandong, Shandong First Medical University, and Shandong Academy of Medical Sciences, Tai'an 271000, China;

[3] Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine, Ruijin Hospital Affiliated to Shanghai Jiao Tong University (SJTU) School of Medicine, Shanghai 200020, China;

[4] CAS Key Laboratory of Molecular Virology & Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences (CAS), Shanghai 200031, China;

[5] School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou 325035, China;

[6] CAS Key Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences (CAS), Beijing 100101, China.

# Contributed equally.

Correspondence:
Guizhen Wu: wugz@ivdc.chinacdc.cn, ORCID 0000-0003-2778-4290
George F. Gao: gaofu@chinacdc.cn, ORCID 0000-0002-3869-615X
William J. Liu: liujun@ivdc.chinacdc.cn, ORCID 0000-0003-3605-4070

Abstract SARS-CoV-2, the causative agent of COVID-19, emerged in December 2019. Its origins remain uncertain. It has been reported that a number of the early human cases had a history of contact with the Huanan Seafood Market. Here we present the results of surveillance for SARS-CoV-2 within the market. From January 1st 2020, after closure of the market, 923 samples were collected from the environment. From 18th January, 457 samples were collected from 18 species of animals, comprising of unsold contents of refrigerators and freezers, swabs from stray animals, and the contents of a fish tank. Using RT-qPCR, SARS-CoV-2 was detected in 73 environmental samples, but none of the animal samples. Three live viruses were successfully isolated. The viruses from the market shared nucleotide identity of 99.99% to 100% with the human isolate HCoV-19/Wuhan/IVDC-HB-01/2019. SARS-CoV-2 lineage A (8782T and 28144C) was found in an environmental sample. RNA-seq analysis of SARS-CoV-2 positive and negative environmental samples showed an abundance of different vertebrate genera at the market. In summary, this study provides information about the distribution and prevalence of SARS-CoV-2 in the Huanan Seafood Market during the early stages of the COVID-19 outbreak.

Keywords:
COVID-19, SARS-CoV-2, Huanan Seafood Market, origin, high-throughput sequencing, virus isolation, sewage

Infections with novel human coronavirus 2019 (HCoV-19) [1,2], named as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (ICTV) [3], can result in coronavirus disease 2019 (COVID-19), characterized by various clinical outcomes from asymptomatic infections to severe pneumonia and even death [4,5]. Globally, as of Feb 28th 2023, over 758 million confirmed cases and over 6.8 million deaths have been reported (covid19.who.int).

Human cases with COVID-19 were first reported in late December 2019, in Wuhan, China, as pneumonia of unknown etiology (PUE). A majority of these early cases were found to be linked to the Huanan Seafood Market (HSM) in Wuhan [4,6], where various animal meats, exotic seafood and live animals were available for purchase. The HSM has been suspected to be the source of the COVID-19 pandemic [7]. Not all of the early human cases had epidemiological links to the market[6,8] and alternative hypotheses for the market association, for example entry of virus into the market via humans or the cold-chain, also exist.

SARS-CoV-2 has high similarity with a few coronaviruses derived from bats in Asian countries including China, Laos, Japan, Cambodia and Thailand, and some scientists have proposed that bats might be the original source of SARS-CoV-2 [1,8-14]. Whether another animal might have acted as an intermediate host to facilitate virus spillover from bats to humans is still unknown [15,16]. An important finding was the discovery of SARS-CoV-2-related coronaviruses from pangolins, in which the spike proteins contained receptor-binding domains (RBD) showing high similarity to the RBD of SARS-CoV-2 [17-19]. Pangolins might be involved in the ecology of coronaviruses, but whether they are the intermediate host for SARS-CoV-2 is unknown, given the current data[20]. A recent study documented the animal species in the HSM between May 2017 and November 2019 and noted that no pangolins or bats were present, but some hypothesized sarbecovirus-susceptible animals, such as raccoon dogs were present [21]. Thus far, the origins of SARS-CoV-2 [22,23] and the role of the HSM in the origins and

spread of SARS-CoV-2 remain unclear. The data from the HSM may provide important information.

The HSM is located in the Jianghan District in the downtown area of Wuhan, the capital city of Hubei Province, and is approximately 800 m away from Hankou Railway Station, a major railway travel hub. It occupies >50,000 m^2, with 678 stalls located close to each other in extremely crowded conditions (Fig. 1A). The market is separated into two zones, the East and West Zones, with seafood and animals mainly sold in the West Zone and livestock meat in the East Zone. Among the 678 stalls of the market, 10 stalls selling domesticated wildlife (1.5%) were identified according to sale records [24], located in the south-western corner of West Zone (8/10) and the north-western corner of East Zone (2/10), respectively (Fig. 1A). According to sale records, during late December 2019, animals or animal products were sold in these 10 animal stalls. Animals included snakes, avian species (chickens, ducks, gooses, pheasants and doves), Sika deer, badgers, rabbits, bamboo rats, porcupines, hedgehogs, salamanders, giant salamanders, bay crocodiles and Siamese crocodiles, etc., among which snakes, salamanders and crocodiles were traded as live animals (described in detail in the Report of WHO-convened global study of origins of SARS-CoV-2 [24]).

The market was closed in the morning of January 1st, 2020, shortly after the identification of the PUE. On the same day, in the early morning, the Chinese Center for Disease Control and Prevention (China CDC) dispatched an epidemiological team, together with experts from Hubei Provincial CDC and Wuhan Municipal CDC, to the HSM to collect environmental samples in order to investigate the potential introduction of SARS-CoV-2 to the market (Fig. 1B). From January 1st 2020 until March 2nd 2020, a total of 923 environmental samples from different locations within and around the market and 457 animal samples, including dead animals in refrigerators and freezers and stray animals and their feces, were collected, with some stray animals sampled until March 30th (Extended Data Tables 1, 2, 3 and Supplementary Table 1). After the closure

of the market, the outside surface of the rolling shutter doors of the stalls and the corridors were disinfected (with 1% bleach mixed with water) throughout January and February 2020. The goods inside the stalls were completely cleared and disinfected until early March 2020.

Out of the 923 environmental samples collected in and around the HSM, 73 were found by the real-time polymerase chain reactions (RT-PCR) to be positive for SARS-CoV-2 with positivity rate of 7.9%. Cycle threshold (CT) values for the RT-PCR ranged from 23.9 to 41.7 (Supplementary Table 2). Among the 828 samples inside the HSM, 64 samples (7.7%) were positive. Of the 64 SARS-CoV-2 positive samples collected inside the HSM, 87.5% (56/64) were collected in the West Zone of the market, in particular streets from no. 1 to 8, with 71.4% (40/56) positive samples identified herein (Fig. 1A). Among the 14 samples from warehouses related to the HSM, five tested positive. This may reflect the nature of SARS-CoV-2 presence in the market during the early phase of the outbreak. Among the 51 samples from sewerage wells (Supplementary Table 1) in the surrounding areas outside the HSM, three tested positive (Supplementary Table 2). Notably, one sample (Env_0601), a floor surface swab, out of the 30 environmental samples collected from Dongxihu Market in Wuhan on January 22nd 2020, also tested positive (Supplementary Table 2, Extended Data Table 4).

Of the 110 samples collected from sewers or sewerage wells in the market, 24 samples were positive for SARS-CoV-2 nucleic acid. All four sewerage wells in the market tested positive. During the onsite investigation of the overground drainage pathway in the HSM, we found that the wastewater in the overground drainage led into the underground drainage inside the market and then flowed into the wells on the edge of the market. We then did a spot-check sampling across all the overground drains according to the principles described in the Methods (Extended Data Fig. 1). Excreta of the upper respiratory tract of infected humans and the potential animal waste would be mixed together into the overground drainage. Thus, these data suggested that either

infected people and/or animals in the market contaminated the sewage or that the contaminated sewage may have further played a role in furthering the virus transmission within the case cluster in the market.

The merchants' activities were assessed against the PCR results of the environmental samples. The sampling covered 19.8% (134/678) of the shops in the market (95% confidence interval (CI): 16.8-23.0%). Of the positive samples, 44 were distributed among 21 shops in the market, 19 of whom were located in the West Zone with the remaining two located in the east area (Fig. 1A). Some vendors sold more than one type of product. While the results provided some indication of the association of cases with different products, no significant differences were observed between different shops, including those selling poultry (22%, 8/37: 95% CI: 9.8-38.2%), cold-chain products (18.4%, 16/87, 95% CI: 10.9-28.1%), aquatic products (17.8%, 13/73, 95% CI: 9.8-28.5%), livestock (14%, 5/36: 95% CI: 4.7-29.5%), seafood products (11%, 6/56: 95% CI: 4-21.9%), wildlife products (11%, 1/9: 95% CI: 0.3-48.2%), and vegetables (25%, 2/8: 95% CI: 3.2-65%) (Extended Data Fig. 2, Extended Data Table 5). The detection of SARS-CoV-2 in multiple shops selling different product types suggested that SARS-CoV-2 may have been circulating in the market, especially the West Zone, for a while in December 2019, leading to an extensive distribution of the virus within the market, which may have been facilitated by the crowded buyers and the contaminated environment.

The 457 animal samples included 188 individuals belonging to 18 species (with some stray animals sampled until March 30th) (Extended Data Table 6). The sources of the samples included unsold goods kept in refrigerators and freezers in the stalls of the HSM, and goods kept in warehouses and refrigerators related to the HSM. Three Chinese giant salamanders, which were found in a fish tank, were alive and swab samples were collected and tested. Samples from stray animals in the market were also collected, comprising swab samples from 10 stray cats, 27 samples of cat feces, one

dog, one weasel, and 10 rats. All the 457 animal samples tested negative for SARS-CoV-2 nucleic acid.

To determine whether there was live virus in the HSM, we inoculated 27 SARS-CoV-2 positive environmental samples collected on January 1st, 2020, into cell lines, including Vero E6 and Huh7.5 cells. Cytopathic effects (CPE) were observed 3 days post inoculation with sample Env_0313 on Vero E6 cells. CPE was also observed 5 days post inoculation on Huh7.5 cells. The electron micrographs of Vero E6 cells after 5 days of post inoculation showed that virus particles were present in both the supernatant and the cells. Negative-stained virus particles and ultra-thin cultured cell sections showed typical coronavirus morphology (Fig. 2). Live viruses were isolated from samples Env_0313, Env_0354 and Env_0126, which were the only three samples with CT values <30 in the PCR. Env_0354 and Env_0126 were swab samples from the ground and Env_0313 were swab samples from a wall. Notably, samples Env_0313 and Env_0354 were from the stalls with confirmed patients. All the results of successful virus isolation from the original samples with low CT values revealed the existence of live SARS-CoV-2 with high titers in the environment of the HSM. Do the high CT values, we did not perform virus isolation based on the samples collected from later time points due.

During later sampling in the HSM in February, we collected samples to investigate the virus RNA persistence in the market. Some of these samples tested positive, especially in the sewage well and even on the walls (Supplementary Table 2). Within the 73 PCR positive samples, 35 samples (27 within the HSM and 8 from the surrounding area) collected in February were still positive for SARS-CoV-2. The long persistence of its genetic material in the environment might reflect high levels of environmental contamination before the market was closed. For the sample Env_0838, collected from a wall on February 20th 2020, a 3-plex PCR test was performed. The viral RNA segment was undetectable in one PCR channel targeting N gene, but could be amplified in the

other two channels targeting the RdRp and E genes, with CT values of 32.59 and 37.34, respectively. This result is reasonable considering the degradation of the viral genome. However, the results also indicate a long persistence of the viral RNA in the environment.

We further performed high-throughput sequencing (Supplementary Table 3) and successfully obtained seven complete or near complete SARS-CoV-2 genome sequences, including three sequences from three environmental samples (Env_0313, Env_0354 and Env_0020), and four sequences from cell supernatants of Env_0313, Env_0354 and Env_0126 (Fig. 3, Supplementary Table 4). A few samples were re-sequenced using a multiplex PCR approach, including Env_0020_seq01, Env_0313_seq04, Env_0313_seq05, Env_0126_seq06, and Env_0354_seq07 (Supplementary Table 3 and 4). The genome sequences of three environmental samples, Env_0126, Env_0313 and Env_0354, were found to be completely identical to the reference strain HCoV-19/Wuhan/IVDC-HB-01/2019 (IVDC-HB-01, GISAID accession number: EPI_ISL_402119) and the human strain Wuhan-Hu-1 (GenBank: NC_045512) (Fig. 3A). The genome sequence of the isolated virus from environmental sample Env_0354 had two synonymous mutations compared to HCoV-19/Wuhan/IVDC-HB-01/2019, with sequence identity of 99.99% (Fig. 3A). Therefore, the SARS-CoV-2 sequences from environmental samples were highly similar to the clinical strains obtained during the early stages of the COVID-19 outbreak.

Previously, SARS-CoV-2 has been proposed to be classified into two major lineages based on the two highly-linked single nucleotide polymorphisms (SNPs): A lineage (8782T and 28144C, or S lineage in another nomenclature of SARS-CoV-2) and B lineage (8782C and 28144T, or L lineage). It has been proposed that A/S lineage most likely is the ancestral lineage, because all of the SARS-CoV-2 related coronaviruses from bats and pangolins possessed 8782T and 28144C [25,26], while Pekar et al. also presented a possibility that both lineages represent separate introduction events[27].

Phylogenetic analysis revealed that most of the environmental strains belong to the B/L lineage and they cluster together with the human strains circulating in the early stage of the pandemic (Fig. 3B, Supplementary Fig. 1). The phylogenetic analysis did not involve the environmental sample Env_0020, the A/S lineage of which was confirmed by the high number of reads mapped to positions 8782 and 28144 in Env_0020 (Supplementary Table 5). However, it should be noted that the genome of Env_0020 is of low quality and there are many discontinuous gaps in the assembled genome. Indeed, though it is difficult to root the SARS-CoV-2 phylogenetic tree, our analysis indicated that the environmental viruses clustered together with the human strains circulating in the early stages of the pandemic.

We conducted RNA-seq analysis using 60 SARS-CoV-2 PCR-positive and 112 SARS-CoV-2 PCR-negative environmental samples from the HSM (Fig. 4A and Supplementary Table 3), in which the bias of sampling and RNA-seq should be considered. We used two approaches for genera identification. The Kraken2 method with all available genes/genomes in the database was used for the identification of all genera, including bacteria, viruses, eukaryota, and archaea. Additionally, the barcoding method using mitochondrial cytochrome c oxidase subunit I (COI) sequences was used specifically for the identification of Chordata genera. Bacteria were the most abundant species in almost all samples and mammal species could be found in most samples, which fit the feature of samples collected from the environment (Fig. 4B and Supplementary Table 6 and 7). *Gallus*, *Homo*, *Anas*, *Sus*, *Bos*, and *Canis* could be detected in most samples (Fig. 4C and Supplementary Table 8), which was in accordance with the environmental feature of the seafood markets in China. We analyzed the mammalian genera in all sequenced samples with kranken2 (detailed in the methods) using different thresholds. A total of 70 mammal genera, which existed in more than 2% samples, were identified with a threshold of 100 reads per million (Fig. 4D). It is important to highlight that the results of the kraken2 analysis (Figure 4D) and the BOLD analysis (Extended Data Figure 3) differ. In particular, the proportion of

reads assigned as raccoon dog differ considerably with the two methods used. This may be due to the heterogeneity of the reference data used by the two methods (BOLD, as for mitochondria, and kraken2 for whole genome). It should be noted that the genera identified using current approaches might be updated with additional reference genomes. As such, this list is not definitive and further in-depth analysis with other methods will be required to provide more information regarding the wildlife species present at the market.

Particularly, we analyzed three samples (Env_0126, Env_0313 and Env_0354) collected on 1st Jan 2020 with high levels of SARS-CoV-2 (Ct value <30) (Fig. 4E). The identified mammal genera in the Env_0313 and Env_0354 samples were related to species in the general food market, such as *Homo*, *Ovis*, *Bos*, *Canis*, *Sus*, and *Felis*. Many mammalian genera were observed in the Env_0126 sample, but the most abundant mammalian genera were also related to the general food market, including *Bos* (77.30%), *Ovis* (19.91%), *Homo* (0.77%), and *Bubalus* (0.57%). *Pipistrellus* (0.002%) and *Lutra* (0.001%) were found also found in this sample, but at extremely low relative abundance, raising the possibility of false detection. Moreover, we also noted that only *Homo*, *Ovis*, *Bos*, and *Sus* reads but not species related to wildlife were found in the Env_0020 samples, the one that belongs to the A/S lineage.

We illustrated the top-ranked genera in four areas of the market, where multiple SARS-CoV-2 PCR-positive samples were detected. As shown in Fig. 4F, the top-ranked genera in these areas were *homo* or other genera that generally exist in food markets. We also noted that *Nyctereutes* could be found in the shop 25 of street 8, while *Atelerix* and *Erinaceus* could be found in shops 15-17 of street 7 (Fig. 4F). These genera were detected in both SARS-CoV-2 positive and SARS-CoV-2 negative samples, and actually more often so in negative ones (Supplementary Table 6-9), and furthermore, this does not allow conclusions about whether these animals were infected with SARS-CoV-2.

We checked samples that might relate to wildlife, such as samples collected in the defeathering machine and areas with the visible blood spots. The most abundant mammal genera of the defeathering machine sample (Env_0584) was *Canis* (Extended data Fig 3). The most abundant mammal species of the visible blood spot sample (Env_0262) were *Bos*, *Sus*, *Ovis* and *Bison*, accordingly (Extended data Fig 3). Additionally, we plotted the distribution of some genera of concern, including *Myotis*, *Erinaceus*, *Mustela*, *Nyctereutes*, *Rhizomys*, *Meles*, and *Melogale*. Most of these samples were distributed in the western district of the market (Extended data Fig 4), where wildlife products were sold, though this also reflects the zone much more intensively sampled and analyzed by RNA-seq. The distribution locations of *Homo*, *Sus*, *Bos*, *Gallus* and *Anas* were also dominant in this area, where the enriched areas of SARS-CoV-2 PCR-positive samples were nearby. The repeated sampling of the locations with PCR-positive results may contribute some bias to the distribution analyses of enriched areas of SARS-CoV-2 PCR-positive samples. Additionally, We plotted the proportions of mammal genera in those SARS-CoV-2 positive samples with high abundance of genera related to wildlife, such as Env_0576 (*Nyctereutes* enriched), Env_0807 (*Lariscus* enriched), Env_0809 (*Erinaceus* enriched), and Env_0585 (*Erinaceus* enriched) (Extended Data Fig. 3).

Of particular note was the difference in the results from PCR and NGS. Among the 60 SARS-CoV-2 PCR-positive samples for RNA-seq analysis, 39 samples tested negative by NGS (no SARS-CoV-2 reads at all) (65.0%), including sample Env_0262. For these NGS-negative samples, the CT values ranged from 31.80 to 37.44. Since the RT-PCR detection assay employed in the very early stage of the pandemic was not formally verified, we believe that there may be some false positives in the PCR detection results in this study. Meanwhile, we also found that SARS-CoV-2 reads could also be detected by NGS in a portion of SARS-CoV-2 PCR negative samples (15.2%), which might be caused by the degradation of SARS-CoV-2 within the PCR target region or

contamination during library building.

In summary, we report the detection of SARS-CoV-2 RNA and live virus in environmental samples from the West Zone of the HSM. We should note that the selection of shops for sampling was biased because shops selling wildlife as well as shops linked to early cases were prioritized for sampling. The origin of the virus cannot be determined from all the analyses available so far. Although gene barcode analysis of animal species in the study suggested that *Myotis*, *Nyctereutes* and *Melogale* – species that have been recognized as potential host species of sarbecoviruses – were present at the market, these barcodes were mostly detected within the SARS-CoV-2 PCR negative environment samples. It remains possible that the market may acted as an amplifier of transmission due to the high number of visitors every day, causing many of the initially identified infection clusters in the early stages of the outbreak [24].

Recent reports traced the outbreak back to the HSM and proposed, after compiling information reported by various sources, including the WHO-China Joint Report and social media, etc. that the market sold live wild animals as recently as 2019 [28]. Another report hypothesized that SARS-CoV-2 spilled over from animals to humans at least twice in November or December 2019, and the raccoon dog was hypothesized to be the intermediate host animal [27]. The evidence provided in this study is not sufficient to support such a hypothesis [29]. Our study confirmed the existence of raccoon dogs, and other hypothesized/potential SARS-CoV-2 susceptible animals, at the market, prior to its closure. However, these environmental samples cannot prove that the animals were infected. Furthermore, even if the animals were infected, our study does not rule out that human-to-animal transmission occurred, considering the sampling time was after the human infection within the market as reported retrospectively[6]. Thus, the possibility of potential introduction of the virus to the market through infected humans, or cold chain products, cannot be ruled out yet.

More work, involving internationally coordinated efforts, is needed to investigate the potential origins of SARS-CoV-2 [24]. Surveillance of wild animals should be enhanced to explore the potential natural and intermediate hosts for SARS-CoV-2 [7,30], if any, which would help to prevent future pandemics caused by animal-origin coronaviruses.

References

1 Tan, W. *et al.* A novel coronavirus genome identified in a cluster of pneumonia cases - Wuhan, China 2019-2020. *China CDC Wkly* **2**, 61-62, doi:10.46234/ccdcw2020.017 (2020).

2 Jiang, S. *et al.* A distinct name is needed for the new coronavirus. *Lancet* **395**, 949, doi:10.1016/S0140-6736(20)30419-0 (2020).

3 Coronaviridae Study Group of the International Committee on Taxonomy of, V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536-544, doi:10.1038/s41564-020-0695-z (2020).

4 Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470-473, doi:10.1016/S0140-6736(20)30185-9 (2020).

5 Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

6 Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* **382**, 1199-1207, doi:10.1056/NEJMoa2001316 (2020).

7 Daszak, P., Olival, K. J. & Li, H. A strategy to prevent future epidemics similar to the 2019-nCoV outbreak. *Biosaf Health* **2**, 6-8, doi:10.1016/j.bsheal.2020.01.003 (2020).

8 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

9 Murakami, S. *et al.* Detection and characterization of bat sarbecovirus phylogenetically related to SARS-CoV-2, Japan. *Emerg Infect Dis* **26**, 3025-3029, doi:10.3201/eid2612.203386 (2020).

10 Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat Commun* **12**, 972, doi:10.1038/s41467-021-21240-1 (2021).

11 Zhou, H. *et al.* A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr Biol* **30**, 2196-2203 e2193, doi:10.1016/j.cub.2020.05.023 (2020).

12 Zhou, H. *et al.* Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* **184**, 4380-4391 e4314, doi:10.1016/j.cell.2021.06.008 (2021).

13 Li, J., Lai, S., Gao, G. F. & Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **600**, 408-418, doi:10.1038/s41586-021-04188-6 (2021).

14 Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330-336, doi:10.1038/s41586-022-04532-4 (2022).

15 Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**,

565-574, doi:10.1016/S0140-6736(20)30251-8 (2020).
16 Wang, J. *et al.* Individual bat viromes reveal the co-infection, spillover and emergence risk of potential zoonotic viruses. *bioRxiv*, doi:10.1101/2022.11.23.517609 (2022).
17 Lam, T. T. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282-285, doi:10.1038/s41586-020-2169-0 (2020).
18 Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286-289, doi:10.1038/s41586-020-2313-x (2020).
19 Niu, S. *et al.* Molecular basis of cross-species ACE2 interactions with SARS-CoV-2-like viruses of pangolin origin. *EMBO J* **40**, e107786, doi:10.15252/embj.2021107786 (2021).
20 He, W. T. *et al.* Virome characterization of game animals in China reveals a spectrum of emerging pathogens. *Cell* **185**, 1117-1129 e1118, doi:10.1016/j.cell.2022.02.014 (2022).
21 Xiao, X., Newman, C., Buesching, C. D., Macdonald, D. W. & Zhou, Z. M. Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Sci Rep* **11**, 11898, doi:10.1038/s41598-021-91470-2 (2021).
22 Wang, Q. *et al.* Tracing the origins of SARS-CoV-2: lessons learned from the past. *Cell Res* **31**, 1139-1141, doi:10.1038/s41422-021-00575-w (2021).
23 Tong, Y. *et al.* The origins of viruses: discovery takes time, international resources, and cooperation. *Lancet* **398**, 1401-1402, doi:10.1016/S0140-6736(21)02180-2 (2021).
24 WHO-convened global study of origins of SARS-CoV-2: China Part. *https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part* (2021).
25 Tang, X. *et al.* On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* **7**, 2 (2020).
26 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).
27 Pekar, J. E. *et al.* The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* **377**, 960-966, doi:10.1126/science.abp8337 (2022).
28 Worobey, M. *et al.* The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* **377**, 951-959, doi:10.1126/science.abp8715 (2022).
29 Maxmen, A. Wuhan market was epicentre of pandemic's start, studies suggest. *Nature* **603**, 15-16, doi:10.1038/d41586-022-00584-8 (2022).
30 Li, H. *et al.* Human-animal interactions and bat coronavirus spillover potential among rural residents in Southern China. *Biosaf Health* **1**, 84-90, doi:10.1016/j.bsheal.2019.10.004 (2019).

Figure legends

Fig. 1. The distribution of the positive environmental samples in the Huanan Seafood Market.

A. As the place of the early cluster of COVID-19 patients, the Huanan Seafood Market is separated into East and West Zones with the Xinhua Road between them. To detect for the presence of SARS-CoV-2 RNA, reverse transcription, quantitative polymerase chain reaction (RT-qPCR) was performed. The locations of the positive samples were marked in the map of the market within orange, while the location of the samples that the live viruses were isolated from were labeled with red. The map also shows locations of stalls where domesticated wildlife products were sold. B. Timeline of environmental and animal samples collected within and around the Huanan Seafood Market. The information of confirmed patients up to December 31st 2019 was referenced from the Report of WHO-convened global study of origins of SARS-CoV-2.

Fig. 2. The SARS-CoV-2 virus isolation from environmental samples of the Huanan Seafood Market.

The electron micrographs of the SARS-CoV-2 viruses isolated from the environmental samples in the Huanan Seafood Market. To determine whether SARS-CoV-2 particles could be visualized from the cell supernatant and lysate, we used transmission electron microscopy (EM) to observe the culture supernatant and ultra-thin section cells based from both VeroE6 and Huh7.5 cells. The electron micrographs showed that virus particles were present in both the supernatant (A, B) and the cells (C, D). Negative-stained virus particles were generally spherical, pleomorphic and 60-140 nm in diameter. Spike protrusions were observed around the particles in a crown (corona) shape (A, B). In ultra-thin cultured cell sections, a group of virus particles can be seen outside the cell (C), and sheets of virus particles can also be observed inside the cells (D). The graphs were the representatives of repeated experiments of electron micrographs.

Fig. 3. Genomic and phylogenetic analyses of SARS-CoV-2 virus genomes from the Huanan Seafood Market.

A. Sequence comparison of the full-length SARS-CoV-2 genomes in the environmental samples. B. Phylogenetic analysis of full-length SARS-CoV-2 genomes from the Huanan Seafood Market and representative strains from the early stage of the COVID-19 pandemic, showing that most environmental strains cluster together with the human strains in the B/L lineage, with Env_0020 in A/S lineage.

Fig. 4. Analysis of environmental samples in the Huanan Seafood Market.

A. Schematic illustration of the experimental design. All 73 SARS-CoV-2 positive samples were included for RNA-seq. A total of 60 RNA-seq libraries were successfully constructed. Additionally, RNA-seq libraries of 112 SARS-CoV-2 negative samples passed library quality control. The kranken2 was used for genus classification. The bowtie2 and sequences in the barcode of life data system was used for the classification of genus in the Mammalia class. B. Heatmap showing the reads distribution of the four domains (Bacteria, Eukaryota, Viruses and Archaea), the *Homo* genus, the *Mammalia* class and the SARS-CoV-2 species. SARS-CoV-2 PCR-positive or -negative were shown in the left panel. C. Positive ratio of illustrated genus in all tested samples. Top ranked genus within the *Mammalia* class were shown. D. Illustration of mammal genera in market using the threshold of 100 reads per millions. The samples were group by SARS-CoV-2 PCR results. The blue bar indicates the positive detected genera. E. Illustration of mammal genera distribution in samples with high viral load. The Env_0020, Env_0313, Env_0354 and Env_0126 were shown.

F. Distribution of the positively detected *Mammal* genera in the market. Samples in four areas where multiple SARS-CoV-2 PCR-positive samples were plotted. The distribution of top mammal genera in each area was shown.

Methods

Sample collection

The Huanan Seafood Market (HSM) was closed in the early morning of January 1st 2020, and at the same time, China CDC began collecting environmental and animal samples. Staff from China CDC entered the market about 30 times before the market's final clean-up on March 2nd 2020, with some stray animals sampled outside the market until March 30th. Environmental samples in the HSM were collected to represent exhaustively as possible, from a wide diversity of surfaces, animals and products (Supplementary Table 2 and Extended Data Table 6) according to different sampling principles, as described in detail in the Joint Report of WHO-convened Global Study of Origins of SARS-CoV-2: China Part [24].

The principles and ranges of in-market sampling covered: (1) environmental samples from stalls related to early cases; (2) environmental samples from doors and floors of all stalls in the blocks where the early cases were located; (3) environmental samples in the East Zone of the market were collected according to blocks; (4) transport carts, trash cans and similar objects; (5) environmental samples from stalls that sold livestock, poultry, farmed wildlife (also called "domesticated wildlife" or "domesticated wildlife products" in this report); (6) samples of sewage and silt from drainage channels and sewerage wells; (7) stray cats, rats and other stray animals in the market; (8) animal products and other commodity samples kept in the cold storages and refrigerators in the market; (9) the market's ventilation and air-conditioning system; and (10) public toilets, public activity rooms and other places where people gathered in the market.

The investigators used full personal protective equipment during the sampling in the market. Commercial products of swabs and virus preservation solution were used for the sampling (Disposable Virus Sampling Tube, V5-S-25, Shen Zhen Zi Jian Biotechnology Co., Ltd., Shenzhen, China). For environmental samples, sampling swabs were applied to smear the floors, walls or surfaces of objects and then preserved them in virus preservation solution.

For animal samples, depending on the type of animal and whether it was alive or frozen, pharyngeal, anal, body surface and body cavity swabs or tissue samples were

collected for nucleic acid testing (NAT). Generally, for alive animal and frozen full bodies, three samples, including pharyngeal, anal and body surface swabs were collected for each animal individuals. And for animal bodies after "bai tiao" disposing (remaining parts of poultry or livestock after removal of hair and viscera), the body cavity swabs were collected.

Drain samples were collected by the use of virus sampling swabs to probe into the silt at the bottom of drainage channels in the market. Wastewater and silt samples were preserved in virus preservation solution. For the sewage well (for the drain water), a container was used to take a silt-water mixture from a location near the bottom of the well, and an appropriate amount of sample was collected by using virus sampling swabs and then preserved in virus preservation solution.

Nucleic acid extraction and SARS-CoV-2 real-time PCR assay

A virus nucleic acid extraction kit (Xi'an Tianlong) was used to extract viral nucleic acid from samples using an automated nucleic acid extraction instrument according to the manufacturer's instructions. Real-time (RT) PCR was performed on extracted nucleic acid samples with a SARS-CoV-2 nucleic acid assay kit. The reagent brands include BioGerm (40/38, cycle number/cut-off value, the same as below), DAAN (45/40) and BGI (40/38).

Virus isolations

Virus isolations were performed in biosafety level (BSL)-3 laboratory in National Institute for Viral Diseases Control and Prevention, China CDC. Samples positive for SARS-CoV-2 were cultured in Vero E6 and Huh7.5 cells on January 11th, 2020. The cell lines were inoculated with positive samples and three blind passages were performed for each sample. The culture supernatant and cell pellet of each passage were harvested for RT PCR. The morphology of viral particles in the cell sections and the supernatant were firstly observed by transmission electron microscope (TEM) on January 22nd, 2020.

Metagenomic sequencing

Metagenomic sequencing was conducted at National Institute for Viral Disease Control

and Prevention, China CDC and Wuhan BGI. Nucleic acid was extracted using Qiagen's viral RNA microextraction kit and human nucleic acid was removed using an enrichment kit to improve the sensitivity of viral RNA detection. Extracted RNA was reverse transcribed into cDNA and segmented into 150-200 bp by enzyme digestion. After repair, fitting, purification, PCR amplification and purification, sample concentration was assayed by DNBSEQ-T7, and an average output of more than 200 million reads was obtained. Sequencing data were compared with those in a SARS-CoV-2 database to determine whether the samples contained coronavirus sequences. For the seven complete SARS-CoV-2 genome sequences, three sequences from environmental samples (Env_0020_seq01, Env_0313_seq02 and Env_0354_seq03) were obtained from DNBSEQ-T7, and four sequences from cell supernatants of Env_0313, Env_0354 and Env_0126 (Fig. 3) were obtained from NextSeq 550 platform. A few samples were re-sequenced using a multiplex PCR approach, including Env_0020_seq01, Env_0313_seq04, Env_0313_seq05, Env_0126_seq06, and Env_0354_seq07 (Supplementary Table 3 and 4). All raw data related to the genomes, including any partial genomes that were sequenced were fully reported and deposited to the public database (Supplementary Table 3 and 4).

Virus genome assembly and phylogenetic analysis

Raw reads were adaptor- and quality- trimmed with the Fastp (version 0.20.0) program. The clean reads were mapped to the SARS-CoV-2 reference genome (GenBank: NC_045512) using Bowtie2. The assembled genomes were merged and checked using Geneious (version 11.1.5) (https://www.geneious.com). The coverage and depth of genomes were calculated with SAMtools v1.10 based on SAM files from Bowtie2.

Reference genomes, IVDC-HB-01 (GISAID: EPI_ISL_402119) and Wuhan-Hu-1 (GenBank: NC_045512), were employed as a query. Multiple sequence alignment of the seven SARS-CoV-2 sequences obtained from this study and reference sequences were performed with Mafft (v7.450). Phylogenetic analyses were performed using RAxML v8.2.9 with 1000 bootstrap replicates, employing the GTR nucleotide substitution model and the Gamma distribution.

Bioinformatic analysis of the species abundances

The Kraken2 (version 2.1.2)[31] was used for species classification with the option '--confidence 0.1'. Sequences of all species in the Nucleotide (nt) database were used for generating the index. The bracken (version 2.5) was used for re-evaluating species abundance. The matrix of species was obtained by using the pavian algorithm[32]. ggplot2 package in R was used for plotting. Read counts of each genus were used for further analysis and plotting. Raw counts of four domains (Archaea, Viruses, Eukaryota, and Bacteria), SARS-CoV-2, *Homo* genus, and *Mammalia* class were shown by heatmap (4B). Two tail unpaired t-test was used for identification of differential genus between SARS-CoV-2 PCR-positive and -negative samples.

For the analysis of the Chordata genera characterization, the reference was generated using the sequence of mitochondrial cytochrome c oxidase subunit I (COI) in the barcode of life data (BOLD) system[33-35]. RNA-seq samples were mapped to the reference sequences by the bowtie2[36] algorithm with the default settings. Read counts of each genus were calculated by the samtools[37]. Read counts over 20 were used as cut-off for the identification of positively enriched genus. Fisher's exact test was used for comparing the differential genus in the Mammalia class between SARS-CoV-2 PCR-positive and -negative samples.

Ethics

The sample collection was determined by China CDC to be part of the emergency responses to the pneumonia of unknown etiology (PUE) and therefore exempt from institutional review board assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All the raw sequencing data and genomes have been uploaded onto the GISAID (China CDC Weekly, 2021, DOI: 10.46234/ccdcw2021.255). The list of accession codes in Supplementary Table 3 and 4. The raw sequence data reported in this paper have also been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics, 2021, DOI: 10.1016/j.gpb.2021.08.001) in National Genomics Data Center (Nucleic Acids Res, 2022, DOI: 10.1093/nar/gkab951), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences that are publicly accessible at https://ngdc.cncb.ac.cn/gsa (GSA: CRA010170). The viral genomes reported in this paper have been deposited in the GenBase in National Genomics Data Center (Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation under accession numbers C_AA002295.1 to C_AA002301.1 that are publicly accessible at https://ngdc.cncb.ac.cn/genbase/. Raw sequence data were also deposited into NCBI BioProject under accession number PRJNA948658 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA948658) and in China National Microbiology Data Center (NMDC) with accession numbers NMDC10018366 (https://nmdc.cn/resource/genomics/sample/detail/NMDC10018366).

Methods References

31 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257, doi:10.1186/s13059-019-1891-0 (2019).

32 Breitwieser, F. P. & Salzberg, S. L. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **36**, 1303-1304, doi:10.1093/bioinformatics/btz715 (2020).

33 Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol Evol* **24**, 110-117, doi:10.1016/j.tree.2008.09.011 (2009).

34 Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. Identification of Birds through DNA Barcodes. *PLoS Biol* **2**, e312, doi:10.1371/journal.pbio.0020312 (2004).

35 Ratnasingham, S. & Hebert, P. D. bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes* **7**, 355-364, doi:10.1111/j.1471-8286.2007.01678.x (2007).

36 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

37 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*

639 **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
640

Acknowledgements

We gratefully acknowledge experts from Wuhan City, Hubei Province and across China who contributed to the study. We gratefully acknowledge the following experts for their invaluable contributions during this study: Prof. Naiying Mao and Prof. Yu Lan from National Institute for Viral Disease Control and Prevention, China CDC; Prof. Huaiqi Jing, and Prof. Qiyong Liu from National Institute for Communicable Disease Control and Prevention, China CDC; Prof. Lei Xu from Tsinghua University; Prof. Yongzhong Jiang, Junqiang Xu, Prof. Xixiang Huo and Dr. Bo Yu from Hubei Provincial CDC; Prof. Yan Xiong from Wuhan Municipal CDC; Prof. Juan Li from Shandong First Medical University; Prof. Weijun Chen and Dr. Honglong Wu from BGI PathoGenesis Pharmaceutical Technology. In addition, we also thank the work and suggestions of the joint team scientists of WHO-convened Global Study of Origins of SARS-CoV-2: China Part. Prof. William J. Liu is supported by the Excellent Young Scientist Program of the National Natural Science Foundation of China (NSFC, 81822040). Dr. Yun Tan's bioinformatics analyses in the study were supported by the ASTRA computing platform in the National Research Center for Translational Medicine (Shanghai) and the Pi computing platform in the Center for High Performance Computing at Shanghai Jiao Tong University.

Author contributions

The study was designed by G-Z.W., W.J.L and G.F.G. The onsite epidemiological survey and sample collection by W.J.L., W.L, Z.J., X.H., J.W., F.W., G.W., K.Q., R.G., J.Z., M.L. W.X. and G.F.G. The nucleic acid extraction and RT-PCR were performed by W.J.L., P.L., W.L, Z.J., X.H., J.W., F.W., K.C. and G.W. Next generation sequencing was performed by W.J.L., P.L., W.L, Z.J., X.H., J.W., F.W., G.W., and W.Z. Complete genome sequencing and analyses were performed by P.L., W.Z., W.S. and W.J.L. The virus isolation was performed by P.L., S.Z., W.Z., W.L., J.S. and Z.X. Data analyses were performed by W.J.L., P.L., Z.J., X.H., W.S., Y.T., S.Z., J.W., F.W., G.W., Y.G., Z.X., Y.Z., J.S., Jing Z., W.Z., W-T.Z., B.Y., J.S., M.Y., W-M.Z., Y.D., G.L., Y.B., W.T.,

and J.H. The manuscript was written by W.J.L., P.L., W.S., Y.T., Gary W., G.F.G. and G-Z.W.

Competing interest declaration

No competing interest exists.

Extended Data Figure Legends

Extended Data Fig. 1. The overground drainage pathway in the Huanan Seafood Market and environmental sample collection.

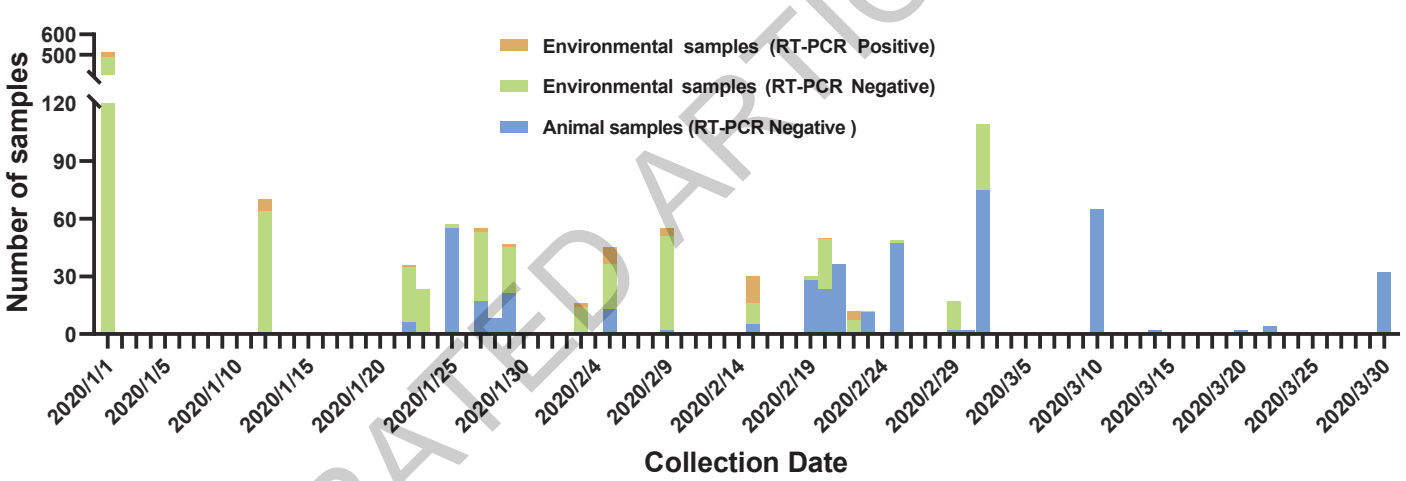
The wastewater in the overground drainage was lead into the underground drainage inside the market and then flow into the wells on the edge of the market. And we did a spot-check sampling across all the overground drainages. To detect for the presence of SARS-CoV-2 RNA, reverse transcription and quantitative polymerase chain reaction (RT-qPCR) were performed. The locations of the positive samples were marked in the map of the market within yellow.

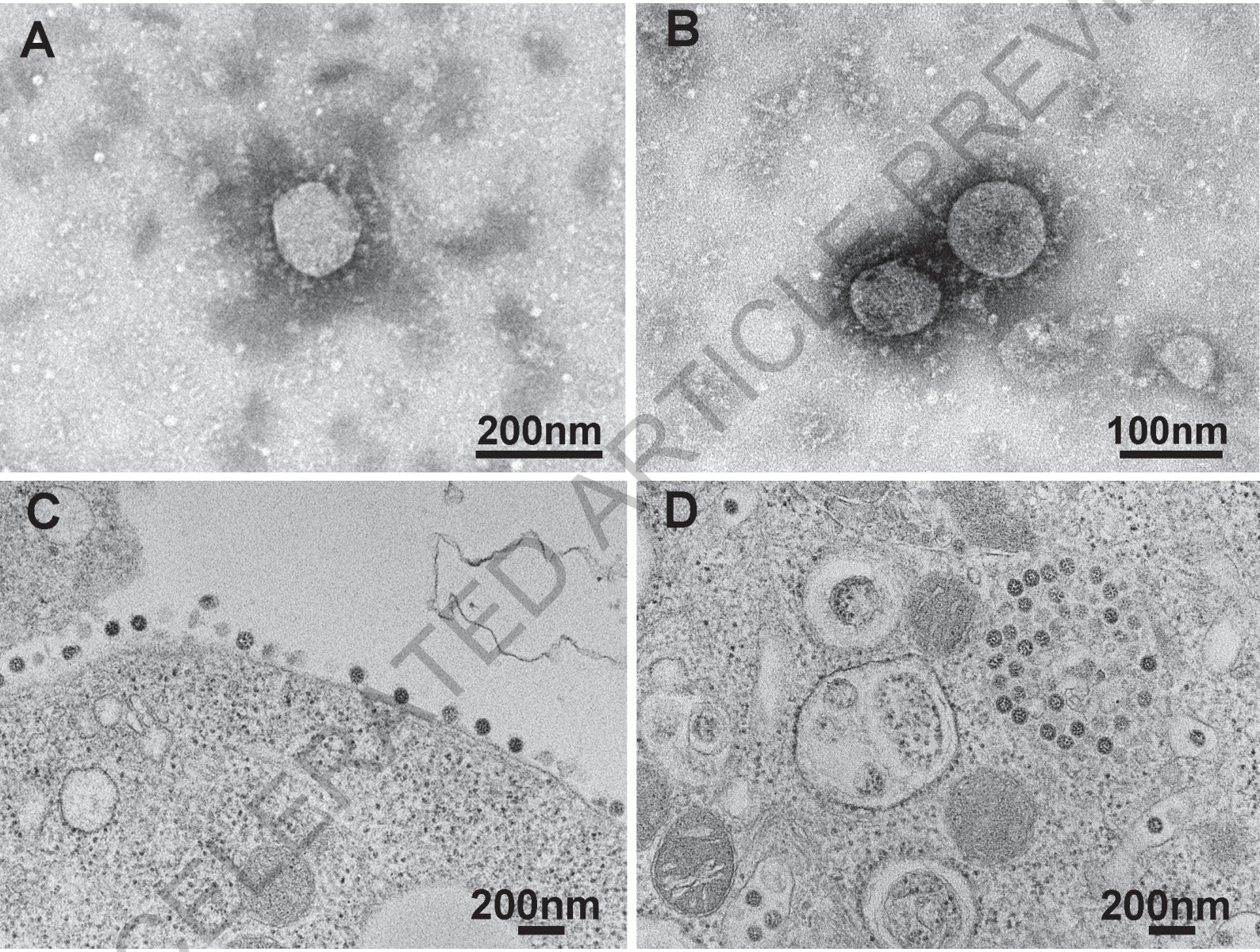
Extended Data Fig. 2. Positive environmental samples associated with different products in the Huanan Seafood Market.

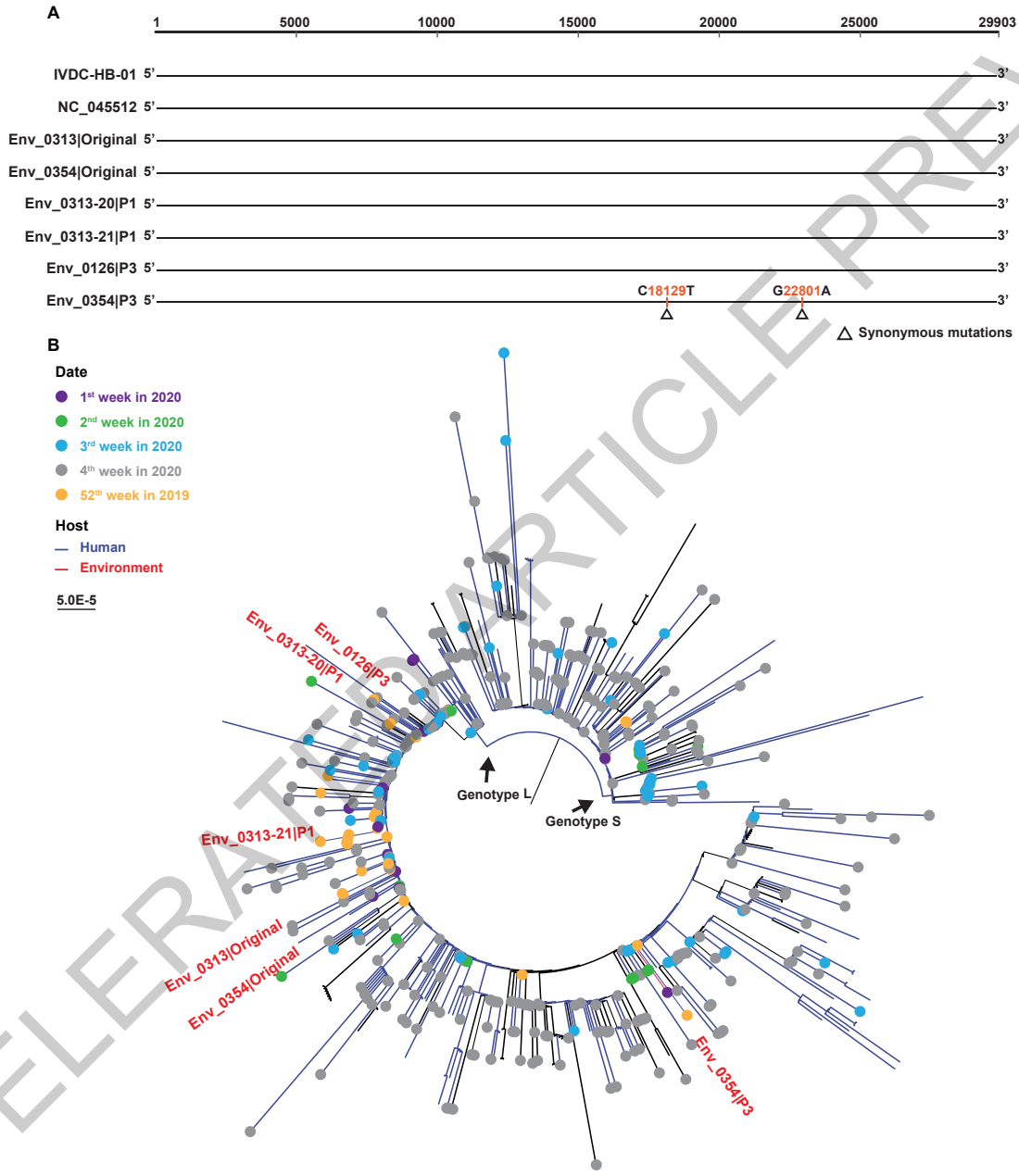
Dots represent the percentage of positive environmental samples associated with each product. Bars represent 95% confidence intervals for the binomials in the text above. Note that the confidence interval (CI) for some products (e.g. vegetables, farmed wildlife) have broad error bars that are likely due to the low number of vendors for these categories in the market. Nine of the 10 vendors selling farmed wildlife have been sampled. Data are represented as percentage in this figure.

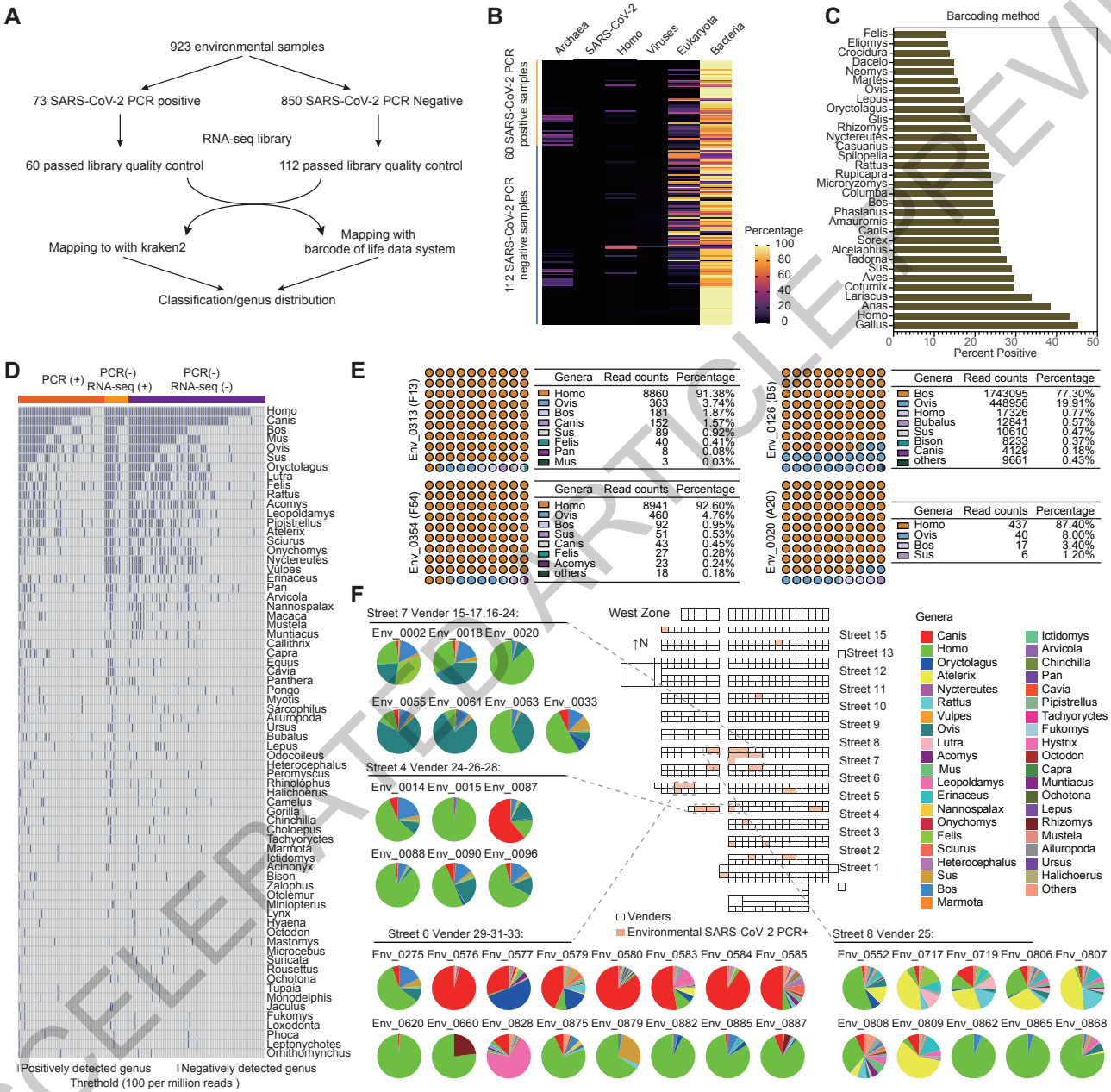
Extended Data Fig. 3. Illustration of mammal genera distribution in samples of concerns. Illustration of mammal genera distribution in samples of concerns. Samples related to the blood spot and the de-feather machine (Env_0262 and Env_0584) and samples enriched with genera related to wildlife (Env_0576, Env_0807, Env_0809, and Env_0585) were plotted. Animal genera identified by the BOLD method were shown

698 in the left panel, while mammal genera identified by the kraken2 method were shown
699 in the right panel.
700 **Extended Data Fig. 4. Distribution of the positively detected Mammal genera in**
701 **the market.** The distribution of SARS-CoV-2 and potential host were plotted by yellow
702 and blue dots, respectively. The density of the distribution of potential host was shown
703 in red, while the SARS-CoV-2 by green.
704
705 **Extended Data Tables**
706 **Extended Data Table 1. Overview of environmental sample sampling and testing**
707 **in the Huanan Seafood Market.**
708 **Extended Data Table 2. The collection logic of the environment samples.**
709 **Extended Data Table 3. The collection logic of the animal samples.**
710 **Extended Data Table 4. The information of the sampling in other markets.**
711 **Extended Data Table 5. Twenty-one shops of RT-PCR positive in the Huanan**
712 **Seafood Market.**
713 **Extended Data Table 6. The animal samples collected in the Huanan Seafood**
714 **Market.**
715

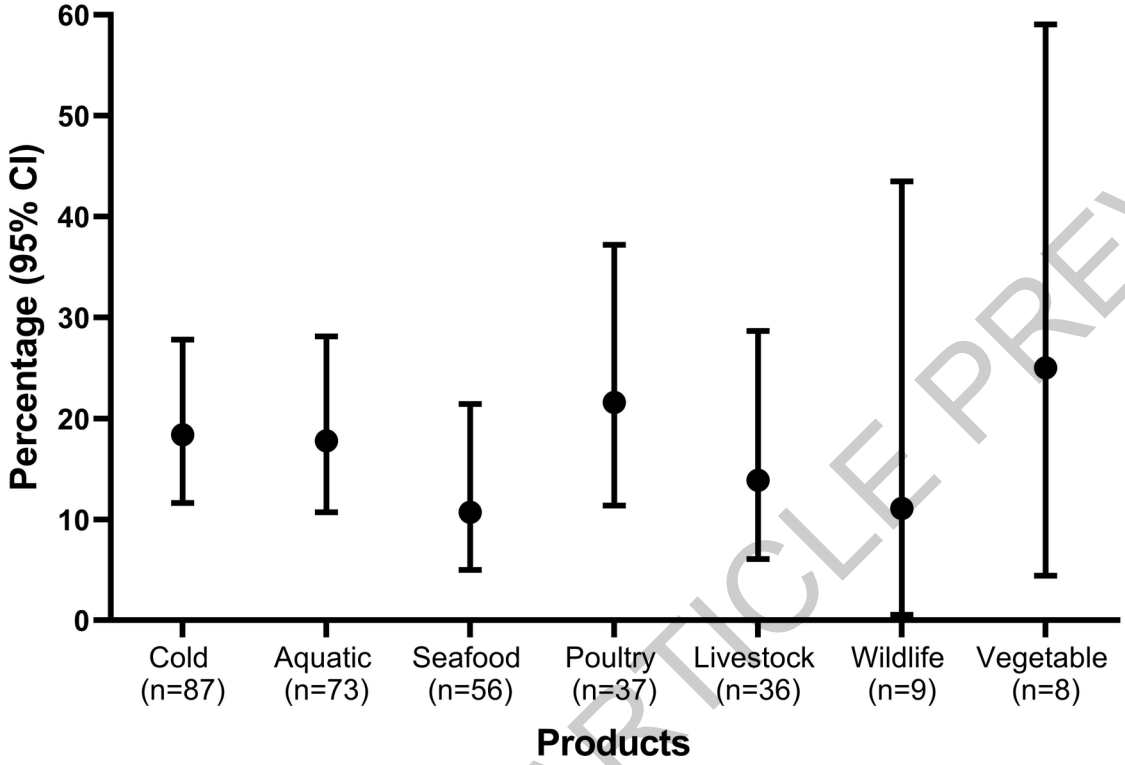




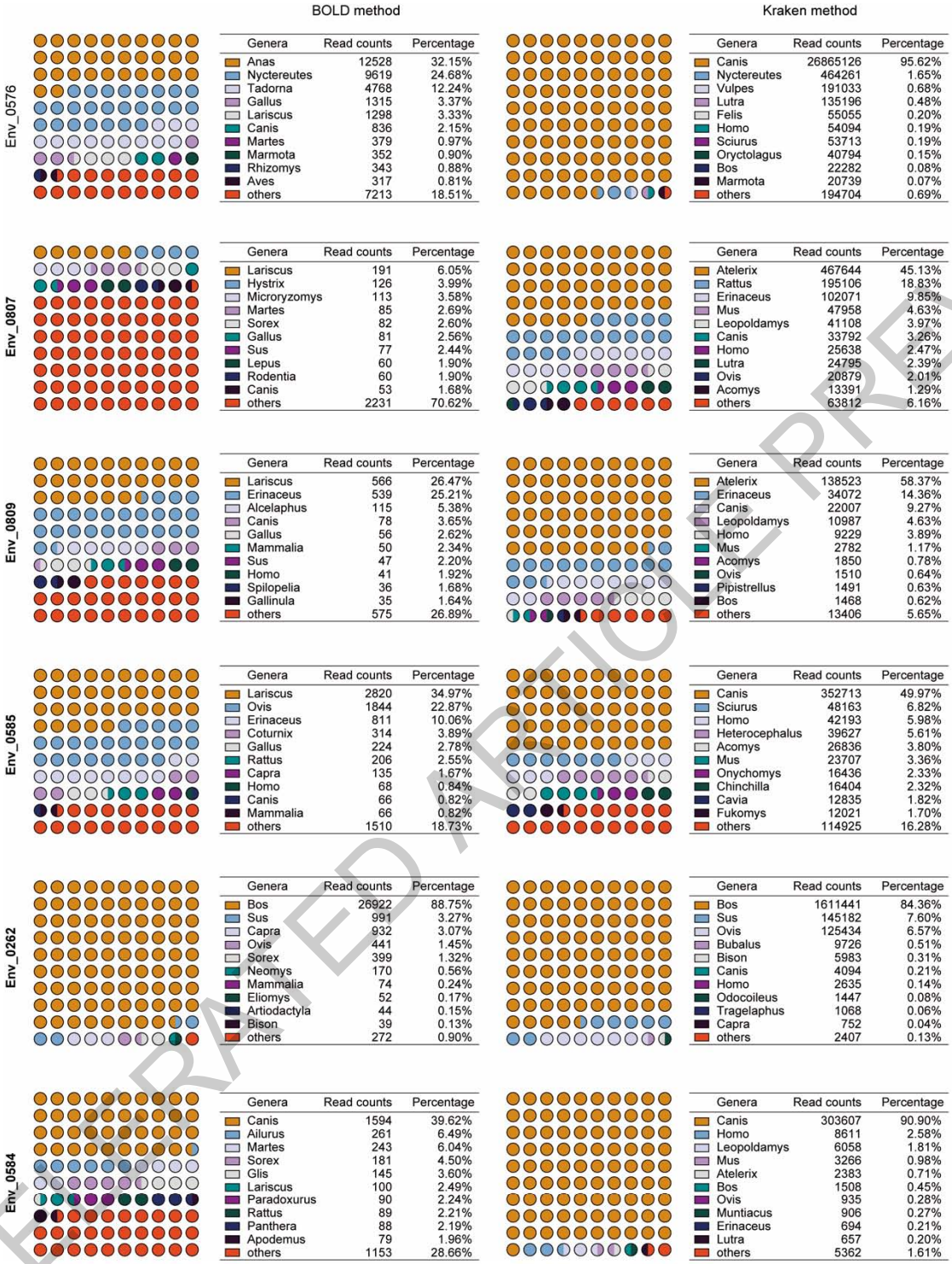




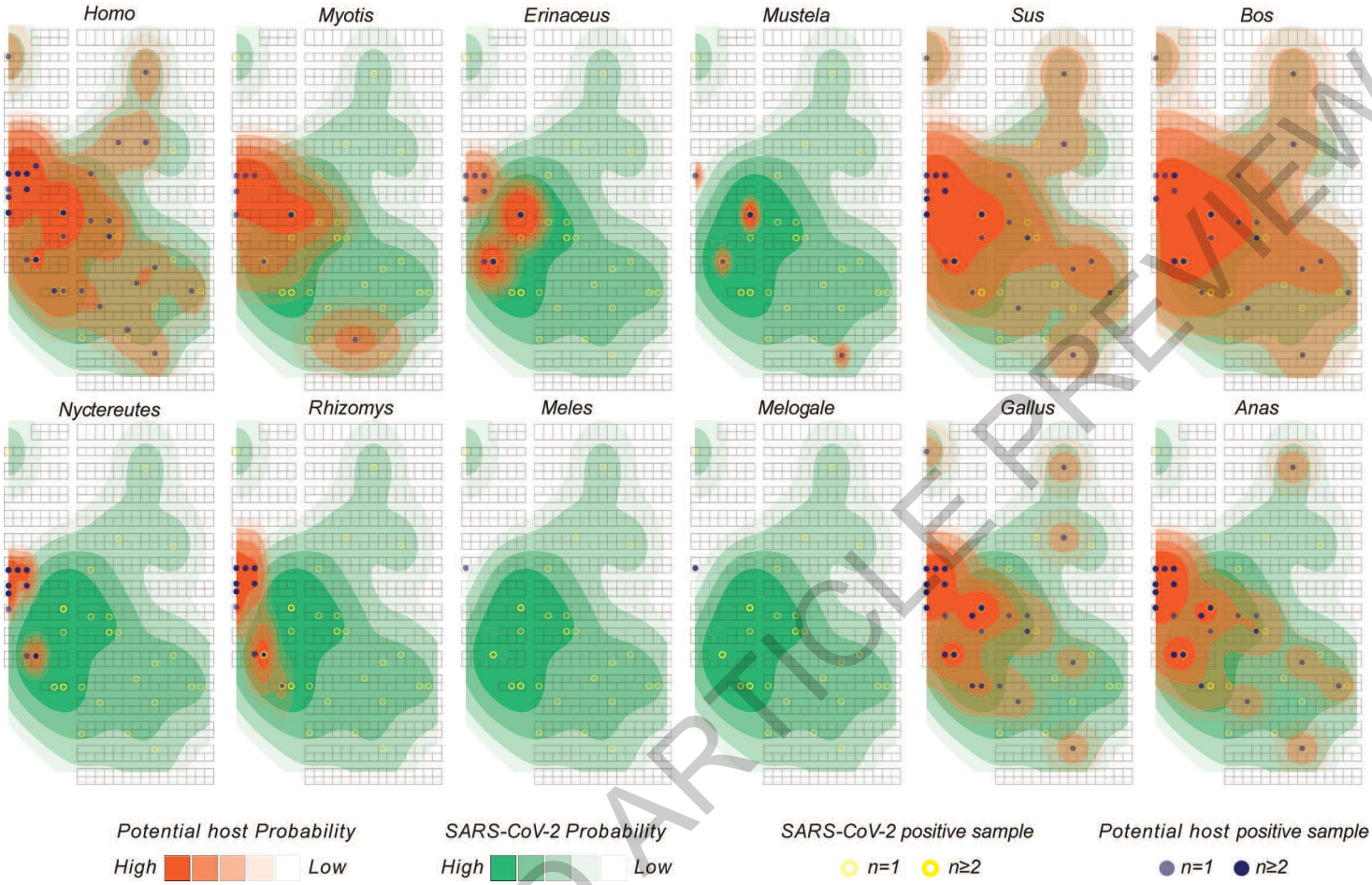
Extended Data Fig. 1



Extended Data Fig. 2



Extended Data Fig. 3



Extended Data Fig. 4

Extended Data Table 1. Overview of environmental sample sampling and testing in the Huanan Seafood Market.

	Number of samples	Number of positive samples by RT-PCR	Number of isolated viruses
Huanan Seafood Market	718	40	3
Warehouses related to the Huanan Seafood Market[a]	14	5	
Other markets in Wuhan and Huanggang[b]	30	1	
Drainage system in the Huanan Seafood Market	110	24	
Sewerage wells in surrounding areas	51	3	
Total	923	73	3

[a] The warehouses related to the Huanan Seafood Market were located out of the market.

[b] The one positive sample outside HSM was collected from Dongxihu Market in Wuhan. More information was provided in Extended Data Table 4.

Extended Data Table 2. The collection logic of the environment samples.

No.	Time	Objective	Sample time	Amount	Sum
1	1,Jan	(1) Environmental samples from stalls related to early cases; (2) Environmental samples from doors and floors of all stalls in the blocks where the early cases were located; (3) Environmental samples in the east wing of the market were collected according to blocks; (4) Transport carts, trash cans and similar objects.	1,Jan	515	515
2	12,Jan	Environmental samples from stalls that sold livestock, poultry, farmed wildlife (also called domesticated wildlife).	12,Jan	70	70
3	22,Jan	Environmental samples from other markets in Wuhan	22,Jan	30	30
4	23,Jan-19,Feb	The outdoor environmental samples from stalls that sold livestock, poultry, farmed wildlife.	23,Jan	23	52
			25,Jan	2	
			3,Feb	16	
			9,Feb	5	
			15,Feb	4	
			19,Feb	2	
5	27,Jan-15,Feb	Samples of sewage and silt from drainage channels and sewerage wells in the market.	27,Jan	38	94
			29,Jan	26	
			9,Feb	9	
			15,Feb	21	
6	5,Feb-9,Feb	Samples of sewage and silt from city sewerage wells around the market.	5,Feb	32	71
			9,Feb	39	
7	20,Feb-2,Mar	(1) Cold storages and refrigerators from stalls that sold livestock, poultry, farmed wildlife in the market; (2) The market's ventilation and air-conditioning system; (3) Public toilets, public activity rooms and other places where people gathered in the market.	20,Feb	27	91
			22,Feb	12	
			23,Feb	1	
			25,Feb	2	
			29,Feb	15	
			2,Mar	34	
Total				923	

Extended Data Table 3. The collection logic of the animal samples.

No.[a]	Time	Objectives	Sample time	Amount	Sum
8	22,Jan	Animal products in other markets.	22,Jan	6	6
9	25,Jan-10,Mar	Animal products and other commodity samples kept in the cold storages and refrigerators in the market.	25,Jan	55	306
			20,Feb	23	
			21,Feb	36	
			23,Feb	5	
			25,Feb	47	
			2,Mar	75	
			10,Mar	65	
10	27,Jan-1,Mar	Live animals captured around the market.	27,Jan	5	17
			5,Feb	3	
			9,Feb	2	
			15,Feb	3	
			29,Feb	2	
			1,Mar	2	
11	18,Jan-30,Mar	Stray cats, mice, cat feces and other stray animals (one dog and one weasel in the market).	18,Jan	1	96
			27,Jan	12	
			28,Jan	8	
			29,Jan	21	
			5,Feb	10	
			15,Feb	2	
			23,Feb	2	
			14,Mar	2	
			20,Mar	2	
			22,Mar	4	
			30,Mar	32	
12	19,Feb-23,Feb	Animal products and other commodity samples kept in the cold storages.	19,Feb	28	32
			23,Feb	4	
Total				457	

[a] The number follows the upper Table for environment samples.

Extended Data Table 4. The information of the sampling in other markets.

District	Number of environment samples [a]	Number of positive environment samples by RT-PCR	Number of animal samples [b]	Number of positive animal samples by RT-PCR
Jiang'han district	7	0	2	0
Jiang'an district	8	0	2	0
Donxihu district	7	1	1	0
Huanggang city	8	0	1	0
Total	30	1	6	0

[a] Swab sample collected from the floor, wall or chopping board.

[b] The heart, liver and large intestine tissues from pigs.

Extended Data Table 5. Twenty-one shops of RT-PCR positive in the Huanan Seafood Market.

Vendors No.	Location	Product types [a]						
		Cold-chain products	Aquatic products	Seafood products	Poultry	Livestock	Wildlife products	Vegetables
1	West	no	no	no	yes	no	no	no
2	West	yes	yes	yes	no	no	no	no
3	West	yes	yes	no	yes	yes	yes	no
4	East	yes	no	no	yes	yes	no	no
5	West	no	no	no	no	no	no	no
6	West	no	yes	no	yes	yes	no	no
7	West	yes	no	no	yes	no	no	no
8	West	yes	yes	yes	yes	no	no	no
9	West	yes	yes	yes	no	no	no	no
10	West	yes	yes	yes	yes	yes	no	no
11	West	yes	yes	no	no	no	no	no
12	West	yes	yes	yes	no	no	no	no
13	West	yes	yes	no	no	no	no	no
14	West	yes	yes	no	no	no	no	no
15	West	yes	yes	no	no	no	no	no
16	West	yes	yes	no	no	no	no	no
17	West	no	no	no	no	no	no	no
18	West	yes	no	no	yes	yes	no	no
19	West	no	no	no	no	no	no	yes
20	West	yes	no	no	no	no	no	yes
21	East	yes	yes	yes	no	no	no	no
Sum of NAT positive vendors		16	13	6	8	5	1	2
Vendors sampled in the study selling such products		87	73	56	37	36	9	8

[a] "yes" indicates product sold by vendors; "no" indicates product not sold by vendors.

Extended Data Table 6. The animal samples collected in the Huanan Seafood Market.

Species	Animal number	Sample number	RT-PCR positive number
Rabbit/Hares	52	104	0
Stray cat	27	80[a]	0
Snake	40	80	0
Hedgehog	16	67	0
Muntjac	6	18	0
Dog	7[b]	17	0
Badger	6	16	0
Bamboo rat	6	15	0
Mouse	10	12	0
Pig	NA[c]	6[d]	0
Chicken	5	5	0
Chinese giant salamander	3	5	0
Crocodile	2	4	0
Wild boar	2	4	0
Soft-shelled turtle	2	3	0
Weasel[e]	1	2	0
Fish	2	2	0
Sheep	1	1	0
Others	NA[f]	16	0
Total	188	457	0

[a] Six of the cats were from the Huanan Seafood Market. And the samples included faeces.

[b] Including one stray dog in the Huanan Seafood Market.

[c] Not applicable due to the processed pork.

[d] Collected from other markets.

[e] The weasel was not sold in the market, but caught alive in the Market.

[f] Not applicable due to the unrecognized “bai tiao” product as described in the methods.

nature portfolio

Corresponding author(s): Guizhen wu

Last updated by author(s): Mar 29, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For RNA-seq, raw datasets were collected with the BGI's Sequencing Systems.
Data analysis	Virus genome assembly and phylogenetic analysis Raw reads were adaptor- and quality- trimmed with the Fastp (version 0.20.0) program. The clean reads were mapped to the SARS-CoV-2 reference genome (GenBank: NC_045512) using Bowtie2. The assembled genomes were merged and checked using Geneious (version 11.1.5) (https://www.geneious.com). The coverage and depth of genomes were calculated with SAMtools v1.10 based on SAM files from Bowtie2. Reference genomes, IVDC-HB-01 (GISAID: EPI_ISL_402119) and Wuhan-Hu-1 (GenBank: NC_045512), were employed as a query. Multiple sequence alignment of the seven SARS-CoV-2 sequences obtained from this study and reference sequences were performed with Mafft v7.450. Phylogenetic analyses were performed using RAxML v8.2.9 with 1000 bootstrap replicates, employing the GTR nucleotide substitution model and the Gamma distribution. Bioinformatic analysis of the species abundances The Kraken2 (version 2.1.2) was used for species classification with the option '--confidence 0.1'. Sequences of all species in the Nucleotide (nt) database were used for generating the index. The bracken (version 2.5) was used for re-evaluating species abundance. The matrix of species was obtained by using the pavian algorithm. ggplot2 package in R was used for plotting. Read counts of each genus were used for further analysis and plotting. Raw counts of four domains (Archaea, Viruses, Eukaryota, and Bacteria), SARS-CoV-2, Homo genus, and Mammalia class were shown by heatmap. Two tail unpaired t-test was used for identification of differential genus between SARS-CoV-2 PCRpositive and -negative samples. For the analysis of the mammalian genus characterization, the reference was generated using the sequence of mitochondrial cytochrome c oxidase subunit I (COI-5P) in the barcode of life data (BOLD) system. RNA-seq samples were mapped to the reference sequences by the bowtie2 algorithm with the default settings. Read counts of each genus were calculated by the samtools. Read counts over 20 were used as

cut-off for the identification of positively enriched genus. Fisher's exact test was used for comparing the differential genus in the Mammalia class between SARS-CoV-2 PCR-positive and -negative samples.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All the raw sequencing data have been uploaded onto the GISAID (China CDC Weekly, 2021, DOI: 10.46234/ccdcw2021.255). The list of accession codes in Extended Data Table 6 and 7. The raw sequence data reported in this paper have also been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics, 2021, DOI: 10.1016/j.gpb.2021.08.001) in National Genomics Data Center (Nucleic Acids Res, 2022, DOI: 10.1093/nar/gkab951), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA010170) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Raw sequence data was deposited into NCBI BioProject under accession number PRJNA948658 (http://www.ncbi.nlm.nih.gov/bioproject/948658) and in China National Microbiology Data Center (NMDC) with accession numbers NMDC10018366 (https://nmdc.cn/resource/genomics/sample/detail/NMDC10018366).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We presented the SARS-CoV-2 detection results of 1380 samples collected from the environment and the animals within the market in early 2020. We further conducted RNA-seq analysis.
Research sample	Environmental samples in the Huanan Seafood Market were collected to represent exhaustively as possible, from a wide diversity of surfaces, animals and products.
Sampling strategy	Please refer to the sample collection in the method section.
Data collection	Please refer to the sample collection in the method section.
Timing and spatial scale	Please refer to the sample collection in the method section and Table 1, Table 2 and Supplementary Table 1.
Data exclusions	For the RNA-Seq analysis. 73 SARS-CoV-2 positive environmental samples were used for RNA-seq library construction. Among these, 60 samples successfully passed the library quality control and were used for analysis. A total of 850 SARS-CoV-2 negative environmental samples were obtained. Among these, 112 samples with high RNA abundance were used for RNA-seq analysis. Finally, 172 samples were used for analysis, and no samples were excluded.
Reproducibility	All samples used in the current study were unique, thus it would not be able to repeat the experiments.

Randomization	Not related.
Blinding	Not related.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Please refer to the sample collection in the method section, Extended Data Table S1 to Table S5, and also Supplementary Table 1.
Location	Please refer to the sample collection in the method section, Extended Data Table S1 to Table S5, and also Supplementary Table 1.
Access & import/export	The sample collection was guided and conducted by China CDC.
Disturbance	Huanan Seafood Market was a large market with more than 600 stalls inside. Thus, it was a tough job to finish the sampling of all the stalls in a short time during the emergency response to COVID-19. Thus, we performed the sample collection according to the ten different sampling principles we summarized in the methods. Especially, early case-related stores and wildlife-related stores were prioritized for sample collection and repeated sampling were also performed in these locations. Thus, it should be noted that these factors may lead to a biased sampling in the market.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	ATCC.
Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	Tested negative.
Commonly misidentified lines (See ICLAC register)	None.