

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://nmcd.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. 70 SARS-CoV-2 positive environmental samples by RT-qPCR were used for RNA-seq library construction. Among these, 57 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, 115 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.