

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 No software was used for data collection. Code and software (including version numbers) used to process sequence-data generated in this study are provided in the Methods and associated linked Github repository (and cited with zenodo doi).

Data analysis All the code and software used for the data analysis is available in GitHub repo (and cited with zenodo link).

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](#)

The Hong Kong SARS-CoV-2 genome sequences and associated metadata generated in this study have been deposited in GenBank (accession numbers are available on GitHub at <https://git.io/JyRjK>) and GISAID (accession numbers are available on GitHub at <https://git.io/JyRD7>). Details of confirmed cases of COVID-19 infection in Hong Kong are available from CHP (<https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent>). SARS-CoV-2 reference genome (Wuhan-Hu-1, GenBank: MN908947.3) is available on GenBank. Sequence data from the other countries/regions were obtained from GISAID (accession numbers and acknowledgements are provided in Supplementary Information Data 4). Public transit data was provided by Octopus Cards Limited (Octopus). We obtained consent

from Octopus to share the aggregate data of transport transactions between January 1, 2020 and May 31, 2021. Our agreement with Octopus prohibits us from further sharing data with third parties, but interested parties may contact Octopus.

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](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	This is a genomic epidemiology study that used all available genomic data to infer origin and transmission dynamics of SARS-CoV-2
Data exclusions	Sequence data was excluded only if they were of low quality or incomplete, as described in the Methods.
Replication	For all phylogenetic analysis, a minimum of two independent Bayesian analysis runs were performed, and their convergence and effective sample sizes were confirmed to be adequate prior to inference. For within-host evolution experiments, twelve samples were sequenced in duplicate to exclude any sequencing artifacts affecting within-host diversity estimates, as stated in Methods.
Randomization	Groups were not compared in this study.
Blinding	Blinding was not relevant as routine surveillance samples with de-identified patient data were utilised in the study.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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