Corresponding author(s): Maciej Boni, Philippe Lemey, Andrew Rambaut, David Robertson

Last updated by author(s): July 7, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	All in Methods section of paper.
Data analysis	Software used: RAxML (v8.2.8), 3SEQ (v1.7), RDP5 (v5), GARD (v2.5.0), BEAST (v1.10.4), BEAGLE (v3), MAFFT (v7.310), Neighbor-Nets (SplitsTree v4.15.1), PHI-Test (SplitsTree v4.15.1), IQTREE (v2.0), TempEst (v1.5.3), Tree Annotator (v1.10.4), FigTree(v1.4.2), Phylogenetic Diversity Analyzer Tool (v0.5), BBMap (v38.75), pheatmap (v1.0.12). All custom code available at https://github.com/plemey/SARSCoV2origins

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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:

- All sequence data are available in GenBank and GISAID, with accession numbers listed in Supplementary Table S4.

All sequence data analyzed in this manuscript are available at https://github.com/plemey/SARSCoV2origins. Note that one of the sequences requires GISAID permissions from original authors, but will be made available either individually or publicly depending on original depositor's request.

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	Sarbecovirus genomes were downloaded from publicly available sources. Non-recombinant regions were identified. Evolutionary rate calibration for coronaviruses was performed using publicly available data sets for SARS, MERS, and OC43. An evolutionary rate was inferred for three putative non-recombinant regions of the sarbecoviruses. Time to the most common ancestor between SARS-CoV-2 and its closest bat-virus relative were computed.
Research sample	All available genomes (n=68) for sarbecoviruses -- including the SARS-CoV-2 genome; 27 genomes of human coronavirus OC43; 35 genomes of MERC-CoV; 69 genomes of SARS-CoV.
Sampling strategy	No sample size calculation was done as it was possible to use all samples (all available sequences) in the analysis.
Data collection	Publicly available sequence data were downloaded from GenBank and GISAID.
Timing and spatial scale	Sarbecovirus genomes were collected over an 18-year period (2002-2020), OC43 sequences from 1968 to 2016, MERS sequences from 2012 to 2015, and SARS-CoV sequences from 2002-2004.
Data exclusions	Sequences without 'sampling year' were excluded. This exclusion criterion was not formally pre-established, but this exclusion is both common and necessary when estimating evolutionary rates from sequence data. It goes without saying that sequences without dates must be excluded.
Reproducibility	N/A;
Randomization	N/A
Blinding	N/A

Did the study involve field work? ☐ Yes ☒ No

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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