nature research

Corresponding author(s): Andreas Bergthaler

Last updated by author(s): Jun 8, 2022

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 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	Microsoft SQL Server database was used to collect and organise data and metadata
Data analysis	BWA-MEM (v 0.7.17): read mapping iVar (v 1.2.3): amplicon primer masking LoFreq (v 2.1.2): realignment and variant calling Bcftools (v 1.9): variant filtering with SnpEff (v 4.3): variant annotation SnpSift (v 4.3): variant annotation R (v 4.0.4): statistical analysis and data visualisation gamlss (v 5.3.4) : R package for model fitting MBA (v 0.0.9) : R package for spatiotemporal interpolation EpiEstim (v 2.2-3): R package for reproduction number estimation SNPgenie (v 2019.10.31): calculation of nucleotide diversity pi The source code for the developed and applied software for variant quantification (VaQuERo) and mutation deconvolution (DeViVa) is available on GitHub (https://github.com/fabou-uobaf/VaQuERo and https://github.com/SebH87 /DeViVa).

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 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:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw sequencing data used during this study, with additional sequencing data produced for samples drawn from Austrian WWTP before December 2020, are available under the following ENA accession number: PRJEB48985. Access to detailed epidemiological case data, collected by the Austrian Agency for Health and Food Safety (AGES; https://covid19-dashboard.ages.at/), can be granted for research purposes in compliance with local regulatory and legal frameworks, to exclude retraceability of individual cases, in due time upon request via the corresponding author. SARS-CoV-2 genome sequences for marker mutation definition can be obtained from GISAID (https://www.gisaid.org/) by registered users agreeing to the effective terms of use (https://www.gisaid.org/registration/terms-of-use/).

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	Deduction of SARS-CoV-2 variant abundance from wastewater samples.
Research sample	SARS-CoV-2 (Taxonomy ID: 2697049) collected from wastewater at treatment plants was used to examine virus occurence in the population residing in the respective catchments to avoid sampling bias and some economic constraints of clinical surveillance.
Sampling strategy	Sampled wastewater treatment plants cover up to 57% of the Austrian population. The selection of monitored WWTPs was guided by the objective to cover as many people in as many parts of the country as possible, with a special emphasis on towns with larger school facilities, as well as access to samples and appropriate logistic chains. No statistical sample size calculation was performed.
Data collection	24 h volume equivalent mixed samples collected by the wastewater treatment plant operator with onsite installed 24-hour refrigerated automatic composite samplers.
Timing and spatial scale	Samples collected in the period from December 2020 and February 2022, from 94 wastewater treatment plants across Austria. Time scale determination is guided by the onset of the large scale monitoring program till the end of observation periode, after arrival of Omicron variants BA.1 and BA.2. Spatial scale selection was guided by the objective to cover as many people in as many parts of the country as possible, with a special emphasis on towns with larger school facilities, as well as access to samples and appropriate logistic chains.
Data exclusions	Data point failed to be successfully sequencing (less than 40% of the virus genome covered by at least 10 reads) were excluded for further analysis, namely Variant detection/quantification, variant deconvolution, nucleotide diversity measurments, and reproductive number estimation.
Reproducibility	Technical reproducibility was evaluated by analysis triplicates from two independent wastewater treatment plants
Randomization	No group comparison was performed, hence no allocation of samples into groups has to be performed.
Blinding	No blinding was performed. Integration with epidemiological case data was only performed in hindsight.

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 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