nature portfolio

Corresponding author(s): Margo Diricks

Last updated by author(s): 05.07.2022

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	Genome assemblies were downloaded directly from https://www.ncbi.nlm.nih.gov/assembly. FastQ files were downloaded from SRA using fastq-dump.2.8.2 and GNU parallel v. 20161222
Data analysis	Assemblies were made using shovill v1.1.0, (skesa v2.4.0 + SPAdes v3.15.0; trimmomatic v0.39) or SeqSphere+ v7.7.5. CgMLST scheme creation (cgMLST target definer v1.5) and analysis was performed in SeqSphere+ v7.7.5. Gene extraction was performed in BioNumerics v7.6 and cgSNP analysis with MTBseq v1.0.4. Bash scripts used for this study are available at https://github.com/ngs-fzb/NTMtools, but include only publicly available tools. The cgMLST scheme is publicly available at cgmlst.org (https://www.cgmlst.org/ncs/schema/22602285/).

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

Accession numbers of all whole genome sequencing datasets analyzed in this study are listed in Supplementary Table 1 and 3. The cgMLST scheme is publicly available at cgmlst.org (https://www.cgmlst.org/ncs/schema/22602285/).

Human research participants

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

Reporting on sex and gender	No human participants were involved in this study
Population characteristics	No human participants were involved in this study
Recruitment	No human participants were involved in this study
Ethics oversight	No human participants were involved in this study

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

Life sciences study design

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

Sample size	For this study, we downloaded assemblies (fastA files) from the NCBI/RefSeq Assembly database for 1,797 isolates (Supplementary Table 1) and Illumina paired-end read datasets (fastQ files) from the sequence read archive (SRA) for 372 isolates (Supplementary Table 3). The total dataset (n=2,169) comprised 1991 unique biosamples.
Data exclusions	Initially, All assemblies with more than 300 contigs (n=13) were removed from further analysis.
Replication	For technical validation, we used same read sets of 30 isolates and created different assemblies with different algorithms. Allele numbers did not differ between these replicates.
Randomization	Samples were allocated into experimental groups (e.g. extra-pulmonary vs pulmonary outbreaks) according to the metadata that was available for these samples.
Blinding	This is not applicable as the associated metadata unambiguously resulted in the allocation of the samples within a specific group.

Behavioural & social sciences study design

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

Study description	Not applicable

Research sample	Not applicable
Sampling strategy	Not applicable
Data collection	Not applicable
Timing	Not applicable
Data exclusions	Not applicable
Non-participation	Not applicable
Randomization	Not applicable

Ecological, evolutionary & environmental sciences study design

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

Study description	Not applicable
Research sample	Not applicable
Sampling strategy	Not applicable
Data collection	Not applicable
Timing and spatial scale	Not applicable
Data exclusions	Not applicable
Reproducibility	Not applicable
Randomization	Not applicable
Blinding	Not applicable

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	Not applicable
Location	Not applicable
Access & import/export	Not applicable
Disturbance	Not applicable

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

Antibodies

Antibodies used	Not applicable
Validation	Not applicable

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Not applicable
Authentication	Not applicable
Mycoplasma contamination	Not applicable
Commonly misidentified lines (See ICLAC register)	Not applicable

Palaeontology and Archaeology

Specimen provenance	Not applicable
Specimen deposition	Not applicable
Dating methods	Not applicable

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	Not applicable

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Not applicable
Wild animals	Not applicable
Reporting on sex	Not applicable
Field-collected samples	Not applicable
Ethics oversight	Not applicable

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Not applicable
Study protocol	Not applicable
Data collection	Not applicable
Outcomes	Not applicable

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Not applicable
Files in database submission	Not applicable
Genome browser session (e.g. UCSC)	Not applicable

Methodology

Replicates	Not applicable
Sequencing depth	Not applicable

Antibodies	Not applicable
Peak calling parameters	Not applicable
Data quality	Not applicable
Software	Not applicable

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Not applicable
Instrument	Not applicable
Software	Not applicable
Cell population abundance	Not applicable
Gating strategy	Not applicable

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	Not applicable
Design specifications	Not applicable
Behavioral performance measures	Not applicable

Acquisition

Imaging type(s)	Not applicable
Field strength	Not applicable
Sequence & imaging parameters	Not applicable
Area of acquisition	Not applicable

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Not applicable
Normalization	Not applicable
Normalization template	Not applicable
Noise and artifact removal	Not applicable
Volume censoring	Not applicable

Statistical modeling & inference

Model type and settings	Not applicable
Effect(s) tested	Not applicable

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	Not applicable
Correction	Not applicable

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	Not applicable
Graph analysis	Not applicable
Multivariate modeling and predictive analysis	Not applicable