nature portfolio

Corresponding author(s): DAPR, Manuscript number:
NCOMMS-25-104422

Last updated by author(s): Jan 13, 2026

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 to collect data for this study
Data analysis	The software used in this study was standard for the field and usage was described in Methods: we used Sickle v1.33, Bowtie 2 v2.29, SAMtools v1.31, VarScan v2.3.9, RAxML v2.8.12, FastBAPS(R code), PastML(https://pastml.pasteur.fr/), BEAST v1.10.4, SNPPar v1.0. Any other scripts used were small custom scripts also described in Methods.

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 data used in this study were obtained from publicly available sources. The 401 BioProject accession numbers and detailed information for the 91,177 Mycobacterium tuberculosis isolates analyzed in this study are provided in Supplementary Table 1.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This study does not involve human participants, vertebrate animals, or other subjects for which sex or gender would be relevant. Therefore, sex and gender were not considered in the study design, no data on sex or gender were collected, and no sex- or gender-based analyses were performed.
Reporting on race, ethnicity, or other socially relevant groupings	This study does not involve human participants, vertebrate animals, or other subjects for which race, ethnicity, or socially relevant groupings would be applicable. No socially constructed or socially relevant categorization variables were used, and therefore no classification, definitions, or confounding controls related to these variables were required.
Population characteristics	This study does not involve human participants. Therefore, no population characteristics (e.g., age, genotype, diagnosis, or treatment information) were collected or analyzed.
Recruitment	This study does not involve human participants. Therefore, no participant recruitment was performed, and issues such as self-selection bias or related biases are not applicable.
Ethics oversight	This study does not involve human participants or vertebrate animals. Therefore, no ethics approval or oversight was required.

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	The research sample comprises a total of 91,177 Mycobacterium tuberculosis whole-genome sequences obtained from publicly available repositories. This global dataset includes isolates from Lineage 3 (n = 7,679) as well as Lineage 2 and Lineage 4 strains, which were incorporated for comparative phylogenetic and evolutionary analyses. The genomes represent isolates collected across 38 countries spanning five continents. All sequence data and associated metadata were derived from public databases, and detailed BioProject accession numbers and isolate information are provided in Supplementary Table 1.
Data exclusions	Genomes were excluded only if they failed predefined quality-control criteria, including insufficient coverage, evidence of contamination, or ambiguous lineage assignment. No data were excluded based on geographic origin, resistance profile, or phylogenetic position.
Replication	All analyses were conducted using reproducible computational workflows. Phylogenetic reconstruction, clustering, demographic inference, and selection analyses were repeated using alternative parameters and methods to confirm robustness, yielding consistent results. All software tools, parameters, and analysis steps are described in the Methods.
Randomization	Randomization was not applicable, as this study analyzed existing genomic datasets and did not involve experimental allocation of samples or organisms.
Blinding	Blinding was not applicable because the study involved computational analyses of genomic data without experimental interventions or outcome assessment.

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
☒ ☐	Plants

Methods

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

Plants

Seed stocks

This study does not involve plants, seed stocks, or other plant material. Therefore, no plant specimens were collected or used.

Novel plant genotypes

This study does not involve plants or the generation of novel plant genotypes. No transgenic, gene-edited, chemically or radiation-mutagenized, or hybrid plant lines were produced or analyzed.

Authentication

This study does not involve plants, seed stocks, or novel genotypes. Therefore, no authentication procedures, mutation effect assessments, or evaluations of potential secondary effects were performed.