Corresponding author(s): Sebastien Gagneux

Last updated by author(s): Mar 30, 2021

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	No software was used for data collection
Data analysis	R v3.3.1 R v3.6.2 Rstudio v1.1.456 phybreak v0.2.0 Trimmomatic v0.33 Seqprep v1.2 Picard v2.1.1 BWA v0.7.12 GATK v3.4.0 Samtools v1.2 VarScan v2.4.1 SnpEFF v4.11 RAxML v8.2.8 NetworkX v2.2 Gephi v0.9.2 cluster v2.0.6 fitdistrplus v1.0-11 epitools v.0.5-10.1

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 <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. 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

The raw sequences were deposited at the European Nucleotide Archive under the BioProject ID PRJEB39561. Accession numbers are listed in Supplementary Table S2. Metadata associated with the genomes is provided in Supplementary Table S3.

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	To study the transmission of multidrug-resistant (MDR) M. tuberculosis isolates, we used a retrospective collection of a total of 2063 M. tuberculosis isolates in the Republic of Georgia between 2011 and 2016. These samples were collected during the routine operation of the Georigan National Tuberculosis Control program and processed at the National Center for Tuberculosis and Lung Diseases in Tbilisi (Georgia) for routine drug susceptibility testing. We subjected these strains to whole genome sequencing. We excluded 450 strains as they were either contaminated, did not show an MDR resistance profile, were mixed infections or failed during sequencing. The final sample set consisted of 1613 MDR M. tuberculosis isolates, representing ca. 70 % of all culture confirmed MDR M. tuberculosis cases in the country between 2011 and 2016. We used the whole genome sequences together with limited, pseudonomized metadata to study transmission of MDR M. tuberculosis and epidemiological factors associated with transmission.
Research sample	The sample set represented a retrospective M. tuberculosis collection. Samples were collected by the Georgian National Tuberculosis Control Program for routine drug susceptibility testing, whereby a sample was frozen and stored. The authors of the study did not have any role in sample collection and the dataset represents a retrospective convenience sample covering an estimated 70 % of the MDR-TB epidemic in Georgia.
Sampling strategy	We aimed at covering the MDR M. tuberculosis epidemic in the Republic of Georgia as comprehensively as possible. The sample set was collated by The Georgian National TB control program and processed at the National Center for Tuberculosis and Lung Diseases in Tbilisi (Georgia) for routine drug susceptibility testing. We were able to sample aproximately 70 % of all culture-confirmed MDR M. tuberculosis cases in Georgia in the timeframe.
Data collection	Samples and associated metadata were processed by the National Centre for Tuberculosis and Lung Diseases in Tbilisi, Georgia in the context of the Georgian National TB control program. The authors of the study were not involved in sample/metadata collection as it was performed in the context of the routine operation of the Georgian National TB control program. The samples were regrown and sent sent to the Swiss Tropical and Public Health Institute and subjected to whole genome sequencing at the Department of Biosystems Science and Engineering of the ETH Zürich.
Timing and spatial scale	The sample set consisted of a retrospective convenience sample of all stored MDR M. tuberculosis isolates collected between 2011 and 2016 by the Georgian National TB control program. The authors of the study did not have any influence on the collection of the samples and simply used what was available to maximize the sampling of the MDR TB epidemic in Georgia.
Data exclusions	A total of 450 strains had to be excluded for various reasons: 1. No MDR genotype, 2. failed whole genome sequencing, 3. potential mixed infections, 4. no isolation date available
Reproducibility	Samples were grown and sequenced once. If sequencing failed due to due to low DNA quality, re-culture and sequencing of the isolate in question was attempted once. The size of the dataset precluded the the growth and sequencing of each sample multiple times.
Randomization	Randomization not sensible/possible as the sample set consisted of a convenience sample aimed at maximizing the number of MDR M. tuberculosis isolates included in the study
Blinding	Blinding was not relevant as the sample set consisted of a retrospective convenience sample.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Sample set consisted of a retrospective convenience sample of MDR-TB isolates obtained from patients in Georgia aimed at maximizing the number of included samples. A summary of the population characteristics can be found in Table 1 in the main text.
Recruitment	No patients were specifically recruited for this study as it was based on samples collected by the Georgian National TB control program during its routine operation. The sample set therefore corresponds to retrospective convenience sample.
Ethics oversight	National Bioethics Council of the country of Georgia Ethikkommission Nordwest- und Zentralschweiz EKNZ, Switzerland

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