

A data compendium associating the genomes of 12,289 Mycobacterium tuberculosis isolates with quantitative resistance phenotypes to 13 antibiotics

Alice Brankin, Kerri M. Malone, Ivan Barilar, Simone Battaglia, Emanuele Borroni, Angela Pires Brandao, Andrea Maurizio Cabibbe, Joshua Carter, Darren Chetty, Daniela María Cirillo, et al.

▶ **To cite this version:**

Alice Brankin, Kerri M. Malone, Ivan Barilar, Simone Battaglia, Emanuele Borroni, et al.. A data compendium associating the genomes of 12,289 Mycobacterium tuberculosis isolates with quantitative resistance phenotypes to 13 antibiotics. PLoS Biology, 2022, 20 (8), pp.e3001721. 10.1371/journal.pbio.3001721 . hal-03837104

HAL Id: hal-03837104
https://hal.science/hal-03837104

Submitted on 3 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS BIOLOGY

METHODS AND RESOURCES

A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics

The CRyPTIC Consortium[¶]*

University of Oxford, Oxford, United Kingdom

¶ Membership of The CRyPTIC Consortium is provided in Text A of S1 File.
* alice.brankin@ndm.ox.ac.uk, kmalone@ebi.ac.uk, philip.fowler@ndm.ox.ac.uk, zi@ebi.ac.uk

Check for updates

OPEN ACCESS

Citation: The CRyPTIC Consortium (2022) A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics. PLoS Biol 20(8): e3001721. https://doi.org/10.1371/journal.pbio.3001721

Academic Editor: Jason Ladner, Northern Arizona University, UNITED STATES

Received: April 6, 2022

Accepted: June 21, 2022

Published: August 9, 2022

Copyright: © 2022 The CRyPTIC Consortium. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from ftp.ebi.ac.uk/pub/databases/cryptic/release_june2022/ The FTP site contains two top level directories: "reuse" and "reproducibility". All data for this study were analysed and visualised using either R or python3 libraries and packages. See github.com/kerrimalone/Brankin_Malone_2022 for codebase.

Funding: This work was supported by Wellcome Trust/Newton Fund-MRC Collaborative Award

Abstract

The Comprehensive Resistance Prediction for Tuberculosis: an International Consortium (CRyPTIC) presents here a data compendium of 12,289 *Mycobacterium tuberculosis* global clinical isolates, all of which have undergone whole-genome sequencing and have had their minimum inhibitory concentrations to 13 antitubercular drugs measured in a single assay. It is the largest matched phenotypic and genotypic dataset for *M. tuberculosis* to date. Here, we provide a summary detailing the breadth of data collected, along with a description of how the isolates were selected, collected, and uniformly processed in CRyPTIC partner laboratories across 23 countries. The compendium contains 6,814 isolates resistant to at least 1 drug, including 2,129 samples that fully satisfy the clinical definitions of rifampicin resistant (RR), multidrug resistant (MDR), pre-extensively drug resistant (pre-XDR), or extensively drug resistant (XDR). The data are enriched for rare resistance-associated variants, and the current limits of genotypic prediction of resistance status (sensitive/resistant) are presented by using a genetic mutation catalogue, along with the presence of suspected resistance-conferring mutations for isolates resistant to the newly introduced drugs bedaquiline, clofazimine, delamanid, and linezolid. Finally, a case study of rifampicin monoresistance demonstrates how this compendium could be used to advance our genetic understanding of rare resistance phenotypes. The data compendium is fully open source and it is hoped that it will facilitate and inspire future research for years to come.

Introduction

Tuberculosis (TB) is a curable and preventable disease; 85% of those afflicted can be successfully treated with a 6-month regimen. Despite this, TB is the world's top infectious disease killer (current SARS-CoV-2 pandemic excepted) with 10 million new cases and 1.2 million deaths estimated in 2019 alone [1]. Furthermore, drug-resistant TB (DR-TB; please see Table A in S1 File for a list of acronyms used throughout the manuscript) is a continual threat; almost half a million cases resistant to the first-line drug rifampicin (RR-TB) were estimated,

(200205/Z/15/Z); and Bill & Melinda Gates Foundation Trust (OPP1133541). Oxford CRyPTIC consortium members are funded/supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC), the views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health, and the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, a partnership between Public Health England and the University of Oxford, the views expressed are those of the authors and not necessarily those of the NIHR, Public Health England or the Department of Health and Social Care. J.M. is supported by the Wellcome Trust (203919/Z/16/Z). Z.Y. is supported by the National Science and Technology Major Project, China Grant No. 2018ZX10103001. K.M.M. is supported by EMBL's EIPOD3 programme funded by the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska Curie Actions. T.C.R. is funded in part by funding from Unitaid Grant No. 2019-32-FIND MDR. R.S.O. is supported by FAPESP Grant No. 17/16082-7. L.F. received financial support from FAPESP Grant No. 2012/51756-5. B.Z. is supported by the National Natural Science Foundation of China (81991534) and the Beijing Municipal Science & Technology Commission (Z201100005520041). N.T.T.T. is supported by the Wellcome Trust International Intermediate Fellowship (206724/Z/17/Z). G.T. is funded by the Wellcome Trust. R.W. is supported by the South African Medical Research Council. J.C. is supported by the Rhodes Trust and Stanford Medical Scientist Training Program (T32 GM007365). A.L. is supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Respiratory Infections at Imperial College London. S.G.L. is supported by the Fonds de Recherche en Santé du Québec. C.N. is funded by Wellcome Trust Grant No. 203583/Z/16/Z. A.V.R. is supported by Research Foundation Flanders (FWO) under Grant No. G0F8316N (FWO Odysseus). G.M. was supported by the Wellcome Trust (098316, 214321/Z/18/Z, and 203135/Z/16/Z), and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF) of South Africa (Grant No. 64787). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of this report. The opinions, findings and conclusions expressed in this manuscript reflect those of the authors alone. L.G. was supported by the Wellcome Trust (201470/Z/16/Z), the National Institute of Allergy

with three-quarters of these estimated to be multidrug-resistant (MDR-TB, resistant to first-line drugs isoniazid and rifampicin) [1]. Worryingly, only 44% of DR-TB cases were officially notified and just over half of these cases were successfully treated (57%) [1].

To address these issues, the World Health Organisation (WHO) is encouraging the development of better, faster, and more targeted diagnostic and treatment strategies through its EndTB campaign [1,2]. Of particular interest is universal drug susceptibility testing (DST). Conventionally, DST relies on lengthy (4 weeks minimum) culture-based methods that require strict biosafety conditions for *Mycobacterium tuberculosis*. The development of rapid genetics-based assays has decreased diagnostic time to as little as 2 hours through the detection of specific resistance conferring mutations, e.g., the Cepheid Xpert MTB/RIF test [3,4]. However, assay bias towards specific genic regions can result in misdiagnosis of resistance, the prescription of ineffective treatment regimens, and subsequent spread of MDR disease, as seen during an MDR outbreak in Eswatini [5–7]. Furthermore, detection of rifampicin resistance is used to infer MDR-TB epidemiologically as rifampicin resistance tends to coincide with resistance to isoniazid [8]. While this *modus operandi* is successful at pragmatically identifying potential MDR cases quickly and effectively, it is not generally true that a single path exists for developing MDR or extensively drug resistant TB (XDR = MDR/RR + resistance to at least 1 fluoroquinolone and either bedaquiline or linezolid).

Whole-genome sequencing (WGS) has the potential to reveal the entirety of the *M. tuberculosis* genetic resistance landscape for any number of drugs simultaneously while enabling a more rapid turnaround time and reduction in cost compared to culture-based DST methods [9]. However, the success of WGS as a diagnostic tool wholly depends on there being a comprehensive and accurate catalogue of resistance-conferring mutations for each drug. Recent advances have shown that genotypic predictions of resistance correlate well with DST measurements for first-line drugs [8]. However, the mechanisms of resistance to second-line drugs along with the new and repurposed drugs (NRDs) are less well understood despite their increased administration in clinics as MDR cases climb [1,10].

To address these shortcomings, the Comprehensive Resistance Prediction for Tuberculosis: an International Consortium (CRyPTIC) has collected *M. tuberculosis* clinical isolates worldwide to survey the genetic variation associated with resistance to 13 antitubercular drugs, specifically the first-line drugs rifampicin, isoniazid, and ethambutol; the second-line drugs amikacin, kanamycin, rifabutin, levofloxacin, moxifloxacin, and ethionamide; and the NRDs bedaquiline, clofazimine, delamanid, and linezolid. Here, we introduce and describe these data in the form of an open-access data compendium of 12,289 isolates, each of which has had its genomic sequence determined and DST profile measured [11]. This compendium is the largest drug screening effort to date for *M. tuberculosis* in a "one isolate–one microscale assay" format across defined compound concentration ranges. A sampling process was designed to enrich for resistant isolates to account for the variable prevalence of resistance found in different countries and the many rare resistance mutations for several drugs. As a result, the compendium is not suitable for measuring prevalence, or estimating "real-world" error rates of resistance prediction tools; rather, it serves as a resource to accelerate antimicrobial resistance (AMR) diagnostic development by enriching mutation catalogues for WGS resistance prediction, improving our understanding of the genetic mechanisms of resistance, and identifying important diagnostic gaps and drug resistance patterns. Indeed, the consortium has begun to address some of these important issues in recent publications using this very compendium [12–16].

and Infectious Diseases of the National Institutes of Health under award number 1R01AI146338, the GOSH Charity (VC0921) and the GOSH/ICH Biomedical Research Centre (www.nihr.ac.uk). A.B. is funded by the NDM Prize Studentship from the Oxford Medical Research Council Doctoral Training Partnership and the Nuffield Department of Clinical Medicine. D.J.W. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant No. 101237/Z/13/B) and by the Robertson Foundation. A.S.W. is an NIHR Senior Investigator. T.M.W. is a Wellcome Trust Clinical Career Development Fellow (214560/Z/18/Z). A.S.L. is supported by the Rhodes Trust. R.J.W. receives funding from the Francis Crick Institute which is supported by Wellcome Trust, (FC0010218), UKRI (FC0010218), and CRUK (FC0010218). T.C. has received grant funding and salary support from US NIH, CDC, USAID and Bill and Melinda Gates Foundation. The computational aspects of this research were supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC (BRC-1215-20008). Parts of the work were funded by the German Center of Infection Research (DZIF). The Scottish Mycobacteria Reference Laboratory is funded through National Services Scotland. The Wadsworth Center contributions were supported in part by Cooperative Agreement No. U60OE000103 funded by the Centers for Disease Control and Prevention through the Association of Public Health Laboratories and NIH/NIAID grant AI-117312. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: E.R. is employed by Public Health England and holds an honorary contract with Imperial College London. I.F.L. is Director of the Scottish Mycobacteria Reference Laboratory. S.N. receives funding from German Center for Infection Research, Excellenz Cluster Precision Medicine in Chronic Inflammation, Leibniz Science Campus Evolutionary Medicine of the LUNG (EvoLUNG)tion EXC 2167. P.S. is a consultant at Genoscreen. T.R. is funded by NIH and DoD and receives salary support from the non-profit organization FIND. T.R. is a co-founder, board member and shareholder of Verus Diagnostics Inc, a company that was founded with the intent of developing diagnostic assays. Verus Diagnostics was not involved in any way with data collection, analysis or publication of the results. T. R. has not received any financial support from Verus Diagnostics. UCSD Conflict of Interest office has reviewed and approved T.R.'s role in Verus

Results

The data compendium

The data compendium derives from a collection of 15,211 isolates for which both genomic and phenotypic data were collected by 23 of the partner countries across the continents of Asia, Africa, South America, and Europe. This collection will herein be referred to as the "full" dataset. An overview of the processing of the isolates is presented in Fig 1, and for a full description, please see Methods.

The 15, 211 isolates originated from 23 different countries (Fig 2). The largest numbers of isolates were contributed by India (n = 4,004), Peru (n = 2,691), South Africa (n = 2,155), Vietnam (n = 1,288), and China (n = 1,121).

Lineage assignment revealed that 99.7% of the 15,211 isolates belong to the 4 main *M. tuberculosis* lineages (L1 to L4). The pie charts in Fig B in S1 File show the proportion of isolates among the different lineages (Table D in S1 File) and sublineages (Table E in S1 File) for each location. Like previous studies, we see a strong association between geolocation and lineage (Pearson's chi-squared test, $p < 2.2 \times 10^{-16}$, Fig C in S1 File) [17,18]. The phylogenetic tree in Fig D in S1 File further highlights the strong population structure of this collection of isolates, with isolates clustering according to lineage. Typically undersampled in current databases and biobanks, the L3 isolates in this study represent the largest collection to date [19].

Although the 15,211 *M. tuberculosis* isolates belonging to the full dataset were plated to determine their minimum inhibitory concentrations (MICs) to 13 antitubercular drugs, regular quality assurance checks detected problems with plate inoculation and reading in 2 laboratories, resulting in the removal of 2,922 isolates. The data compendium therefore represents a total of 12,289 isolates with matched phenotypic and genotypic data for further analysis (Fig 1). Due to wells being skipped and other phenomena that prevent an MIC being measured, 88.1% of the isolates had a phenotype for all 13 drugs on the plate. For each drug, the number of isolates with an MIC measurement, and the associated quality of the reading, is presented in Table 1.

Resistance classification and distribution

Unsurprisingly, given its size and bias towards the collection of resistant isolates, resistance to each of the 13 drugs is represented within the compendium (Fig 3A). The drugs with the highest percentage of resistant isolates are the first-line drugs isoniazid and rifampicin (49.0% and 38.7%, respectively). Of the second-line drugs, levofloxacin had the highest proportion of resistant isolates in the dataset (17.6%) and amikacin the lowest (7.3%). A low proportion of isolates were resistant to the NRDs, bedaquiline (0.9%), clofazimine (4.4%), delamanid (1.6%), and linezolid (1.3%).

Of the 12,289 isolates represented by the compendium, 6,814 (55.4%) were resistant to at least 1 drug (Fig 3B). For the purpose of describing 5 broader resistance categories present in the dataset, we assumed that all MICs that could not be read had susceptible phenotypes. These 5 resistance categories comprise the following: isoniazid and rifampicin susceptible with resistance to another antitubercular drug, isoniazid resistant but rifampicin susceptible, RR/MDR, pre-XDR (RR/MDR + fluoroquinolone resistance), and XDR (RR/MDR + fluoroquinolone resistance + resistance to a group A agent: bedaquiline or linezolid). Consequently, the calculated prevalence of MDR, XDR, etc., in the dataset (Fig 3) are likely underestimates. Of the isolates resistant to one or more drugs, 22.8% were resistant to isoniazid and not rifampicin, 68.8% were either RR or MDR, and 8.4% were resistant to at least 1 antitubercular drug, but not isoniazid or rifampicin (Fig 3B). Of the RR/MDR isolates, 38.8% were pre-XDR and

Diagnostics Inc. T.R. is a co-inventor of a provisional patent for a TB diagnostic assay (provisional patent #: 63/048.989). T.R. is a co-inventor on a patent associated with the processing of TB sequencing data (European Patent Application No. 14840432.0 & USSN 14/912,918). T.R. has agreed to "donate all present and future interest in and rights to royalties from this patent" to UCSD to ensure that he does not receive any financial benefits from this patent. S.S. is working and holding ESOPs at HaystackAnalytics Pvt. Ltd. (Product: Using whole genome sequencing for drug susceptibility testing for Mycobacterium tuberculosis). G.F.G. is listed as an inventor on patent applications for RBD-dimer-based CoV vaccines. The patents for RBD-dimers as protein subunit vaccines for SARS-CoV-2 have been licensed to Anhui Zhifei Longcom Biopharmaceutical Co. Ltd, China.

Abbreviations: AMI, amikacin; AMR, antimicrobial resistance; AMyGDA, Automated Mycobacterial Growth Detection Algorithm; BDQ, bedaquiline; CFZ, clofazimine; CRyPTIC, Comprehensive Resistance Prediction for Tuberculosis: an International Consortium; CTAB, cetyltrimethylammonium bromide; DLM, delamanid; DR-TB, drug-resistant TB; DST, drug susceptibility testing; ECOFF, epidemiological cutoff; EMB, ethambutol; ETH, ethionamide; FN, false negative; FP, false positive; GCP, genotype confidence percentile; Hr-TB, isoniazid resistant and rifampicin susceptible; INH, isoniazid; KAN, kanamycin; LEV, levofloxacin; LZD, linezolid; MDR, multidrug resistant; ME, major error rate; MIC, minimum inhibitory concentration; MXF, moxifloxacin; NPV, negative predictive value; NRD, new and repurposed drug; PPV, positive predictive value; pre-XDR, pre-extensively drug resistant; RFB, rifabutin; RIF, rifampicin; RRDR, rifampicin resistance–determining region; RMR, rifampicin monoresistant; RR, rifampicin resistant; SNP, single nucleotide polymorphism; TB, tuberculosis; TDR, totally drug resistant; TN, true negative; TP, true positive; VCF, variant call format; VME, very major error rate; WGS, whole-genome sequencing; WHO, World Health Organisation; XDR, extensively drug resistant.

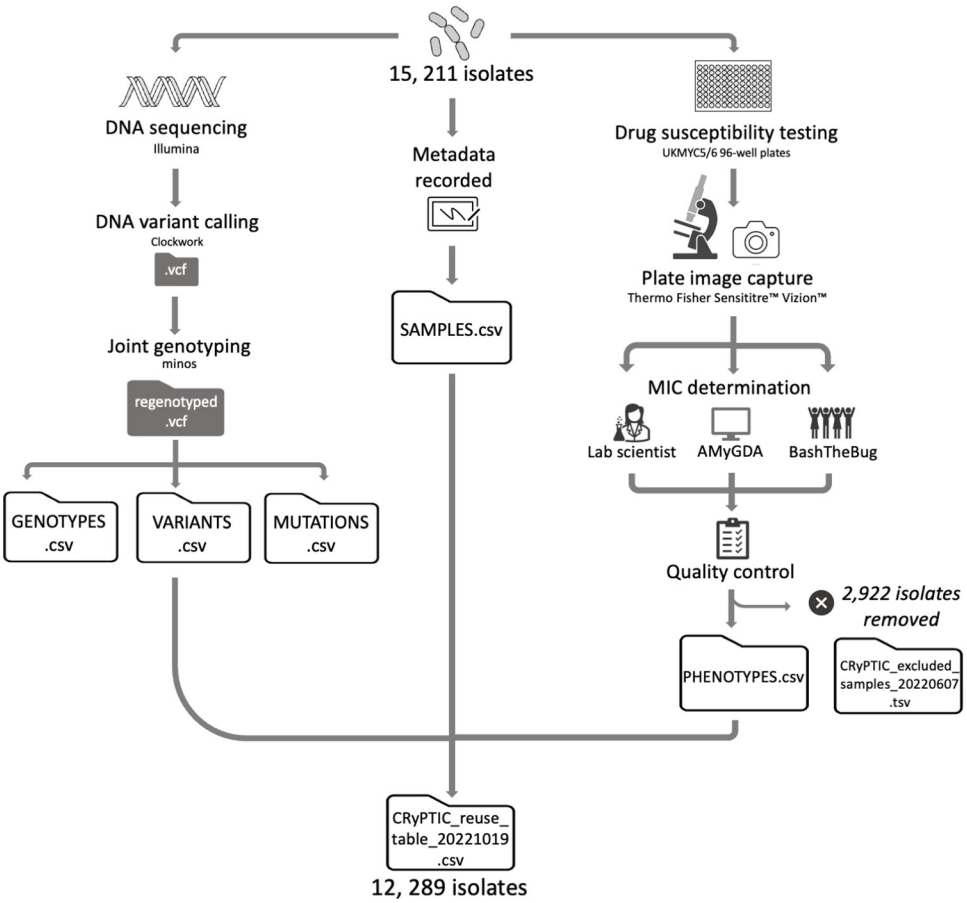


Fig 1. Processing sequencing and minimum inhibitory concentration data for 15,211 *Mycobacterium tuberculosis* isolates ("full" dataset). Briefly: Each isolate was DNA sequenced using an Illumina machine and plated onto 96-well plates (UKMYC5/6) containing 5–10× doubling dilutions of 13 antitubercular drugs for DST. Associated metadata (including country of origin and processing laboratory) was recorded. DNA variant calling and analysis was performed using Clockwork and Minos [47]). After 14 days, MIC measurements were taken by a trained scientist using Vizion, and the plate was photographed to also measure the MIC using the automated AMyGDA software and citizen scientists from BashTheBug [45]. After quality control procedures, phenotypic MIC data for 2,922 isolates were removed. The compendium therefore contains 15,211 isolates with WGS data ("full dataset"), 12,289 of which have matched quality assessed phenotypic data ("data compendium"). The raw sequence, VCFs, MICs, and binary resistance calls for the data compendium are presented in "CRyPTIC_reuse_table_20211019.csv" via an FTP site (see Methods), and the raw sequence and VCF files for those samples present in the full dataset are presented in "CRyPTIC_excluded_samples_20220607.tsv" via the same FTP site (see Methods). The data tables GENOTYPES.csv, VARIANTS.csv, MUTATIONS.csv, SAMPLES.csv, and PHENOTYPES.csv used for the analyses presented in this manuscript are also accessible via the FTP site (see Methods). DST, drug susceptibility testing; MIC, minimum inhibitory concentration; WGS, whole-genome sequencing.

https://doi.org/10.1371/journal.pbio.3001721.g001

3.0% were XDR. Two of the XDR isolates returned a resistant phenotype to all 13 of the drugs assayed (Table E in S1 File) and therefore could be reasonably described as totally drug resistant (TDR). One such isolate belonged to L4 and was contributed by South Africa, and the other belonged to L2 with an unknown country of origin contributed by Sweden.

The proportion of drug susceptible isolates collected differed between countries (Fig 3C, Methods). In countries that contributed more than 100 resistant isolates, each of the broad phenotypic resistance categories in Fig 3B were seen except for Peru, Vietnam, and Nepal, which did not contribute any XDR isolates (Fig 3C). Vietnam and Brazil sampled a high proportion of non-MDR/RR resistant phenotypes; 73.9% and 55.1% of resistant isolates

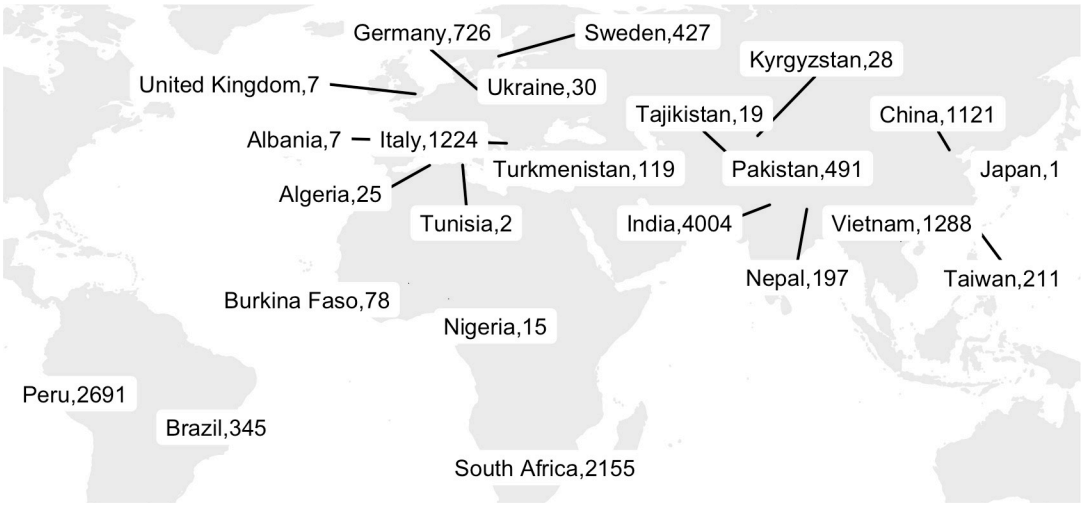


Fig 2. Geographical distribution of 15,211 CRyPTIC *Mycobacterium tuberculosis* clinical isolates ("full" dataset). The total number of isolates contributed by each country is depicted. Where the origin of an isolate was not known, the collection site identity was assigned (this occurred for 269 isolates in Germany, 17 isolates in India, 6 isolates in Peru, 885 isolates in Italy, 510 isolates in South Africa, 357 isolates in Sweden, 208 isolates in Taiwan, 1 isolate in Brazil, and 4 isolates in the UK). The base layer map was sourced from Natural Earth, which is in the public domain (see http://www.naturalearthdata.com/about/terms-of-use/), and created using the R packages "ggmap" and "maps" (see github.com/kerrimalone/Brankin_Malone_2022/).

https://doi.org/10.1371/journal.pbio.3001721.g002

contributed by these countries, respectively, were neither MDR nor RR. For Nepal and India, an especially high proportion of the MDR/RR isolates contributed were fluoroquinolone resistant (92.9% and 69.8%, respectively), which has been previously observed for this geographical region [1].

Table 1. Quality metrics for phenotype data.

	MIC measurements	HIGH QUALITY	MEDIUM QUALITY	LOW QUALITY
INH	12,070	9,519	1,351	1,200
RIF	12,099	8,955	1,356	1,788
EMB	12,158	7,506	1,355	3,297
LEV	12,163	7,774	1,354	3,035
MXF	12,194	6,785	1,353	4,056
AMI	12,072	8,973	1,350	1,749
KAN	12,130	9,333	1,355	1,442
BDQ	12,068	8,536	1,355	2,177
CFZ	12,049	7,763	1,352	2,934
DLM	11,927	8,095	1,349	2,483
LZD	12,189	7,141	1,355	3,693
ETH	12,132	8,821	1,355	1,956
RFB	12,150	10,042	1,352	756
TOTAL	**157,401**	**109,243**	**17,592**	**30,566**

Stated for each drug is the total number of MIC measurements stratified into "high" quality (at least 2 MIC measurement methods agree), "medium" quality (either Vizion and AMyGDA disagree, or the scientist recorded a MIC measurement using Vizion but did not store the plate picture), or "low" quality (all 3 MIC measurements methods disagree) phenotype classifications as described in Methods.

AMI, amikacin; BDQ, bedaquiline; CFZ, clofazimine; DLM, delamanid; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; LEV, levofloxacin; LZD, linezolid; MXF, moxifloxacin; RFB, rifabutin; RIF, rifampicin.

https://doi.org/10.1371/journal.pbio.3001721.t001

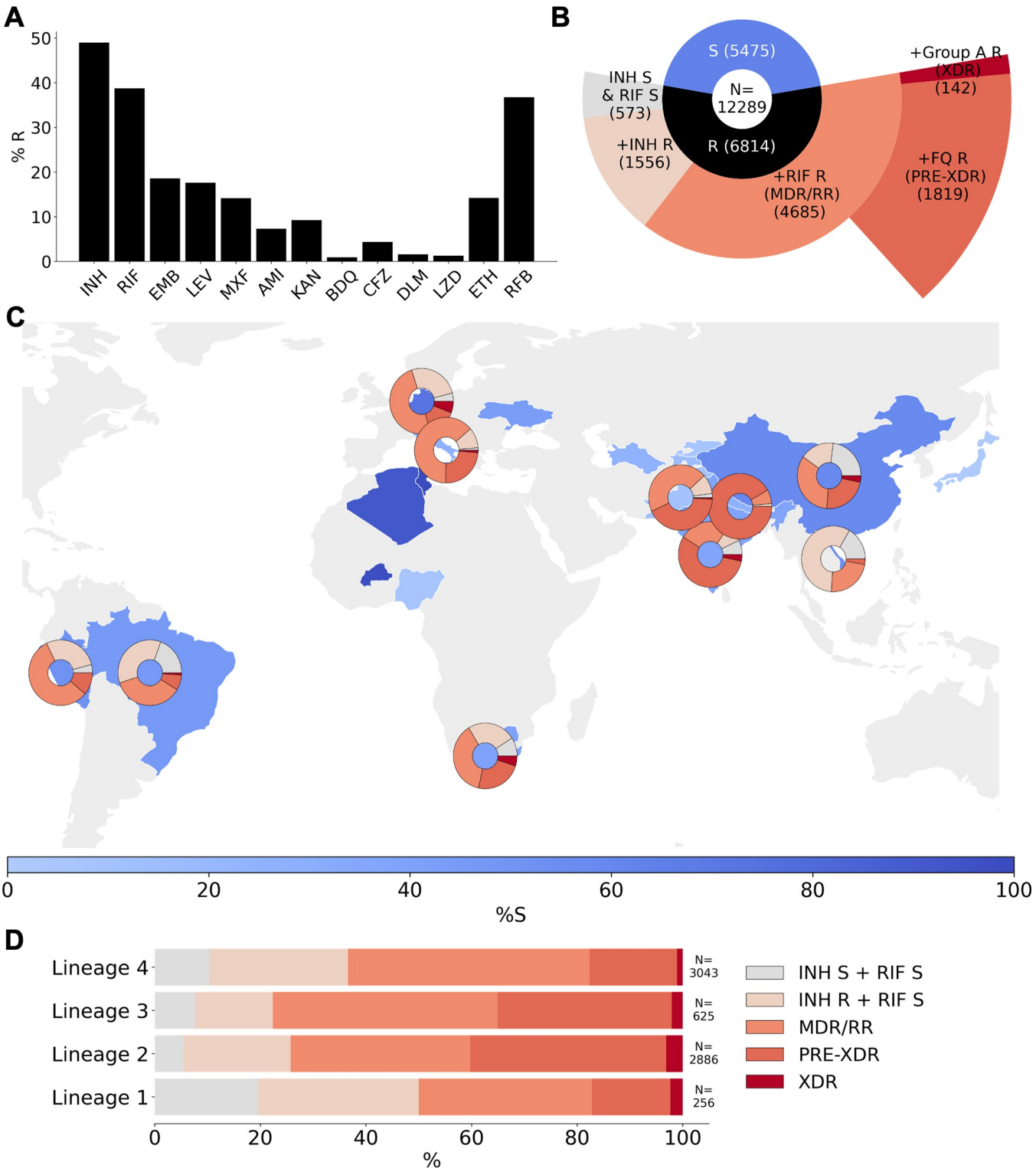


Fig 3. Drug phenotype data for the CRyPTIC compendium. (A) Frequency of resistance to each of 13 drugs in the data compendium. The total number of isolates with a binary phenotype (of any quality) for the corresponding drug is presented in Table 1. **(B)** Phenotypes of the 12,289 isolates with a binary phenotype for at least 1 drug. **(C)** Geographical distribution of phenotypes of 12,289 compendium isolates. Intensity of blue shows the percentage of isolates contributed that were categorised as susceptible to all 13 drugs ("%S"). Donut plots show the proportions of resistant phenotypes identified in (B) for countries contributing > = 100 isolates with drug resistance. **(D)** Proportions of resistance phenotypes in the 4 major *Mycobacterium tuberculosis* lineages. N is the number of isolates of the lineage called resistant to at least one of the 13 drugs. The base layer map in (C) was sourced from Natural Earth, which is in the public

domain (see http://www.naturalearthdata.com/about/terms-of-use/), and created using the python library "geopandas" (see github.com/kerrimalone/Brankin_Malone_2022/). AMI, amikacin; BDQ, bedaquiline; CFZ, clofazimine; CRyPTIC, Comprehensive Resistance Prediction for Tuberculosis: an International Consortium; DLM, delamanid; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; LEV, levofloxacin; LZD, linezolid; MDR, multidrug resistant; MXF, moxifloxacin; pre-XDR, pre-extensively drug resistant; RFB, rifabutin; RIF, rifampicin; RR, rifampicin resistant; XDR, extensively drug resistant.

https://doi.org/10.1371/journal.pbio.3001721.g003

Of the 6,814 resistant isolates, 256 were from L1, 2,886 were from L2, 625 were from L3, and 3,043 were from L4. All 5 broader categories of resistance were represented in the 4 major *M. tuberculosis* lineages (Fig 3D). We note that the relative proportions of resistance categories will have been influenced by the different local sampling approaches since lineage distributions are typically geographically distinct (Fig B in S1 File). Bearing this in mind, we observe that in the compendium, L3 isolates contained the most MDR/RR isolates as a proportion of resistant isolates (77.6%), L2 isolates contained the most pre-XDR isolates as a proportion of MDR/RR isolates (54.2%), and L2 contained the most XDR isolates as a proportion of MDR/RR isolates (4.7%).

Co-occurrence of drug resistance among the compendium isolates

As we measured MICs to 13 drugs in parallel, we can ask whether, and how often, co-occurrence of drug resistance occurs among the isolates. We found isolates with all possible 2-drug resistant combinations in this dataset (Fig 4A and Table G in S1 File). With the exception of correlations between drugs in the same class (rifabutin versus rifampicin, moxifloxacin versus levofloxacin), isoniazid resistance was the most strongly associated with resistance to each of the other drugs. Resistance to any of the drugs was also strongly associated with resistance to rifampicin. Of the second-line drugs, levofloxacin and moxifloxacin were more commonly seen as a second resistant phenotype than the injectable drugs kanamycin and amikacin.

Resistance to both drugs in the aminoglycoside class was common in the dataset; 90.4% of amikacin-resistant isolates were also resistant to kanamycin, although significantly fewer kanamycin-resistant isolates were resistant to amikacin (72.0%, $p < 0.00001$) (Fig 4A). In a similar fashion, a smaller proportion of RR isolates were resistant to rifabutin than rifabutin-resistant isolates that were resistant to rifampicin (91.3%, 96.8%, $p < 0.00001$), while a smaller proportion of levofloxacin-resistant isolates were resistant to moxifloxacin than moxifloxacin-resistant isolates that were resistant to levofloxacin (78.5%, 97.6%, $p < 0.00001$). Differences in drugs of the same class are also well documented by in vitro studies [20–22].

Isolates resistant to the NRDs bedaquiline, clofazimine, delamanid, and linezolid were most likely to also be resistant to isoniazid, followed by rifampicin and rifabutin. The NRDs were less commonly seen as a second resistance phenotype and the smallest proportional resistance combinations involved the NRDs (e.g., 1.5% of isoniazid-resistant isolates were bedaquiline resistant). Within the NRDs, however, co-occurrence of resistance was proportionally higher; bedaquiline, linezolid, and delamanid resistance was commonly seen with clofazimine resistance (52.4%, 34.2%, and 26.3% of isolates having coresistance with clofazimine, respectively).

Additional antibiotic resistance in isolates with non-MDR or MDR phenotypic backgrounds

To further investigate drug resistance patterns among the isolates, we examined in more detail the correlation structure of phenotypes by conditioning on different resistance backgrounds including isoniazid and rifampicin susceptible, isoniazid resistant and rifampicin susceptible, rifampicin resistant and isoniazid susceptible, MDR, pre-XDR, and XDR (Fig 4B–4F). We found that a greater proportion of isolates that were susceptible to isoniazid and rifampicin

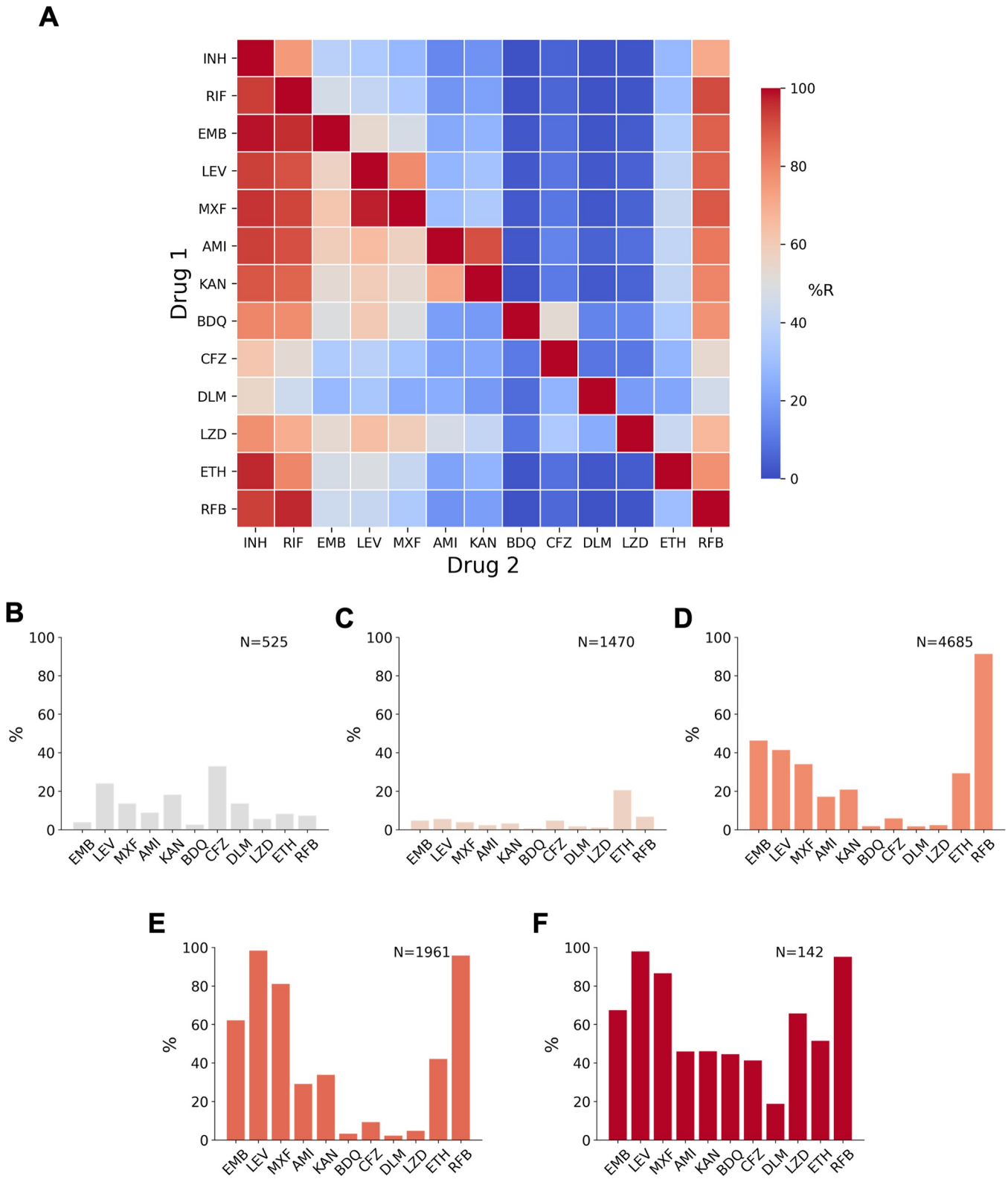


Fig 4. Co-occurrence of resistance to 1 drug conditional on resistance to another drug, or to resistance background. (A) The heatmap shows the probability of an isolate being resistant to Drug 2 if it is resistant to Drug 1, percentages are given in Table F in S1 File. **(B-F)** Percentage of isolates that are resistant to another of the 13 drugs in a background of **(B)** isoniazid susceptible + rifampicin susceptible (but resistant to at least one other antitubercular drug), **(C)** isoniazid resistant + rifampicin susceptible, **(D)** MDR/RR, **(E)** Pre-XDR, and **(F)** XDR. Only samples with definite phenotypes for RIF in MDR

backgrounds and RIF and INH in non-MDR backgrounds and the additional drug are included. AMI, amikacin; BDQ, bedaquiline; CFZ, clofazimine; CRyPTIC, Comprehensive Resistance Prediction for Tuberculosis: an International Consortium; DLM, delamanid; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; LEV, levofloxacin; LZD, linezolid; MDR, multidrug resistant (resistant to first-line drugs isoniazid and rifampicin); MXF, moxifloxacin; pre-XDR, pre-extensively drug resistant (MDR/RR + fluoroquinolone resistant); RFB, rifabutin; RIF, rifampicin; RR, rifampicin resistant; XDR, extensively drug resistant (MDR/RR + resistant to at least 1 fluoroquinolone and either bedaquiline or linezolid).

https://doi.org/10.1371/journal.pbio.3001721.g004

were resistant to the second-line drugs levofloxacin (24.1%), kanamycin (18.1%), moxifloxacin (13.7%), and amikacin (8.9%) than the first-line drug ethambutol (3.8%) (Fig 4B). The proportion of isolates resistant to clofazimine or levofloxacin was particularly high (32.9% and 24.1%, respectively), and more isolates were resistant to these 2 drugs than ethambutol in an isoniazid-resistant and rifampicin-susceptible background but not in MDR/RR isolates (Fig 4C–4F).

MDR/RR isolates were most commonly resistant (excluding rifabutin) to the first-line drug ethambutol (46.3%), closely followed by levofloxacin (41.4%). As expected, the proportion of fluoroquinolone resistance was higher in MDR/RR isolates than non-MDR isolates [23], and we found a greater proportion of isolates were resistant to levofloxacin than moxifloxacin, a pattern seen in all other backgrounds (Fig 4C–4F). For the aminoglycosides, a greater percentage of MDR/RR isolates were kanamycin resistant (21.8%) than amikacin resistant (18.1%), a trend seen in all other backgrounds.

For isolates with an XDR phenotype, a higher proportion were resistant to linezolid than bedaquiline (66.7% compared to 44.6%) and 11.3% of XDR isolates were resistant to both bedaquiline and linezolid (Fig 4F). XDR isolates were also resistant to the other NRDs, clofazimine (41.3%) and delamanid (18.8%). In non-XDR backgrounds, the most common NRD resistance seen was clofazimine (Fig 4B–4E).

Genetic-based predictions of resistance

The compendium was sampled in order to collect as broad as possible a representation of resistance mutations and mechanisms—it is therefore expected to be enriched for rare resistance mutations compared with a prospectively sampled dataset from 1 country. To establish a baseline measure of how much unexplained resistance there was in the compendium, based on a standard catalogue of resistance mutations, we compared genetic-based predictions of susceptibility and resistance to the binary phenotypes derived from MICs for 8 drugs and for all isolates in this compendium (FN and VME columns in Table 2). Since these data are enriched for resistance, the calculated error rates are not representative of how well such a method would perform in routine clinical use. The results were broadly in line with prior measurements on a smaller (independent) set [24]. The hybrid catalogue used in this study does not make predictions for rifabutin, linezolid, bedaquiline, delamanid, or clofazimine; indeed, this is one of the main aims of the consortium and new catalogues published by CRyPTIC, and WHO will begin to address this shortcoming (Fig 5) [25].

Table 3 shows the top mutations found among isolates phenotypically resistant to first- and second-line drugs. As expected, *rpoB* S450L was the most frequent mutation associated with rifampicin resistance, and *katG* S315T was the most common mutation associated with isoniazid resistance. Mutations in *gyrA* dominate among fluoroquinolone-resistant isolates; D94G and A90V are the 2 most frequently occurring mutations for levofloxacin and moxifloxacin.

Resistance to new and repurposed drugs

As previously stated, relatively few isolates are resistant to the NRDs, bedaquiline ($n = 109$), clofazimine ($n = 525$), delamanid ($n = 186$), and linezolid ($n = 156$). South Africa contributed the greatest number of isolates resistant to bedaquiline, clofazimine, and linezolid (Fig 5A),

Table 2. Predicting phenotypic resistance using genetics.

	TP	FP	TN	FN	VME	ME	PPV	NPV
INH	5,493	142	5,622	224	0.039	0.025	0.961	0.975
RIF	4,535	435	6,669	107	0.023	0.061	0.977	0.939
EMB	1,919	513	6,702	111	0.055	0.071	0.945	0.929
LEV	1,689	255	8,104	184	0.098	0.031	0.902	0.969
MXF	1,358	504	9,022	160	0.105	0.053	0.895	0.947
AMI	632	84	10,117	163	0.205	0.008	0.795	0.992
KAN	735	124	9,043	197	0.211	0.014	0.789	0.986
ETH	971	114	9,183	511	0.345	0.012	0.655	0.988

Statistics on how much resistance can be explained in a dataset enriched for rare resistance mutations using a standard resistance catalogue that predates the CRyPTIC project. TP, the number of phenotypically resistant samples that are correctly identified as resistant ("true positives"); FP, the number of phenotypically susceptible samples that are falsely identified as resistant ("false positives"); TN, the number of phenotypically susceptible samples that are correctly identified as susceptible ("true negatives"); FN, the number of phenotypically resistant samples that are incorrectly identified as susceptible ("false negative"); VME, very major error rate (false-negative rate), 0–1; ME, major error rate (false-positive rate), 0–1; PPV, positive predictive value, 0–1; NPV, negative predictive value, 0–1.
AMI, amikacin; EMB, ethambutol; ETH, ethionamide; INH, isoniazid; KAN, kanamycin; LEV, levofloxacin; MXF, moxifloxacin; RIF, rifampicin.

https://doi.org/10.1371/journal.pbio.3001721.t002

while China and India contributed the most isolates resistant to delamanid. Since the collection protocol differed between laboratories, it is not possible to infer any differences in the relative prevalence of resistance to the NRDs in these countries. The results of a survey of all nonsynonymous mutations in genes known or suspected to be involved in resistance to these 4 drugs (*rv0678*, *mmpL5*, *pepQ*, *ddn*, *rplC*, *rrl*, etc.) are depicted in Fig 5B–5E [27–31]. Mutations known to be associated with sensitivity were ignored, along with mutations that occurred at a frequency of $\geq$5% among all isolates (as 0.05% of total isolates are resistant to the NRDs). In contrast to first- and second-line drugs, there are no mutations within a single gene/small group of genes that can fully explain resistance to any NRD, and indeed no single gene (if there were, they would be visible as complete rings around the phylotrees in Fig 5). Note that the role of most of these mutations in resistance remains undetermined.

Case study on rifampicin monoresistance

Around 1% of TB cases are rifampicin monoresistant (RMR), and the frequency is increasing [1,32]. The WHO does not recommend isoniazid for RMR treatment, despite it being effective; this is likely due to the reliance on assays such as Xpert MTB/RIF, which cannot distinguish between RMR and MDR. Use of isoniazid could improve treatment outcomes for RMR patients, which are currently similar to that of MDR TB, including a higher risk of death compared to drug-susceptible infections [33,34]. Due to its low natural prevalence, RMR has been poorly studied to date, but increasingly large clinical TB datasets, such as the one presented here, make its study now feasible.

For this case study, we defined RMR as any isolate that is rifampicin resistant and isoniazid susceptible, and discounted isolates with no definite phenotype for either drug. Of the 4,655 RR isolates in the compendium that also had a phenotype for isoniazid, 302 (6.5%) were RMR. These isolates were contributed by 12 different countries, and we found South African and Nigerian contributions had a significantly higher proportion of RMR isolates than that of the total dataset at 17.5% ($p < 0.00001$) and 27.3% ($p = 0.00534$), respectively (Fig 6A), compared with 6.5% for the total dataset. We note that these proportions are influenced by sampling strategies, but the higher contribution of RMR isolates from South Africa is consistent with previous studies [32].

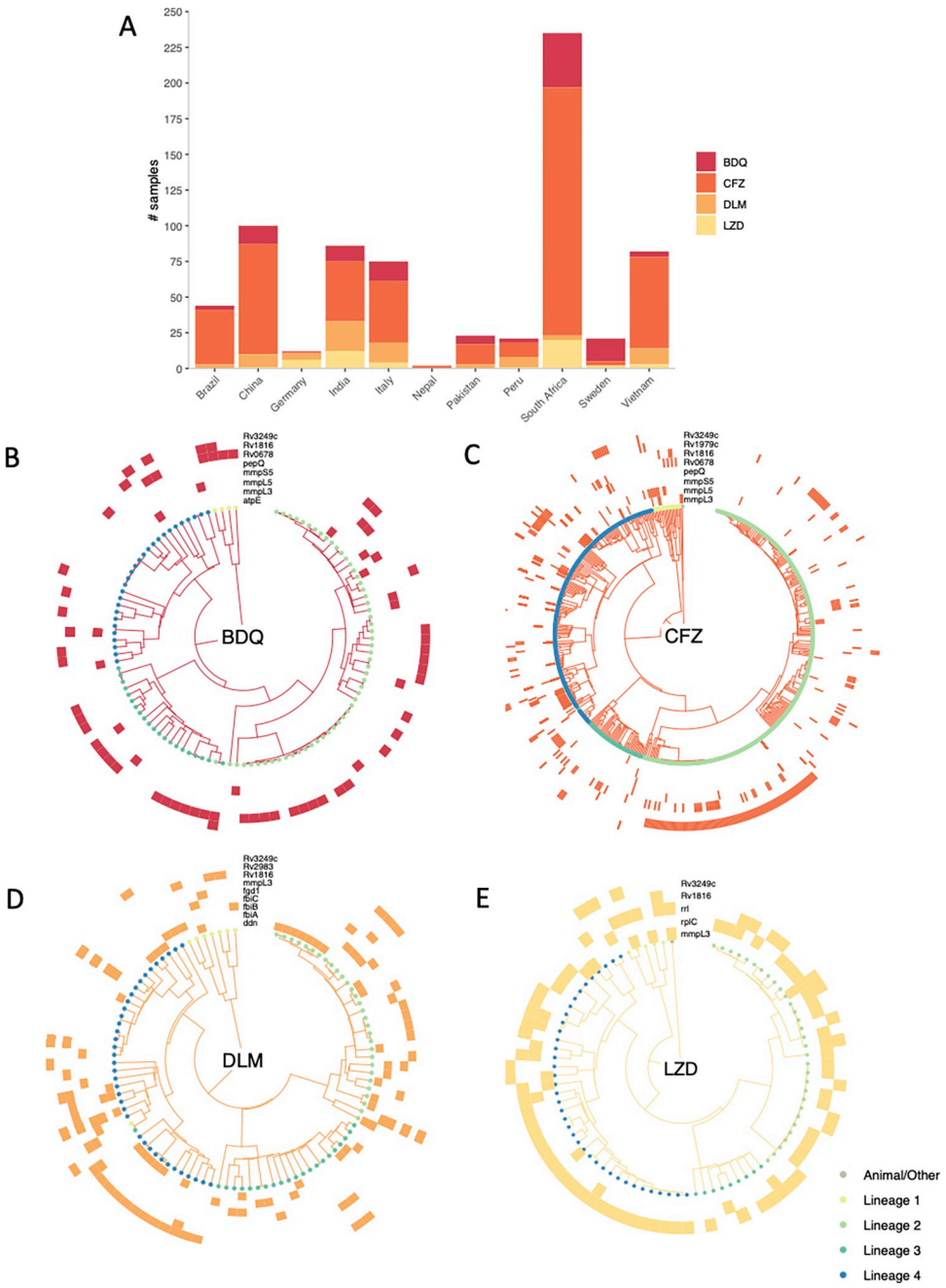


Fig 5. Resistance to bedaquiline, clofazimine, delamanid, and linezolid among *Mycobacterium tuberculosis* compendium isolates. (A) The prevalence (within these data) of resistance to BDQ, CFZ, DLM, and LZD per country or origin or collection site. Phylotrees are shown for isolates phenotypically resistant to **(B)** BDQ, **(C)** CFZ, **(D)** DLM, and **(E)** LZD. Tip point colours denote lineage. Each outer track represents a gene thought to be associated with resistance and coloured blocks denote the presence of a nonsynonymous mutation in the relevant gene for a given isolate. Mutations in these genes that are either associated with sensitivity or present in >5% of the collection of isolates as a whole were ignored. BDQ, bedaquiline; CFZ, clofazimine; DLM, delamanid; LZD, linezolid.

https://doi.org/10.1371/journal.pbio.3001721.g005

Rifampicin monoresistance is incorrectly predicted by current diagnostics

A widely used, WHO-endorsed diagnostic tool, the Xpert MTB/RIF assay, uses a proxy whereby any SNP detected in the "rifampicin-resistance determining region" (RRDR) of *rpoB*

Table 3. The top mutations associated with phenotypic drug resistance.

	GENE	VARIANT	n	%
Rifampicin	*rpoB*	S450L	2,914	62.2
	rpoB	D435V	506	10.8
	rpoB	H445D	202	4.3
	rpoB	H445Y	147	3.1
	rpoB	D435Y	112	2.4
Isoniazid	*katG*	S315T	3,668	70.0
	fabG1	c-15t	829	15.8
	fabG1	g-17t	176	3.4
	fabG1	t-8c	154	2.9
	inhA	I194T	56	1.1
Ethambutol	*embB*	M306V	1,131	50.0
	embB	M306I	1,001	44.3
	embB	Q497R	449	19.9
	embB	G406A	164	7.2
	embB	G406D	105	4.6
Kanamycin	*rrs*	a1401g	660	58.9
	eis	c-14t	70	6.2
	eis	g-10a	53	4.7
Amikacin	*rrs*	a1401g	660	74.7
	rrs	g1484t	7	0.8
Levofloxacin	*gyrA*	D94G	783	36.5
	gyrA	A90V	487	22.7
	gyrA	D94N	157	7.3
	gyrA	D94A	133	6.2
	gyrA	S91P	92	4.3
Moxifloxacin	*gyrA*	D94G	783	45.4
	gyrA	A90V	487	28.2
	gyrA	D94N	157	9.1
	gyrA	D94A	133	7.7
	gyrA	D94Y	70	4.1
Ethionamide	*fabG1*	c-15t	829	48.0
	fabG1	L203L	124	7.2

Depicted is a survey of the resistance-associated mutations present in the data compendium [8,26]. "VARIANT": nonsynonymous amino acid mutations are denoted by upper case letters, while nucleotide substitutions for noncoding sequences are denoted by lower case letters. Negative numbers denote substitutions in promoter regions; "GENE": genic region of interest in which "Variant" can be found; "*n*": number of phenotypically resistant isolates with "VARIANT"; "%": percentage of total phenotypically resistant isolates with "VARIANT".

https://doi.org/10.1371/journal.pbio.3001721.t003

results in a prediction of MDR. However, the suitability of the proxy is dependent upon prevalence of RMR in the population [32]. We tested the reliability of this on the 4,655 RR isolates in our dataset that had a phenotype for isoniazid (Fig 6B).

Of these isolates, 4,353 (93.5%) were MDR and 302 (6.5%) were RMR. Of the MDR isolates, 187 had no RRDR mutation, and therefore, 4.0% of isolates in this study would be predicted as false negative MDR by the Xpert MTB/RIF assay. Of the RMR isolates, 276 had a mutation in the RRDR of *rpoB* and so the Xpert MTB/RIF assay proxy would incorrectly predict 5.9% of the RR isolates as false positive MDR cases. However, overall, the Xpert MTB/RIF assay proxy correctly predicts 89.5% of the RR isolates as MDR and 0.6% of the isolates as non-MDR in

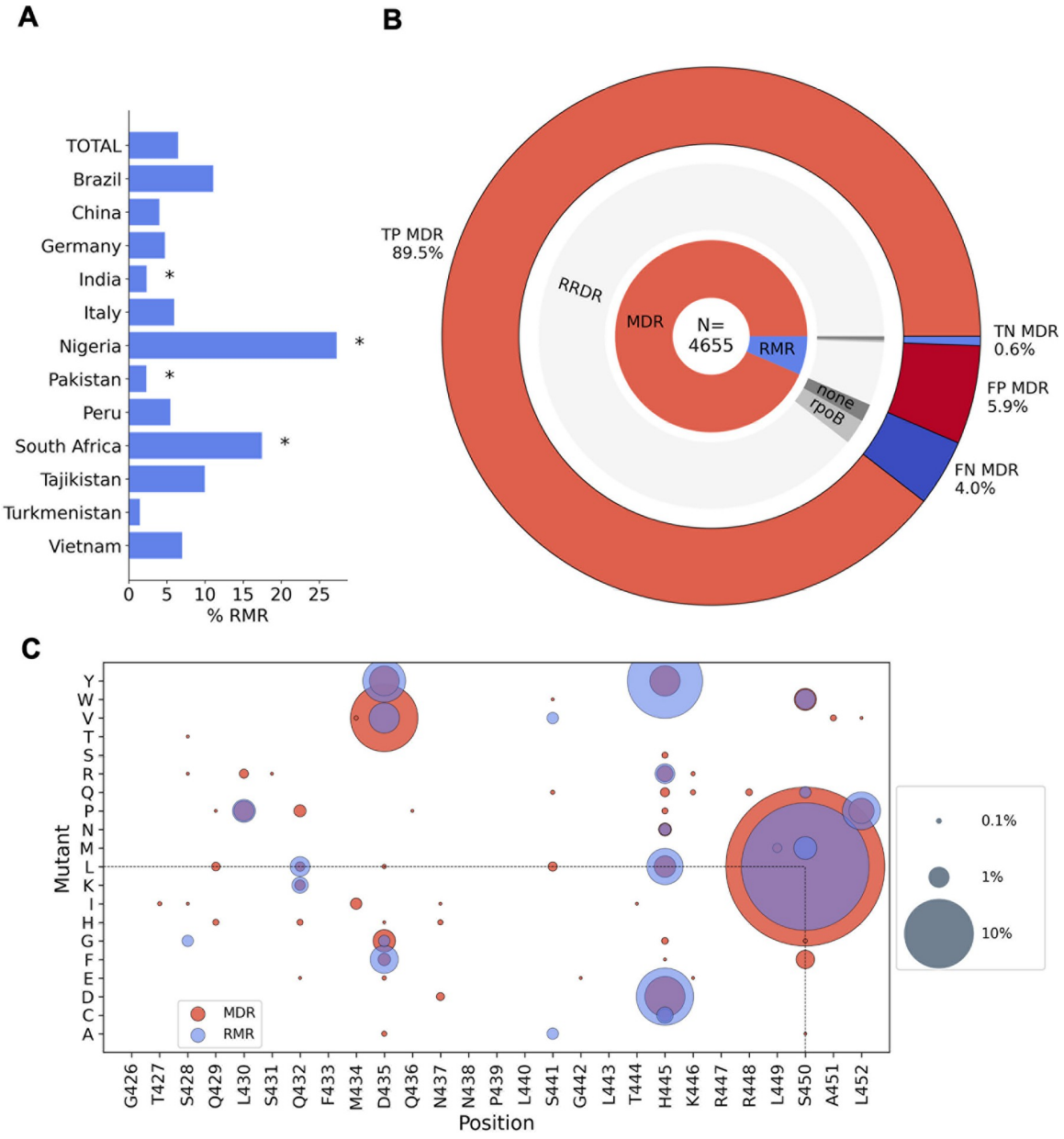


Fig 6. Rifampicin monoresistance. (A) Percentage of RR isolates that are RMR by country of isolate origin. * indicates RMR proportions that were significantly different from that of the total dataset using a 2-tailed z-test with 95% confidence. **(B)** MDR predictions for RR isolates made using the Xpert MTB/RIF assay proxy. N is the total number of RR isolates. The inner ring shows the proportion of RR isolates that are RMR and MDR. The middle ring represents the proportions of RMR and MDR isolates that have an SNP (synonymous or nonsynonymous) in the RRDR of *rpoB* ("RRDR"), no RRDR SNP but a SNP elsewhere in the *rpoB* gene ("*rpoB*"), and no *rpoB* mutations ("none"). The outer ring shows the expected TP, TN, FP, and FN MDR predictions of Xpert MTB/RIF assay, based on the SNPs present in the RR isolates. **(C)** Nonsynonymous mutations found in the RRDR of *rpoB* in RMR isolates and MDR isolates. Presence of a coloured spot indicates that the mutation was found in RMR/MDR isolates, and spot size corresponds to the proportion of RMR or MDR isolates carrying that mutation. FN, false negative; FP, false positive; MDR, multidrug resistant; RRDR, rifampicin resistance–determining region; RMR, rifampicin monoresistant; RR, rifampicin resistant; SNP, single nucleotide polymorphism; TN, true negative; TP, true positive.

https://doi.org/10.1371/journal.pbio.3001721.g006

this dataset, which suggests it is a reasonably successful diagnostic tool with >90% accuracy for MDR classification of RR isolates. As our dataset is oversampled for resistance, it likely contains a higher prevalence of RMR than the global average, and, hence, the Xpert MTB/RIF

assay is likely to perform better on more representative data. However, the analysis shows how the increasing global levels of RMR TB cases could increase the number of false positive MDR diagnoses by the Xpert MTB/RIF assay, denying isoniazid treatment to a greater number of patients who would then be moved on to less effective drugs.

There are genetic differences between rifampicin monoresistant and multidrug resistant isolates

We have analysed the matched phenotypic and genotypic data of the compendium to examine whether there were any differences in the genetic determinants of rifampicin resistance between RMR and MDR isolates as was seen in a recent study of South African isolates [35]. The proportion of RMR isolates with no *rpoB* mutation (5.3%; Fig 6B) was significantly higher than that of MDR isolates (1.8%, $p < 0.00001$). This suggests that non-target-mediated resistance mechanisms, such as up-regulation of rifampicin-specific efflux pumps, could be more influential in providing protection against rifampicin in RMR isolates than in MDR isolates.

The majority of RMR and MDR isolates contained one or more single nucleotide polymorphisms (SNPs) in *rpoB*, with the most having at least 1 mutation in the RRDR. To date, several nonsynonymous RRDR mutations have been found in RMR *M. tuberculosis* isolates, including the resistance conferring mutations S450L, H445D, and D435Y, which are also seen in MDR isolates [36,37]. For both RMR and MDR isolates in this dataset, the most common *rpoB* RRDR mutation seen was S450L (63.6% and 41.1% of isolates, respectively; Fig 6C). Five mutations were present in RMR isolates that were not seen in MDR isolates: S428G, S441A, S441V, S450M, and S450Q; however, these were seen at low prevalence (<2%) of RMR isolates. We found more RMR isolates had His445 mutated than MDR isolates (27.8% of RMR and 9.5% of MDR, $p < 0.00001$), and mutations at Ser450 and Asp435 were more prevalent in MDR isolates than RMR isolates (43.7% of RMR and 65.8% of MDR ($p < 0.00001$), and 9.3% of RMR and 15.5% of MDR ($p = 0.00328$), respectively).

In RMR isolates, we observed 27 different *rpoB* mutations that fall outside the RRDR; 11 were found in RMR but not MDR isolates and all were seen at <2% prevalence (Fig E in S1 File). The most common non-RRDR mutation in both MDR and RMR isolates was a cytosine to thymine mutation 61 bases upstream of the *rpoB* start codon (10.1% and 8.6% of isolates, respectively). The resistance conferring mutations, *rpoB* I491F, V695L, and V170F, were seen at low proportions (<2% of isolates) with no significant difference between MDR and RMR isolates.

Discussion

This compendium of *M. tuberculosis* clinical isolates is the result of an extensive global effort by the CRyPTIC consortium to better map the genetic variation associated with drug resistance. Through its sheer size and by oversampling for resistance, the compendium gives an unparalleled view of resistance and resistance patterns among the panel of 13 antitubercular compounds studied. This study serves to summarise the data within the compendium and to highlight the existence of the open-access resource to the wider community to help better inform future treatment guidelines and steer the development of improved diagnostics.

Starting with first-line drugs, molecular based diagnostic assays have vastly improved the detection of and the speed at which we find DR-TB cases, resulting in improved quality of care for patients. However, relying solely on these diagnostic methods has several drawbacks. Aside from the Xpert MTB/RIF assay potentially increasing false positive MDR diagnoses as discussed earlier in the RMR case study, the assay assumes isoniazid resistance upon detection of rifampicin resistance. Thus, less is known about the prevalence of mono-isoniazid resistance

or "true" cases of MDR (confirmed rifampicin and isoniazid resistance) [1] and with large datasets such as this compendium, we can further investigate these important and rarer clinical phenotypes (like that of RMR in our case study). Another example of a rarer phenotype is that of isoniazid-resistant and rifampicin-susceptible (Hr-TB) isolates; a greater number of these were contributed by CRyPTIC countries than RMR isolates (n = 1,470 versus n = 302), a pattern also recently observed in a global prevalence study [38]. A modified 6-month treatment regimen is now recommended for Hr-TB (rifampicin, ethambutol, levofloxacin, and pyrazinamide), and as a result of inadequate diagnosis, many of the 1.4 million global Hr-TB estimated cases would have received inadequate and unnecessarily longer treatment regimens [1,39]. Encouragingly, CRyPTIC isolates with an Hr-TB background exhibited relatively low levels of resistance to other antitubercular drugs, including those in the augmented regimen (Fig 4C). However, without appropriate tools to assess and survey this, we will continue to misdiagnose and ineffectively treat these clinical cases. In 2018, CRyPTIC and the 100,000 Genomes project demonstrated that genotypic prediction from WGS correlates well with culture-based phenotype for first-line drugs, which is reflected in our summary of the genetic catalogue applied to this dataset (Table 3) [8]. While predictions can be made to a high level of sensitivity and specificity, there is still more to learn, as exemplified by the isolates in the compendium that despite being resistant to rifampicin and isoniazid could not be described genetically (Table 2). This shortfall, along with the limitations of molecular based diagnostic assays, highlights the need for continual genetic surveillance and shines a favourable light on a WGS-led approach.

A strength of this compendium lies with the data collated for second-line drugs. A greater proportion of drug-resistant isolates had additional resistance to fluoroquinolones than second-line injectable drugs (Fig 4A). This could be due to more widespread use of fluoroquinolones as well as their ease of administration and hence them being recommended over injectables for longer MDR treatment regimens [1]. Concerningly, we found that resistance to levofloxacin and moxifloxacin, and kanamycin and amikacin, were more common than resistance to the mycobacterial specific drug ethambutol in an isoniazid- and rifampicin-susceptible background (Fig 4B), suggesting a level of preexisting resistance to second-line drugs. This concurs with a systematic review that found patients previously prescribed fluoroquinolones were 3 times more likely to have fluoroquinolone-resistant TB [40]. Careful stewardship of fluoroquinolones, both in TB and other infectious diseases, will be paramount for the success of treatment regimens. Despite variability in sample collection, we observed high proportions of fluoroquinolone-resistant MDR/RR isolates from some countries and therefore suggest that MDR treatment regimens could be improved by optimisation on a geographic basis.

Further treatment improvement could also be made by the selection of appropriate drugs from each class. For example, The WHO recommends switching from kanamycin to amikacin when treating MDR TB patients [39], and the compendium supports this recommendation as we saw more resistance to kanamycin than amikacin in all phenotypic backgrounds. For fluoroquinolones, more isolates were resistant to levofloxacin than moxifloxacin in all phenotypic backgrounds, suggesting moxifloxacin may by the most appropriate fluoroquinolone to recommend, although we note this conclusion is critically dependent on the validity of the cutoff, here an ECOFF, used to infer resistance. However, the amenability of drugs to catalogue-based genetic diagnostics is also an important consideration, and our data suggest levofloxacin resistance could be predicted more reliably than moxifloxacin, with fewer false positives predicted (Table 2). Testing for fluoroquinolone resistance using molecular diagnostic tests remains limited. Global data from the past 15 years suggest that the proportion of MDR/RR TB cases resistant to fluoroquinolones sits at around 20%, with these cases primarily found in regions of high MDR-TB burden [1]. While recently approved tools, such as the Cepheid Xpert MTB/XDR cartridge, will permit both isoniazid and fluoroquinolone testing to be increased, the

same pitfalls are to be encountered regarding targeted diagnostic assays [41]. In contrast, the genetic survey in this study demonstrates the potential of WGS for genetic prediction of resistance to second-line drugs, and studies within the consortium to investigate this are underway.

The data compendium has facilitated the first global survey of resistance to NRDs. Reassuringly, prevalence of resistance to the NRDs was substantially lower than for first- and second-line agents in the dataset (Fig 3A), and resistance to the new drugs bedaquiline and delamanid was less common than the repurposed drugs clofazimine and linezolid in an MDR/RR background (Fig 4C). However, the presence of higher levels of delamanid and clofazimine resistance than ethambutol resistance in the isoniazid- and rifampicin-susceptible background does suggest some preexisting propensity towards NRD resistance (Fig 4B).

Coresistance between NRDs was seen in isolates in the compendium, the most common being isolates resistant to both bedaquiline and clofazimine. This link is well documented and has been attributed to shared resistance mechanisms such as nonsynonymous mutations in *rv0678*, which were found in both clofazimine- and bedaquiline-resistant isolates in the compendium [31] (Fig 5B and 5C). Increased clofazimine use could further increase the prevalence of *M*. *tuberculosis* isolates with clofazimine and bedaquiline coresistance, limiting MDR treatment options including using bedaquiline as the backbone of a shorter MDR regimen [42]. Therefore, proposed usage of clofazimine for other infectious diseases should be carefully considered.

WHO recommends against the use of bedaquiline and delamanid in combination to prevent the development of coresistance, which could occur relatively quickly [43]; the rate of spontaneous evolution of delamanid resistance *in vitro* has been shown to be comparable to that of isoniazid, and, likewise, bedaquiline resistance arises at a comparable rate to rifampicin resistance [44]. In this compendium, 12.9% of bedaquiline-resistant isolates were resistant to delamanid, and 7.1% of delamanid-resistant isolates were resistant to bedaquiline. Several scenarios could account for this, including the presence of shared resistance mechanisms. For example, as bedaquiline targets energy metabolism within the cell, changes to cope with energy/nutrient imbalances upon the acquisition of resistance-associated ATPase pump mutations may result in cross resistance to delamanid in a yet unknown or unexplored mechanism [13]. It is imperative that genetic determinants of resistance are fully explored for the NRDs, as these are our current treatments of last resort, with special attention given to those mechanisms that could be shared with other agents. In the meantime, careful stewardship and phenotypic and genotypic surveillance of the NRDs should be implemented, including linezolid and clofazimine, which are now group A and B drugs, respectively, for MDR treatment [1].

Several research avenues are being actively explored by the CRyPTIC consortium that make further use of this compendium, including the following: (i) relating genetic mutations to quantitative changes in the MICs of different drugs [13]; (ii) genome-wide association studies [15]; (iii) training machine learning models that can predict resistance [14]; and (iv) exploration of the genetic determinants of resistance to second line and NRDs [16]. Collectively, these studies share the same aim of facilitating the implementation of WGS-directed resistance prediction in the clinic. Finally, we urge other researchers to explore and analyse this large dataset of *M*. *tuberculosis* clinical isolates and hope it will lead to a wave of new and inciteful studies that will positively serve the TB community for years to come.

Methods

Ethics

Approval for CRyPTIC study was obtained by Taiwan Centers for Disease Control IRB No. 106209, University of KwaZulu Natal Biomedical Research Ethics Committee (UKZN BREC)

(reference BE022/13) and University of Liverpool Central University Research Ethics Committees (reference 2286), Institutional Research Ethics Committee (IREC) of The Foundation for Medical Research, Mumbai (Ref nos. FMR/IEC/TB/01a/2015 and FMR/IEC/TB/01b/2015), Institutional Review Board of P.D. Hinduja Hospital and Medical Research Centre, Mumbai (Ref no. 915-15-CR [MRC]), scientific committee of the Adolfo Lutz Institute (CTC-IAL 47-J / 2017) and in the Ethics Committee (CAAE: 81452517.1.0000.0059) and Ethics Committee review by Universidad Peruana Cayetano Heredia (Lima, Peru) and LSHTM (London, UK), Institutional Review Board at Pham Ngoc Thach Hospital, HCMC, Vietnam and Oxford Tropical Research Ethics Committee, UK, University of the Witwatersrand, Johannesburg Human Research Ethics Committee (Medical) (M160667), Technical Scientific Council (CTC-IAL no. 47-J / 2017) and Research Ethics Committee (CAAE 81452517.1.0000.0059) of Adolfo Lutz Institute, and University off Cape Town Faculty of Health Sciences Research Ethical Committee approvals (HREC 012/2007, 057/2013).

The CRyPTIC study involves analysis of microbiological isolates only—there is no associated data on patients. The study aggregates isolates from previous studies (which had previously obtained IRB approval) and also collected its own samples. Each IRB listed above assessed the protocol and saw no need for individual consent as only microbiological isolates were being analysed, and no personally identifiable information or host genetic data was used. In the remaining jurisdictions (IML Gauting (Germany), Public Health Scotland, Public Health Sweden, San Raffaele Scientific Institute, Italy), no IRB approval (and no individual patient consent) was required for studies analysing routinely collected microbiological isolates only.

Sample collection

This study was designed to identify as many drug resistance mechanisms and mutations as possible. Given that for many drugs there is a long tail of rare mutations present at different frequencies in different countries, sample collection was biased towards collecting resistant isolates, with (wherever possible) temporally and geographically matched susceptibles. Therefore, with 4 exceptions, all collecting sites either sequenced all culturable isolates, a random subsample of all culturable isolates (tailored to budget) or sequenced a subsample of all resistant with matched susceptible samples. The exceptions were as follows: 102 samples from South Africa, which were a clinical cohort recruited on the basis of the health service classifying them as RR; the first 1,000 out of 2,944 samples from Peru were a historical freezer collection with heterogeneous sampling process (and the remainder were sampled prospectively, at random); Brazilian samples combined all stored resistant samples, with prospectively sampled pan-susceptibles; in addition to sampling prospectively and retrospectively (from freezers) enriching for resistance and matched susceptibles, at IML Gauting (Germany) and National Institute for Communicable Diseases (Johannesburg), all isolates resistant to a NRD were included. Chinese isolates were collected according to the following strategy: collecting subsites were randomly selected from the 72 counties, and then all culture-positive isolates were included. A broad breakdown of sampling approaches is included in Table B in S1 File.

Plate assay

The CRyPTIC consortium designed 2 versions of the Sensititre MYCOTB plate (Thermo Fisher Scientific, USA) named the "UKMYC5" and "UKMYC6" microtitre plates [11,12]. These plates contain 5 to 10 doubling dilutions of 13 antibiotics (rifampicin (RIF), rifabutin (RFB), isoniazid (INH), ethambutol (EMB), levofloxacin (LEV), moxifloxacin (MXF), amikacin (AMI), kanamycin (KAN), ethionamide (ETH), clofazimine (CFZ), linezolid (LZD),

delamanid (DLM), and bedaquiline (BDQ)). DLM and BDQ were provided by Otsuka Pharmaceutical and Janssen Pharmaceuticals, respectively. The UKMYC5 plate also contained para-aminosalicylic acid (PAS), but the MICs were not reproducible, and, hence, it was excluded from the UKMYC6 plate design and is not included in any subsequent analysis [11].

A standard operating protocol for sample processing was defined by CRyPTIC as previously described [11,12]. Clinical samples were subcultured using 7H10 agar plates, Lowenstein–Jensen tubes, or MGIT tubes. Bacterial cell suspensions (0.5 McFarland standard, saline Tween) prepared from (no later than) 14-day-old colonies were diluted 100X in 10 ml enriched 7H9 broth prior to plate inoculation. A semiautomated Sensititre Autoinoculator (Thermo Fisher Scientific, USA) was used to inoculate 100 μl prepared cell suspensions (1.5×10^5 CFU/ml [5×10^4 CFU/ml—5×10^5 CFU/ml]) into each well of a UKMYC5/6 microdilution plate. The plate was sealed and incubated for 14 days at 37˚C. Quality control runs were performed periodically using *M*. *tuberculosis* H37Rv ATCC 27294, which is sensitive to all drugs on the plates.

Minimum inhibitory concentration (MIC) measurements

MICs for each drug were read after incubation for 14 days by a laboratory scientist using a Thermo Fisher Sensititre Vizion digital MIC viewing system [11]. The Vizion apparatus was also used to take a high contrast photograph of the plate with a white background, from which the MIC was measured again using the Automated Mycobacterial Growth Detection Algorithm (AMyGDA) software [45]. The AMyGDA algorithm was specifically developed to automate and perform quality control of MIC measurements and to facilitate machine learning studies within the consortium. AMyGDA detects the boundaries of each well using a Hough transform for circles and measures growth as the number of dark pixels within the area contained by this boundary.

All images where the MICs measured by Vizion and AMyGDA were different were uploaded to a citizen science project, BashTheBug, on the Zooniverse platform [46]. Each image was then classified by $\geq$11 volunteers and the median classification taken. MICs were then classified as high (at least 2 methods concur on the MIC), medium (either a scientist recorded a MIC measurement using Vizion but did not store the plate picture, or Vizion and AMyGDA disagree and there is no BashTheBug measurement), or low (all 3 methods disagree) quality.

To ensure adequate data coverage for this study, we took the MIC from the Vizion reading provided by the trained laboratory scientist if it was annotated as having medium or low quality.

Binary phenotype classification

Binary phenotypes (resistant/susceptible) were assigned from the MICs by applying epidemiological cutoff (ECOFF) values [12]; samples with MICs at or below the ECOFF are, by definition, wild-type and hence assigned to be susceptible to the drug in question [12]. Samples with MICs above the ECOFF are therefore classified as resistant (Fig A and Table C in S1 File). Please see [12] for the body of work supporting the use of the ECOFF relative to the compendium isolates and Table C in S1 File for the ECOFFs for each drug tested.

Genomic data processing and variant calling

Clinical samples were subcultured either using Lowenstein–Jensen tubes, 7H10 agar plates, or MGIT tubes for (no more than) 14 days prior to DNA extraction using either the FastPrep-24 instrument (MP Biomedicals) for cell lysis and ethanol precipitation or the

cetyltrimethylammonium bromide (CTAB) method. Paired-end libraries were prepared using a Nextera XT DNA sample preparation kit (Illumina, San Diego, CA, USA) and were sequenced on Illumina instruments. The resulting FASTQ files were processed using the bespoke pipeline Clockwork (v0.8.3, github.com/iqbal-lab-org/clockwork; [47]). Briefly, all raw sequencing files were indexed into a relational database with which Clockwork proceeds. Human, nasopharyngeal flora, and human immunodeficiency virus–related reads were removed, and remaining reads were trimmed (adapters and low-quality ends) using Trimmomatic and mapped with BWA-MEM to the *M. tuberculosis* H37Rv reference genome (NC000962.3) [48,49]. Read duplicates were removed. Genetic variants were called independently using Cortex and SAMtools, 2 variant callers with orthogonal strengths (SAMtools, a high-sensitivity SNP caller, and Cortex, a high-specificity SNP and indel caller) [50,51]. The 2 call sets were merged to produce a final call set, using the Minos adjudication tool (v0.11.0) to resolve loci where the 2 callers disagreed, by remapping reads to an augmented genome containing each alternative allele [24]. Default filters of a minimum depth of 5×, a fraction of supporting reads of 0.9 (Minos) and a genotype confidence percentile (GCP) filter of 0.5 were applied. The GCP filter is a normalised likelihood ratio test, giving a measure of confidence in the called allele compared with the other alternatives, and is described in [24]. This produced one variant call format (VCF) file per sample, each only describing positions where that sample differed from the reference.

These filtered VCFs were then combined to produce a single nonredundant list of all variants seen in the cohort. All samples were then processed a second time with Minos, remapping reads to a graphical representation of all the segregating variation within the cohort, generating VCF files that had an entry at all variable positions (thus for all samples, most positions would be genotyped as having the reference allele). These "regenotyped" VCFs were later used to calculate pairwise distances (see below). Please refer to Supplemental Method A in S1 File for commands used to generate the per-sample and regenotyped VCF files.

To remove untrustworthy loci, a genome mask was applied to the resulting VCF files (regions identified with self-blast matches in [52] comprising of 324,971 bp of the reference genome). Furthermore, positions with less than 90% of total samples passing default Clockwork/Minos variant call filters (described above) were filtered out, comprising 95,703 bp of the genome, of which 55,980 bp intersect with the genome mask.

Resistance prediction using a genetic catalogue

A hybrid catalogue of genetic variants associated with resistance to first- and second-line drugs based on existing catalogues was created and can be found at github.com/oxfordmmm/tuberculosis_amr_catalogues/blob/public/catalogues/NC_000962.3/NC_000962.3_CRyPTIC_v1.311_GARC1_RUS.csv [8,26]. We specifically did not use the recent WHO catalogue to avoid circularity and overtraining, as that catalogue was developed (via prior literature, expert rules, and a heuristic algorithm) based partially on these isolates [25]).

The resulting VCF file for each isolate (see "Genomic data processing and variant calling" section above) was compared to the genetic catalogue to determine the presence or absence of resistance-associated mutations for 8 drugs: RIF, INH, EMB, LEV, MXF, AMI, KAN, and ETH. We did not apply the approach used in [8] to make a prediction if a novel mutation was detected in a known resistance gene, as we simply wanted to measure how well a pre-CRyPTIC catalogue could predict resistance in the compendium. These results (found in PREDICTIONS.csv; see "Data availability" section for access) were then compared to the binary phenotypes (see "Binary phenotype classification" section for how these were defined) with the following metrics calculated: TP, the number of phenotypically resistant samples that are

correctly identified as resistant ("true positives"); FP, the number of phenotypically susceptible samples that are falsely identified as resistant ("false positives"); TN, the number of phenotypically susceptible samples that are correctly identified as susceptible ("true negatives"); FN, the number of phenotypically resistant samples that are incorrectly identified as susceptible ("false negative"); VME, very major error rate (false-negative rate), 0 to 1; ME, major error rate (false-positive rate), 0 to 1; PPV, positive predictive value, 0 to 1; NPV, negative predictive value.

Phylogenetic tree construction

A pairwise genetic distance matrix was constructed for 15,211 isolates by comparing pairs of regenotyped VCF files (see "Genomic data processing and variant calling" section above for more details). A neighbourhood-joining tree was constructed from the distance matrix using *quicktree* [53]. Tree visualisation and annotation were performed using the R library *ggtree* [54]. *M. tuberculosis* lineages were assigned using Mykrobe and are represented by the coloured dots at the branch termini of the tree [24]. For isolates that had "mixed" lineage classification (i.e., 2 lineages were found present in the sample by Mykrobe, $n = 225$, 1.5%), the first of the 2 lineages was assigned to the isolate. *ggtree* was also used to construct the trees depicting BDQ-, CFZ-, and DLM-resistant isolates.

The data

Data are available from ftp.ebi.ac.uk/pub/databases/cryptic/release_june2022/.

The FTP site contains 2 top level directories: "reuse" and "reproducibility". All data for this study were analysed and visualised using either R or python3 libraries and packages. See github.com/kerrimalone/Brankin_Malone_2022 for codebase.

"reuse" directory

We point the reader to this directory to gain access to CRyPTIC project data. "CRyPTIC_reuse_table_20221019.csv" contains genotypic and phenotypic data relating to the figures and summaries listed in this manuscript and is what we present as a general use reference table for most future projects. It includes binary phenotypes (R/S), MICs, phenotype quality metrics, and ENA sample IDs for 12,288 compendium isolates (see "Quality assurance of the minimum inhibitory concentrations for 13 drugs" section below in Results for filters applied to obtain this final set of isolated). It also includes file paths to each isolate's VCF file and "regenotyped" VCF file (VCF files that have an entry at all variable positions; see "Genomic data processing and variant calling" section above for more). "CRyPTIC_excluded_samples_20220607.tsv" contains the ENA accession numbers and file paths to each isolate's VCF file and "regenotyped" VCF file for the 2,922 samples that were sequenced but have no relative MIC data.

"reproducibility" directory

This directory contains the data used for multiple CRyPTIC project publications referenced throughout this manuscript. As stated above, each project has taken slightly different subsets of these data as documented in those papers. For example, see how tables such as "MUTATIONS.csv" and "GENOTYPES.csv" were used and filtered, (along with others) in this study to obtain the reuse file "CRyPTIC_reuse_table_20221019.csv" in Fig 1. Again, for optimal use of CRyPTIC data in your own project, please refer to "CRyPTIC_reuse_table_20221019.csv" in the "reuse" directory. All data for this study were analysed and visualised using either R or python3 libraries and packages. See github.com/kerrimalone/Brankin_Malone_2022 for codebase.

Author Contributions

Conceptualisation: Daniela M Cirillo, Derrick W. Crook, Philip W Fowler, Sarah Hoosdally, Ana Lúıza Gibertoni Cruz, Nazir A. Ismail, Stefan Niemann, Zamin Iqbal, Tim E.A. Peto, A Sarah Walker, Timothy M Walker

Data Curation: Philip W Fowler, Sarah J Hoosdally, Ana Lúıza Gibertoni Cruz, Alice Brankin, Kerri M. Malone, Zamin Iqbal, Martin Hunt and Jeff Knaggs

Formal Analysis: Alice Brankin, Kerri M. Malone

Funding acquisition: Camilla Rodrigues, David Moore, Derrick W. Crook, Daniela M. Cirillo, Zamin Iqbal, Nazir A. Ismail, Nerges Mistry, Stefan Niemann, Tim E.A. Peto, Guy Thwaites, A. Sarah Walker, Timothy M Walker, Daniel J. Wilson

Investigation: Alice Brankin, Kerri M. Malone

Project Administration: Daniela M. Cirillo, Derrick W Crook, Philip W Fowler, Sarah Hoosdally, Zamin Iqbal, Tim E.A. Peto, Aysha Roohi,

Resources: The CRyPTIC Consortium

Software: Philip W Fowler, Martin Hunt, Jeff Knaggs, Brice Letcher

Supervision: Daniela M. Cirillo, Derrick W Crook, Philip W Fowler, Zamin Iqbal, Tim E.A. Peto, Daniel J. Wilson

Validation: Emanuele Borroni, Daniela M Cirillo, Philip W. Fowler, Clara Grazian, Sarah J. Hoosdally, Martin Hunt, Timothy E. A. Peto, Paola M. V. Rancoita

Visualisation: Alice Brankin, Kerri M. Malone

Writing - original draft preparation: Alice Brankin, Kerri M. Malone

Writing - review and editing: Alice Brankin, Kerri M. Malone, Philip W. Fowler, Zamin Iqbal

Supporting information

S1 File. Supporting information. Text A. Acknowledgements. Text B. Lineages of the *M. tuberculosis* isolates of the compendium. Table A. Acronyms used in this manuscript. Table B. Sampling strategies at different collection sites. Table C. Epidemiological cutoff values (ECOFFs) used to binarize MIC measurements into resistant and susceptible. Table D. Lineages–v- geographical location of origin/contribution for CRyPTIC isolates. Table E. Sublineages–v- geographical location of origin/contribution for CRyPTIC isolates. Table F. Sample information for isolates classified as resistant to all 13 drugs tested. Table G. Co-occurrence of antibiotic resistance in CRyPTIC *M. tuberculosis* isolates. Supplemental Method A. Generating per-sample and regenotyped VCF files. Fig A. Per-drug MIC distributions of isolates plated on CRyPTIC designed variations on the Thermo Fischer Sensititre MYCOTB MIC plate; UKMYC5 (A) and UKMYC6 (B). Fig B. Geographical distribution of 15,211 CRyPTIC *M. tuberculosis* clinical isolates. Fig C. A significant association between country and lineage can be seen in the CRyPTIC data. Fig D. Phylogenetic tree of CRyPTIC *M. tuberculosis* clinical isolates. Fig E. Nonsynonymous mutations found outside the RRDR of rpoB in RMR isolates and MDR isolates.
(DOCX)

Acknowledgments

We thank Faisal Masood Khanzada and Alamdar Hussain Rizvi (NTRL, Islamabad, Pakistan), Angela Starks and James Posey (Centers for Disease Control and Prevention, Atlanta, USA),

and Juan Carlos Toro and Solomon Ghebremichael (Public Health Agency of Sweden, Solna, Sweden), Iñaki Comas and Álvaro Chiner-Oms (Instituto de Biología Integrativa de Sistemas, Valencia, Spain; CIBER en Epidemiología y Salud Pública, Valencia, Spain; Instituto de Biomedicina de Valencia, IBV-CSIC, Valencia, Spain). We thank the Wadsworth Center Applied Genomic Technologies Core Facility and the Wadsworth Center Bioinformatics Core for additional support for sequencing and analysis and SYNLAB Holding Germany GmbH for its direct and indirect support of research activities in the Institute of Microbiology and Laboratory Medicine Gauting. N.R. thanks the Programme National de Lutte contre la Tuberculose de Madagascar.

References

1. WHO. Global Tuberculosis Report 2020. 2021.
2. World Health Organisation. EndTB Campaign. Available from: www.who.int/teams/global-tuberculosis-programme/the-end-tb-strategy.
3. Shinnick TM, Starks AM, Alexander HL, Castro KG. Evaluation of the Cepheid Xpert MTB/RIF assay. Expert Rev Mol Diagn. 2015; 15:9–22. https://doi.org/10.1586/14737159.2015.976556 PMID: 25373876
4. Boehme CC, Nicol MP, Nabeta P, Michael JS, Gotuzzo E, Tahirli R, et al. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. Lancet. 2011; 377:1495–1505. https://doi.org/10.1016/S0140-6736(11)60438-8 PMID: 21507477
5. Makhado NA, Matabane E, Faccin M, Pinçon C, Jouet A, Boutachkourt F, et al. Outbreak of multidrug-resistant tuberculosis in South Africa undetected by WHO-endorsed commercial tests: an observational study. Lancet Infect Dis. 2018; 18:1350–1359. https://doi.org/10.1016/S1473-3099(18)30496-1 PMID: 30342828
6. Beckert P, Sanchez-Padilla E, Merker M, Dreyer V, Kohl TA, Utpatel C, et al. MDR M. tuberculosis outbreak clone in Eswatini missed by Xpert has elevated bedaquiline resistance dated to the pre-treatment era. Genome Med. 2020; 12:104. https://doi.org/10.1186/s13073-020-00793-8 PMID: 33239092
7. Sanchez-Padilla E, Merker M, Beckert P, Jochims F, Dlamini T, Kahn P, et al. Detection of Drug-Resistant Tuberculosis by Xpert MTB/RIF in Swaziland, New England. J Med. 2015; 372:1181–1182. https://doi.org/10.1056/NEJMc1413930 PMID: 25785984
8. The CRyPTIC Consortium, 100 000 Genomes ProjecT. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018; 379:1403–1415. https://doi.org/10.1056/NEJMoa1800474 PMID: 30280646
9. Pankhurst LJ, del Ojo EC, Votintseva AA, Walker TM, Cole K, Davies J, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med. 2016; 4:49–58. https://doi.org/10.1016/S2213-2600(15)00466-X PMID: 26669893
10. Kalokhe AS, Shafiq M, Lee JC, Ray SM, Wang YF, Metchock B, et al. Multidrug-resistant tuberculosis drug susceptibility and molecular diagnostic testing. Am J Med Sci. 2013; 345:143–148. https://doi.org/10.1097/MAJ.0b013e31825d32c6 PMID: 22885627
11. Rancoita PM v, Cugnata F, Gibertoni Cruz AL, Borroni E, Hoosdally SJ, Walker TM, et al. Validating a 14-Drug Microtiter Plate Containing Bedaquiline and Delamanid for Large-Scale Research Susceptibility Testing of Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2018;62. https://doi.org/10.1128/AAC.00344-18 PMID: 29941636
12. The CRyPTIC Consortium. Epidemiological cutoff values for a 96-well broth microdilution plate for high-throughput research antibiotic susceptibility testing of *M. tuberculosis*. Eur Respir J. 2022:2200239. https://doi.org/10.1183/13993003.00239-2022 PMID: 35301246
13. The CRyPTIC Consortium. Quantitative measurement of antibiotic resistance in Mycobacterium tuberculosis reveals genetic determinants of resistance and susceptibility in a target gene approach. bioRxiv. 2021. https://doi.org/10.1101/2021.09.14.460353
14. The CRyPTIC Consortium. Predicting Susceptibility to First- and Second-line Tuberculosis Drugs by DNA sequencing and Machine Learning. In preparation.
15. The CRyPTIC Consortium. Genome-wide association studies of global Mycobacterium tuberculosis resistance to thirteen antimicrobials in 10,228 genomes. bioRxiv. 2021. https://doi.org/10.1101/2021.09.14.460272

16. Sonnenkalb L, Carter J, Spitaleri A, Iqbal Z, Hunt M, Malone K, et al. Deciphering Bedaquiline and Clofazimine Resistance in Tuberculosis. An Evolutionary Medicine Approach. bioRxiv. 2021:2021.03.19.436148. https://doi.org/10.1101/2021.03.19.436148

17. Coscolla M, Gagneux S. Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol. 2014; 26:431–444. https://doi.org/10.1016/j.smim.2014.09.012 PMID: 25453224

18. Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol. 2018; 16:202–213. https://doi.org/10.1038/nrmicro.2018.8 PMID: 29456241

19. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, et al. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021; 12:6099. https://doi.org/10.1038/s41467-021-26248-1 PMID: 34671035

20. Dijkstra JA, van der Laan T, Akkerman OW, Bolhuis MS, de Lange WCM, Kosterink JGW, et al. In Vitro Susceptibility of Mycobacterium tuberculosis to Amikacin, Kanamycin, and Capreomycin. Antimicrob Agents Chemother. 2018; 62. https://doi.org/10.1128/AAC.01724-17 PMID: 29311078

21. Maitre T, Petitjean G, Chauffour A, Bernard C, el Helali N, Jarlier V, et al. Are moxifloxacin and levofloxacin equally effective to treat XDR tuberculosis? J Antimicrob Chemother. 2017; 72:2326–2333. https://doi.org/10.1093/jac/dkx150 PMID: 28535203

22. Berrada ZL, Lin S-YG, Rodwell TC, Nguyen D, Schecter GF, Pham L, et al. Rifabutin and rifampin resistance levels and associated rpoB mutations in clinical isolates of Mycobacterium tuberculosis complex. Diagn Microbiol Infect Dis. 2016; 85:177–181. https://doi.org/10.1016/j.diagmicrobio.2016.01.019 PMID: 27036978

23. Ho J, Jelfs P, Sintchenko V. Fluoroquinolone resistance in non-multidrug-resistant tuberculosis—a surveillance study in New South Wales, Australia, and a review of global resistance rates. Int J Infect Dis. 2014; 26:149–153. https://doi.org/10.1016/j.ijid.2014.03.1388 PMID: 25086437

24. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. Wellcome Open Res. 2019; 4:191. https://doi.org/10.12688/wellcomeopenres.15603.1 PMID: 32055708

25. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, et al. The 2021 WHO Catalogue of Mycobacterium Tuberculosis Complex Mutations Associated with Drug Resistance: A New Global Standard for Molecular Diagnostics. SSRN Electronic Journal. 2021. https://doi.org/10.2139/ssrn.3923444

26. Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J. 2017; 50:1701354. https://doi.org/10.1183/13993003.01354-2017 PMID: 29284687

27. Singh R, Manjunatha U, Boshoff HIM, Ha YH, Niyomrattanakit P, Ledwidge R, et al. PA-824 Kills Nonreplicating Mycobacterium tuberculosis by Intracellular NO Release. Science. 1979; 2008(322):1392–1395. https://doi.org/10.1126/science.1164571 PMID: 19039139

28. Hartkoorn RC, Uplekar S, Cole ST. Cross-Resistance between Clofazimine and Bedaquiline through Upregulation of MmpL5 in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2014; 58:2979–2981. https://doi.org/10.1128/AAC.00037-14 PMID: 24590481

29. Degiacomi G, Sammartino JC, Sinigiani V, Marra P, Urbani A, Pasca MR. In vitro Study of Bedaquiline Resistance in Mycobacterium tuberculosis Multi-Drug Resistant Clinical Isolates. Front Microbiol. 2020; 11:559469. https://doi.org/10.3389/fmicb.2020.559469 PMID: 33042066

30. Almeida D, Ioerger T, Tyagi S, Li S- Y, Mdluli K, Andries K, et al. Mutations in pepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2016; 60:4590–4599. https://doi.org/10.1128/AAC.00753-16 PMID: 27185800

31. Nimmo C, Millard J, van Dorp L, Brien K, Moodley S, Wolf A, et al. Population-level emergence of bedaquiline and clofazimine resistance-associated variants among patients with drug-resistant tuberculosis in southern Africa: a phenotypic and phylogenetic analysis. Lancet Microbe. 2020; 1:e165–e174. https://doi.org/10.1016/S2666-5247(20)30031-8 PMID: 32803174

32. Mvelase NR, Balakrishna Y, Lutchminarain K, Mlisana K. Evolving rifampicin and isoniazid mono-resistance in a high multidrug-resistant and extensively drug-resistant tuberculosis region: a retrospective data analysis. BMJ Open. 2019; 9:e031663. https://doi.org/10.1136/bmjopen-2019-031663 PMID: 31699736

33. Villegas L, Otero L, Sterling TR, Huaman MA, van der Stuyft P, Gotuzzo E, et al. Prevalence, Risk Factors, and Treatment Outcomes of Isoniazid- and Rifampicin- Mono-Resistant Pulmonary Tuberculosis in Lima, Peru. PLoS ONE. 2016; 11:e0152933. https://doi.org/10.1371/journal.pone.0152933 PMID: 27045684

34. Park S, Jo K-W, Lee S d, Kim WS, Shim TS. Treatment outcomes of rifampin-sparing treatment in patients with pulmonary tuberculosis with rifampin-mono-resistance or rifampin adverse events: A

retrospective cohort analysis. Respiratory Medicine. 2017; 131:43–48. https://doi.org/10.1016/j.rmed.2017.08.002 PMID: 28947041

35. Salaam-Dreyer Z, Streicher EM, Sirgel FA, Menardo F, Borrell S, Reinhard M, et al. Rifampicin-Monoresistant Tuberculosis Is Not the Same as Multidrug-Resistant Tuberculosis: a Descriptive Study from Khayelitsha, South Africa. Antimicrobial Agents and Chemotherapy. 2021; 65. https://doi.org/10.1128/AAC.00364-21

36. Pang Y, Lu J, Wang Y, Song Y, Wang S, Zhao Y. Study of the Rifampin Monoresistance Mechanism in Mycobacterium tuberculosis. Antimicrob Agents Chemother. 2013; 57:893–900. https://doi.org/10.1128/AAC.01024-12 PMID: 23208715

37. Kigozi E, Kasule GW, Musisi K, Lukoye D, Kyobe S, Katabazi FA, et al. Prevalence and patterns of rifampicin and isoniazid resistance conferring mutations in Mycobacterium tuberculosis isolates from Uganda. PLoS ONE. 2018; 13:e0198091. https://doi.org/10.1371/journal.pone.0198091 PMID: 29847567

38. Dean AS, Zignol M, Cabibbe AM, Falzon D, Glaziou P, Cirillo DM, et al. Prevalence and genetic profiles of isoniazid resistance in tuberculosis patients: A multicountry analysis of cross-sectional data. PLoS Med. 2020; 17:e1003008. https://doi.org/10.1371/journal.pmed.1003008 PMID: 31961877

39. Heysell SK, Ahmed S, Rahman MT, Akhanda MW, Gleason AT, Ebers A, et al. Hearing loss with kanamycin treatment for multidrug-resistant tuberculosis in Bangladesh. Eur Respir J. 2018; 51:1701778. https://doi.org/10.1183/13993003.01778-2017 PMID: 29348152

40. Migliori GB, Langendam MW, D'Ambrosio L, Centis R, Blasi F, Huitric E, et al. Protecting the tuberculosis drug pipeline: stating the case for the rational use of fluoroquinolones. Eur Respir J. 2012; 40:814–822. https://doi.org/10.1183/09031936.00036812 PMID: 22653774

41. WHO. Update on the use of nucleic acid amplification tests to detect TB and drug-resistant TB: rapid communication. 2021.

42. Olaru ID, Heyckendorf J, Andres S, Kalsdorf B, Lange C. Bedaquiline-based treatment regimen for multidrug-resistant tuberculosis. Eur Respir J. 2017; 49:1700742. https://doi.org/10.1183/13993003.00742-2017 PMID: 28529207

43. WHO. World Health Organization treatment guidelines for drug-resistant tuberculosis, 2016 update. 2016.

44. Huitric E, Verhasselt P, Koul A, Andries K, Hoffner S, Andersson DI. Rates and Mechanisms of Resistance Development in *Mycobacterium tuberculosis* to a Novel Diarylquinoline ATP Synthase Inhibitor. Antimicrob Agents Chemother. 2010; 54:1022–1028. https://doi.org/10.1128/AAC.01611-09 PMID: 20038615

45. Fowler PW, Gibertoni Cruz AL, Hoosdally SJ, Jarrett L, Borroni E, Chiacchiaretta M, et al. Automated detection of bacterial growth on 96-well plates for high-throughput drug susceptibility testing of Mycobacterium tuberculosis. Microbiology (N Y). 2018; 164:1522–1530. https://doi.org/10.1099/mic.0.000733 PMID: 30351270

46. Fowler PW, Wright C, Spiers-Bowers H, Zhu T, Baeten EML, Hoosdally SW, et al. BashTheBug: a crowd of volunteers reproducibly and accurately measure the minimum inhibitory concentrations of 13 antitubercular drugs from photographs of 96-well broth microdilution plates. bioRxiv. 2021; 2021.07.20.453060. https://doi.org/10.1101/2021.07.20.453060

47. Hunt M, Letcher B, Malone KM, Nguyen G, Hall MB, Colquhoun RM, et al. Minos: variant adjudication and joint genotyping of cohorts of bacterial genomes. bioRxiv. 2021; 2021.09.15.460475. https://doi.org/10.1101/2021.09.15.460475

48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

49. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

51. Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. Nat Genet. 2012; 44:226–232. https://doi.org/10.1038/ng.1028 PMID: 22231483

52. Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. Lancet Respir Med. 2014; 2:285–292. https://doi.org/10.1016/S2213-2600(14)70027-X PMID: 24717625

53. Howe K, Bateman A, Durbin R. QuickTree: building huge Neighbour-Joining trees of protein sequences. Bioinformatics. 2002; 18:1546–1547. https://doi.org/10.1093/bioinformatics/18.11.1546 PMID: 12424131

54. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinform. 2020; 69. https://doi.org/10.1002/cpbi.96 PMID: 32162851