Genome Medicine

RESEARCH

Open Access

Detection of a historic reservoir of bedaquiline/clofazimine resistance-associated variants in *Mycobacterium tuberculosis*

Camus Nimmo[1,2,3*], Arturo Torres Ortiz[1,4], Cedric C. S. Tan[1], Juanita Pang[1,2], Mislav Acman[1], James Millard[3,5,6], Nesri Padayatchi[7], Alison D. Grant[3,8], Max O'Donnell[7,9], Alex Pym[3], Ola B. Brynildsrud[10], Vegard Eldholm[10], Louis Grandjean[2,11,12], Xavier Didelot[13], François Balloux[1*†] and Lucy van Dorp[1*†]

Abstract

Background Drug resistance in tuberculosis (TB) poses a major ongoing challenge to public health. The recent inclusion of bedaquiline into TB drug regimens has improved treatment outcomes, but this advance is threatened by the emergence of strains of *Mycobacterium tuberculosis* (*Mtb*) resistant to bedaquiline. Clinical bedaquiline resistance is most frequently conferred by off-target resistance-associated variants (RAVs) in the *mmpR5* gene (*Rv0678*), the regulator of an efflux pump, which can also confer cross-resistance to clofazimine, another TB drug.

Methods We compiled a dataset of 3682 *Mtb* genomes, including 180 carrying variants in *mmpR5*, and its immediate background (i.e. *mmpR5* promoter and adjacent *mmpL5* gene), that have been associated to borderline (henceforth intermediate) or confirmed resistance to bedaquiline. We characterised the occurrence of all nonsynonymous mutations in *mmpR5* in this dataset and estimated, using time-resolved phylogenetic methods, the age of their emergence.

Results We identified eight cases where RAVs were present in the genomes of strains collected prior to the use of bedaquiline in TB treatment regimes. Phylogenetic reconstruction points to multiple emergence events and circulation of RAVs in *mmpR5*, some estimated to predate the introduction of bedaquiline. However, epistatic interactions can complicate bedaquiline drug-susceptibility prediction from genetic sequence data. Indeed, in one clade, Ile67fs (a RAV when considered in isolation) was estimated to have emerged prior to the antibiotic era, together with a resistance reverting *mmpL5* mutation.

†François Balloux and Lucy van Dorp contributed equally to this work.

*Correspondence:
Camus Nimmo
camus.nimmo@crick.ac.uk
François Balloux
f.balloux@ucl.ac.uk
Lucy van Dorp
lucy.dorp.12@ucl.ac.uk
Full list of author information is available at the end of the article

Conclusions The presence of a pre-existing reservoir of *Mtb* strains carrying bedaquiline RAVs prior to its clinical use augments the need for rapid drug susceptibility testing and individualised regimen selection to safeguard the use of bedaquiline in TB care and control.

Keywords Tuberculosis, Phylogenetics, Bedaquiline, Drug resistance, AMR

Background

Drug-resistant tuberculosis (DR-TB) currently accounts for 450,000 of the 10 million new tuberculosis (TB) cases reported annually [1]. Treatment outcomes for multidrug-resistant TB (MDR-TB), resistant to at least rifampicin and isoniazid, have historically been poor, with treatment success rates of only 50–60% in routine programmatic settings [2, 3]. The discovery of bedaquiline, a diarylquinoline antimycobacterial active against ATP synthase, which is highly effective against *Mycobacterium tuberculosis* (*Mtb*) [4], was reported in 2004. Following clinical trials, which confirmed reduced time to culture conversion in patients with DR-TB [5], in 2012, bedaquiline received an accelerated Food and Drug Administration (FDA) licence for use in DR-TB [6].

Cohort studies of patients treated with bedaquiline-containing regimens against MDR-TB report success rates of 70–80% [7, 8]. Similar results have been achieved for extensively drug-resistant TB (XDR-TB, traditionally defined as MDR-TB strains with additional resistance to fluoroquinolones and injectables), where treatment outcomes without bedaquiline are even worse [9, 10]. In light of these promising results, the World Health Organization (WHO) now recommends that bedaquiline be included in all MDR-TB regimens [11]. It has played a central role in the highly successful ZeNix [12] and TB-PRACTECAL [13] trials of bedaquiline, pretomanid and linezolid (+/−moxifloxacin) six-month all-oral regimens for DR-TB. These are now incorporated in WHO guidance. In addition, bedaquiline is positioned as a key drug in multiple phase III clinical trials for drug-susceptible TB (SimpliciTB, ClinicalTrials.gov NCT03338621; TRUNCATE-TB [14]).

Resistance in *Mtb* is typically reported shortly after the introduction of a novel TB drug and often appears sequentially [15, 16]. For example, mutations conferring resistance to isoniazid — one of the first antimycobacterials — tend to emerge prior to resistance to rifampicin, the other major first-line drug. These also predate resistance mutations to second-line drugs, so termed because they are used clinically to treat patients infected with strains already resistant to first-line drugs. This was observed, for example, in KwaZulu-Natal, South Africa, where resistance-associated mutations accumulated over decades prior to their identification, leading to a major outbreak of extensively drug-resistant TB (XDR-TB) [16]. Unlike other major drug-resistant bacteria, *Mtb* reproduces strictly clonally and systematically acquires resistance by chromosomal mutations rather than via horizontal gene transfer or recombination [17]. This allows genome-based phylogenetic reconstructions to infer the timings of emergence and subsequent spread of variants in *Mtb* for which there is evidence of an association with phenotypic resistance in at least some genetic backgrounds, termed resistance-associated variants (RAVs).

A number of mechanisms have been implicated in bedaquiline resistance. For example, mutations conferring resistance have been selected in vitro, located in the *atpE* gene encoding the F1F0 ATP synthase, the target of bedaquiline [18]. Off-target resistance-conferring mutations have also been found in *pepQ* in a murine model and potentially in a small number of patients [19]. However, the primary mechanism of resistance observed in clinical isolates has been identified in the context of off-target RAVs in the *mmpR5 (Rv0678)* gene, a negative repressor of expression of the MmpL5 efflux pump. Loss of function of MmpR5 leads to pump overexpression [20] and increased minimum inhibitory concentrations (MIC) to bedaquiline, along with the recently repurposed antimycobacterial clofazimine, fusidic acid, the azole class of antifungal drugs (which also have antimycobacterial activity), as well as to the novel therapeutic class of DprE1 inhibitors in clinical trials [21, 22]. Aligned with this mechanism of resistance, coincident mutations leading to loss of function of the MmpL5 efflux pump can negate the resistance-inducing effect of MmpR5 loss of function [23].

A range of single nucleotide polymorphisms (SNPs) and frameshift *mmpR5* mutations have been associated with resistance to bedaquiline and are often present as heteroresistant alleles in patients [24–35]. In contrast to most other RAVs in *Mtb*, which often cause many-fold increases in MIC and clear-cut resistance, *mmpR5* variants may be associated with normal MICs or subtle increases in bedaquiline MIC, although they may still be clinically important [36]. These increases may not cross the current WHO critical concentrations used to classify resistant versus susceptible strains (0.25 µg/mL on Middlebrook 7H11 agar, or 1 µg/mL in Mycobacteria Growth Indicator Tube [MGIT] liquid media). The second version of the WHO tuberculosis drug resistance catalogue

identifies 86 individual bedaquiline RAVs (Group 1 and Group 2 assignment, https://iris.who.int/handle/10665/374061 accessed January 2024) [35]. Bedaquiline has a long terminal half-life of up to 5.5 months [6], leading to the possibility of subtherapeutic concentrations where adherence is suboptimal or treatment is interrupted, which could act as a further driver of resistance.

Bedaquiline and clofazimine cross-resistance has now been reported across three continents following the rapid expansion in usage of both drugs [25, 30, 37–39] and is associated in some cases with poor adherence to therapy and inadequate regimens. However, baseline isolates in 8/347 (2.3%) patients from phase IIb bedaquiline trials demonstrated *mmpR5* RAVs and high bedaquiline MICs in the absence of prior documented use of bedaquiline or clofazimine [40]. This suggests that bedaquiline RAVs may have been in circulation prior to the usage of either of these drugs, which may be expected in the case of mutations which do not have major fitness consequences [41]. While there have been isolated clinical reports from multiple geographical regions, the global extent of bedaquiline resistance emergence and spread has not yet been investigated.

In this study, we characterise and date the emergence of variants in *mmpR5*, including those implicated as bedaquiline RAVs, in the two global *Mtb* lineage 2 (L2) and lineage 4 (L4) lineages, which include the majority of drug resistance strains [15]. Phylogenetic analyses of two datasets comprising 1514 *Mtb* L2 and 2168 L4 whole genome sequences revealed the emergence and spread of multiple *mmpR5* variants associated to resistance or borderline (intermediate) resistance to bedaquiline prior to its first clinical use. This pre-existing reservoir of bedaquiline/clofazimine-resistant *Mtb* strains suggests *mmpR5* RAVs exert a relatively low fitness cost which could be rapidly selected for as bedaquiline and clofazimine are more widely used in TB treatment.

Methods

Sample collection

In this study, we curated large representative datasets of *Mtb* whole genome sequences encompassing the global genetic and geographic distribution of lineages 2 (L2) and L4 (Fig. 1, Additional file 1: Table S1, Additional file 2: Fig S1). The dataset was enriched to include all available sequenced isolates with *mmpR5* variants, which in some

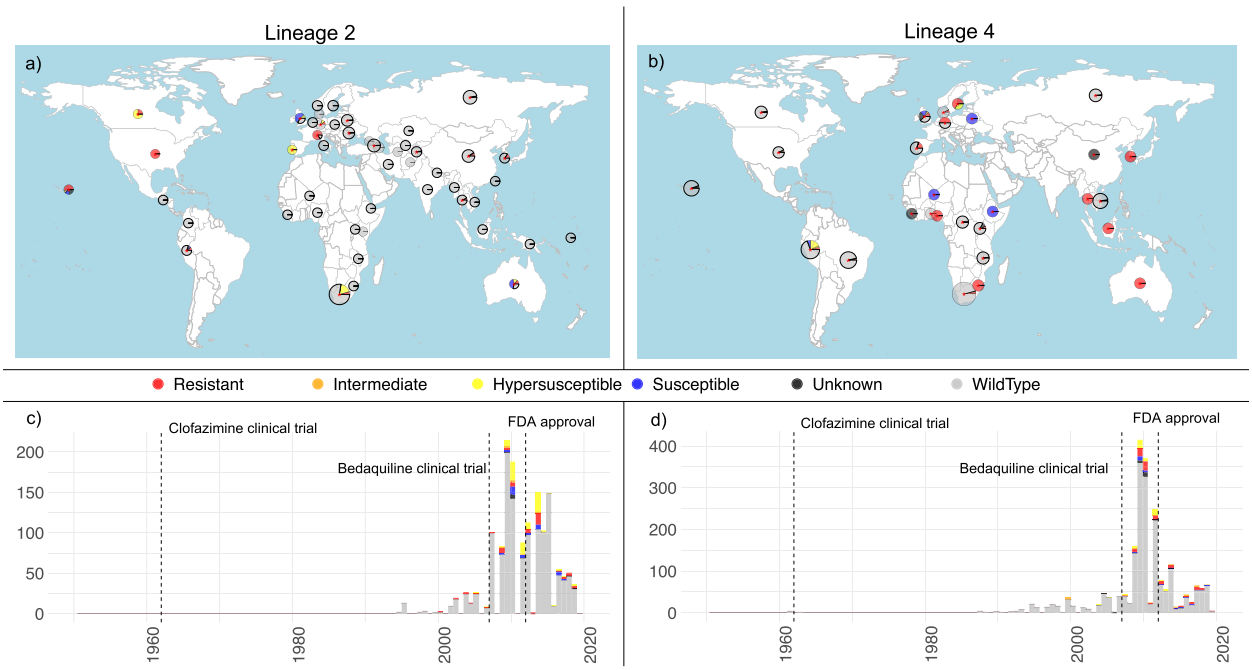

Fig. 1 Compiled global *Mtb* genomic datasets. Panels **a** and **b** provide the geographic location of isolates included in the lineage 2 and lineage 4 datasets respectively. Pies are scaled by the number of samples per country (raw data available in Additional file 1: Table S1) with the colours providing the fraction of genomes with any nonsynonymous/frameshift variants detected in *mmpR5* (coloured as per the legend). Countries comprising samples with known RAVs are highlighted with a red asterisk. Genomic data for which no associated metadata on the geographic location of sampling was available are shown in the Pacific Ocean. Panels **c** and **d** provide the collection dates associated to each genome in the lineage 2 and lineage 4 datasets respectively highlighting those with any variants in *mmpR5* (colour, as per legend). Lineage 4 *Mtb* obtained from eighteenth century mummies are excluded from this plot but included in all analyses. The vertical dashed lines indicate the dates of the first clinical trials for clofazimine, bedaquiline and FDA approval of bedaquiline for clinical use

cases included isolates with no, or limited, published metadata. In all other cases, samples for which metadata on the geographic location and date of collection was available were retained. To ensure high-quality consensus alignments we required that all samples mapped with a minimum percentage cover of 96% and a mean coverage of 30× to the H37Rv reference genome (NC_000962.3). We excluded any samples with evidence of mixed strain infection as identified by the presence of lineage-specific SNPs to more than one sublineage [42] or the presence of a high proportion of heterozygous alleles [43]. The total number of samples included in these datasets, and their source is shown in Additional file 2: Table S2. An index of all samples is available in Additional file 1: Table S1.

A large global dataset of 1669 L4 *Mtb* sequences has been previously constructed, which we used as the basis for curating our L4 dataset [44]. We refer to this as the 'base dataset' for L4. For L2, we constructed a 'base dataset' by screening the Sequence Read Archive (SRA) and European Nucleotide Archive (ENA) using BIGSI [45] for the *rpsA* gene sequence containing the L2 defining variant *rpsA* a636c [42] with a 100% match. This search returned 6307 *Mtb* submissions, of which 1272 represented unique samples that had the minimum required metadata. Metadata from three studies were also added manually as they were not included in their respective SRA submissions but were available within published studies [46–48].

For isolates with only information on the year of sample collection, we set the date to be equal to the middle of the year. For those with information on the month but not the date of collection we set the date of collection to the first of the month. For sequenced samples which were missing associated metadata (32 L2 genomes and 19 L4 genomes), we attempted to estimate an average time of sample collection to impute a sampling date. To do, so we computed the average time between the date of collection and sequence upload date for all samples with associated dates separately in each of the L2 and L4 datasets (Additional file 2: Fig S1). For L2 we estimated a mean lag time of 4.7 years (0.5–12.6 years 95% CI). For L4, having excluded three sequences obtained from eighteenth century mummies from Hungary [49], we estimated a mean lag time of 6.9 years (0.6–19.1 years 95% CI). The estimated dates, where required, are provided in Additional file 1: Table S1.

To enrich the datasets for isolates with *mmpR5* variants, we included further sequences from our own studies in KwaZulu-Natal, South Africa [50, 51], other studies of drug-resistant TB in southern Africa [16, 44, 52–55], and Peru [56, 57]. We additionally supplement the Peruvian data with 163 previously unpublished isolates. In these cases, and to facilitate the most accurate possible estimation of the date of resistance emergence, we included samples with *mmpR5* variants as well as genetically related sequences without *mmpR5* variants.

To identify further published raw sequencing data with *mmpR5* variants from studies where bedaquiline/clofazimine resistance may have been previously unidentified, we screened the NCBI SRA for sequence data containing 85 previously published *mmpR5* variants [28–30, 50, 51, 58, 59] with BIGSI [45]. BIGSI was employed against a publicly available indexed database of complete SRA/ENA bacterial and viral whole genome sequences current to December 2016 (available here: http://ftp.ebi.ac.uk/pub/software/bigsi/nat_biotech_2018/all-microbial-index-v03/), and also employed locally against an updated in-house database which additionally indexed SRA samples from January 2017 until January 2019. Samples added using this approach are flagged 'BIGSI' in Additional file 1: Table S1. We also used the PYGSI tool [60] to interrogate BIGSI with the *mmpR5* sequence adjusted to include every possible single nucleotide substitution. In each instance, we included 30 bases upstream and downstream of the gene as annotated on the H37Rv *Mtb* reference genome. Samples added following the PYGSI screen are flagged 'PYGSI' in Additional file 1: Table S1. A breakdown of the different datasets used is provided in Additional file 2: Table S2.

Reference mapping and variant calling

Original fastq files for all included sequences were downloaded and paired reads mapped to the H37Rv reference genome with bwa mem v0.7.17 [61]. Mapped reads were sorted and de-duplicated using Picard Tools v2.20 followed by indel realignment with GATK v3.8 [62]. Alignment quality and coverage was recorded with Qualimap v2.21 [63]. Variant calling was performed using bcftools v1.9, based on reads mapping with a minimum mapping quality of 20, base quality of 20, no evidence of strand or position bias, a minimum coverage depth of 10 reads, and a minimum of four reads supporting the alternate allele, with at least two of them on each strand. Moreover, SNPs that were less than 2 bp away from an indel were excluded from the analysis. Similarly, only indels 3 bp apart of other indels were kept.

All sites with insufficient coverage to identify a site as variant or reference were excluded (marked as 'N'), as were those in or within 100 bases of PE/PPE genes, or in insertion sequences or phages. SNPs present in the alignment with at least 90% frequency were used to generate a pseudoalignment of equal length to the H37Rv. Samples with more than 10% of the alignment represented by ambiguous bases were excluded. Those positions with more than 10% of ambiguous bases across all the samples were also removed. To avoid bias on the tree structure,

positions known to be associated with drug resistance were not included.

A more permissive variant calling pipeline was used to identify *mmpR5* variants, as they are often present at < 100% frequency with a high incidence of frameshift mutations. Here we instead employed FreeBayes v1.2 [64] to call all variants present in the *mmpR5* gene (or up to 100 bases upstream) that were present at ≥ 5% frequency (alternate allele fraction − *F* 0.05) and supported by at least four reads including one on each strand. Using this more permissive variant calling strategy we also systematically screened for all mutations in the efflux pump proteins *mmpS5-mmpL5* operon.

Classification of resistance variants

All raw fastq files were screened using the rapid resistance profiling tool TBProfiler [65, 66] against a curated whole genome drug resistance mutations library. This allowed rapid assignment of polymorphisms associated with resistance to different antimycobacterial drugs and categorisation of MDR and XDR *Mtb* status (Additional file 2: Fig S2), together with statistical assessment of the co-occurrence of mutations conferring resistance to different drug classes (Additional file 2: Fig S3–S7).

Classification of *mmpR5* variants

The diverse range of *mmpR5* variants and paucity of widespread MIC testing means limited data is available to infer the phenotypic consequences of identified *mmpR5* variants. This was true of our dataset aside from a subset of data sampled in Peru for which 30 L4 isolates from Peru were subjected to MIC testing using the UKMYC6 plate and a further nine were evaluated for MICs reported by the Cryptic consortium [67]. The approach we used was to assign whether nonsynonymous variants confer a normal or raised MIC based on published phenotypic tests for strains carrying that variant. A full list of the literature reports used to classify each mutation is provided in Additional file 2: Table S3. We also introduced an intermediate category to describe isolates with MICs at the critical concentration (e.g. 0.25 μg/mL on Middlebrook 7H11 agar), where there is an overlap of the MIC distributions of *mmpR5* mutated and wild-type isolates with uncertain clinical implications [36]. We assumed that all other disruptive frameshift and stop mutations would confer resistance considering the role of *mmpR5* as a negative repressor, where loss of function should lead to efflux pump overexpression, unless evidence existed in the literature or at other relevant sites in the genome to suggest otherwise. This allowed us to identify two frameshifts of currently unclear effect (Additional file 2: Table S3). All other promoters and previously unreported missense mutations were categorised as unknown (Additional file 2: Table S3). We identified cases of *mmpR5* variants in genomes collected prior to 2007 (Additional file 2: Table S4). Where *mmpR5* mutations were accompanied by an *mmpS5* or *mmpL5* loss of function mutation, we recorded these (Additional File 2: Fig S8–S9, Table S5) and assumed that they would confer susceptibility (or hypersusceptibility) to bedaquiline [23]. Resistance profiles of sequences containing *mmpR5* variants were denoted as either "S" for susceptible, "RR" for rifampicin-resistant and "preXDR" for fluoroquinolone-resistant. For mutations for which phenotypic status could not be ascertained a machine learning model was employed to explore potentially predictive features using a set of *mmpR5* variants of known phenotypic effect (Additional file 2: Note S1; Additional file 2: Fig S10).

Global phylogenetic inference

The alignments for phylogenetic inference were masked for the *mmpR5* region using bedtools v2.25.0. All variant positions were extracted from the resulting global phylogenetic alignments using snp-sites v2.4.1 [68], including a L4 outgroup for the L2 alignment (NC_000962.3) and a lineage 3 (L3) outgroup for the L4 alignment (SRR1188186). This resulted in a 67,585 SNP alignment for the L4 dataset and 29,205 SNP alignment for the L2 dataset. A maximum likelihood phylogenetic tree was constructed for both SNP alignments using RAxML-NG v0.9.0 [69] specifying a GTR + G substitution model, correcting for the number of invariant sites using the ascertainment flag (ASC_STAM) and specifying a minimum branch length of 1×10^{-9} reporting 12 decimal places (–precision 12).

Estimating the age of emergence of *mmpR5* variants

To test whether the resulting phylogenies can be time-calibrated we first dropped the outgroups from the phylogeny and rescaled the trees so that branches were measured in unit of substitutions per genome. We then computed a linear regression between root-to-tip distance and the time of sample collection using BactDating [70], which additionally assesses the significance of the regression based on 10,000 date randomisations. We obtained a significant temporal correlation for both the L2 and L4 phylogenies, both with and without imputation of dates for samples with missing metadata (Additional file 2: Fig S11).

We employed the Bayesian method BactDating v1.01 [70], run without updating the root (updateRoot = F), a mixed relaxed gamma clock model and otherwise default parameters to both global datasets. The MCMC chain was run for 1×10^7 iterations and 3×10^7 iterations. BactDating results were considered only when MCMC chains converged with an Effective Sample Space (ESS) of at

least 100. The analysis was applied to the datasets both with and without considering imputed and non-imputed collection dates (Additional file 2: Table S6).

To independently infer the evolutionary rates associated with each of our datasets, we sub-sampled both the L4 and L2 datasets to 200 isolates, selected so as to retain the maximal diversity of the tree using Treemmer v0.3 [71]. As before, we excluded all variants currently implicated in drug resistance from the alignments. This resulted in a dataset for L4 comprising 25,104 SNPs and spanning 232 years of sampling and for L2 comprising 8221 SNPs and spanning 24 years of sampling. In both cases the L3 sample SRR1188186 was used as an outgroup given this has an associated collection date. Maximum likelihood trees were constructed using RaXML-NG v0.9.0 [69], as previously described, and a significant temporal regression was obtained for both sub-sampled datasets (Additional file 2: Fig S12).

BEAST2 v2.6.0 [72] was run on both subsampled SNP alignments allowing for model averaging over possible choices of substitution models [73]. All models were run with either a relaxed or a strict prior on the evolutionary clock rate for three possible coalescent demographic models: exponential, constant and skyline. To speed up the convergence, the prior on the evolutionary clock rate was given as a uniform distribution (limits 0 to 10) with a starting value set to 10^{-7}. In each case, the MCMC chain was run for 500,000,000 iterations, with the first 10% discarded as burn-in and sampling trees every 10,000 chains. The convergence of the chain was inspected in Tracer 1.7 and through consideration of the ESS for all parameters (ESS > 200). The best-fit model to the data for these runs was assessed through a path sampling analysis [74] specifying 100 steps, 4 million generations per step, alpha = 0.3, pre-burn-in = 1 million generations, burn-in for each step = 40%. For both datasets, the best supported strict clock model was a coalescent Bayesian skyline analysis. The rates (mean and 95% HPD) estimated under these subsampled analyses (L2 7.7×10^{-8} [$4.9\times10^{-8}-1.03\times10^{-7}$] substitutions per site per year; L4 7.1×10^{-8} [$6.2\times10^{-8}-7.9\times10^{-8}$] substitutions per site per year) were used to rescale the maximum likelihood phylogenetic trees generated across the entire L2 and L4 datasets, by transforming all branch lengths of the tree from per unit substitution to per unit substitutions per site per year using the R package Ape v5.3 [75]. This resulted in an estimated tMRCA of 1332CE (945CE–1503CE) for L2 and 853CE (685CE–967CE) for L4 (Fig. 2).

The resulting phylogenetic trees were visualised and annotated for place of geographic sampling and *mmpR5* variant status using ggtree v1.14.6 [76]. All nonsynonymous/frameshift mutations in *mmpR5* were considered, with the phenotypic status assigned in Additional file 2: Table S3. For the purpose of this analysis, and to be conservative, 'unknown' variants classified using XGBoost were still considered 'unknown' (Additional file 2: Note S1). Clades carrying shared variants in *mmpR5* were identified and the distributions around the age of the node (point estimates – mean – and 95% HPDs) were extracted from the time-stamped phylogeny. For isolated samples (single emergences) exhibiting variants in *mmpR5*, the time of sample collection was extracted together with the date associated with the upper bound on the age of the next closest node of the tree, allowing for the mutation to have occurred anywhere over the length of the terminal branch (Fig. 3, Additional file 2: Fig S13–S14). For the Peruvian clade Bayesian skyline analysis was implemented through the skylineplot analysis functionality available in Ape v5.3 [75] (Additional file 2: Fig S15).

Results

The global diversity of *Mtb* lineage L2 and L4

To investigate the global distribution of *Mtb* isolates with variants in *mmpR5*, we curated two large datasets of whole genomes from the two dominant global lineages L2 and L4. Both datasets were enriched for samples with variants in *mmpR5* following a screen for variants in public sequencing repositories (see Methods) and retaining those samples uploaded with accompanying full metadata for geolocation and time of sampling (Fig. 1, Additional file 1: Table S1, Additional file 2: Fig S1, Table S2). The final L2 dataset included 1,514 isolates collected over 24.5 years (between 1994 and 2019) yielding 29,205 SNPs. The final L4 dataset comprised 2,168 sequences collected over 232 years, including three samples from eighteenth century Hungarian mummies [49], encompassing 67,585 SNPs. Both datasets included recently generated data from South Africa (155 L2, 243 L4) [50, 51] and new whole genome sequencing data from Peru (9 L2, 154 L4).

Consistent with previous studies [44, 77, 78], both datasets are highly diverse (Fig. 2). As a nonrecombining clonal organism, identification of mutations in *Mtb* can provide a mechanism to predict phenotypic resistance from a known panel of genotypes [65, 79]. Based on genotypic profiling [65], 911 strains within the L2 dataset were classified as MDR-TB (60%) and 295 (20%) as XDR-TB. Within the L4 dataset, 911 isolates were classified as MDR-TB (42%) and 115 as XDR-TB (5%). The full phylogenetic distribution of resistance profiles is provided in Additional file 2: Fig S2. As is commonplace with genomic datasets, the proportion of drug-resistant strains exceeds their actual prevalence, due to the

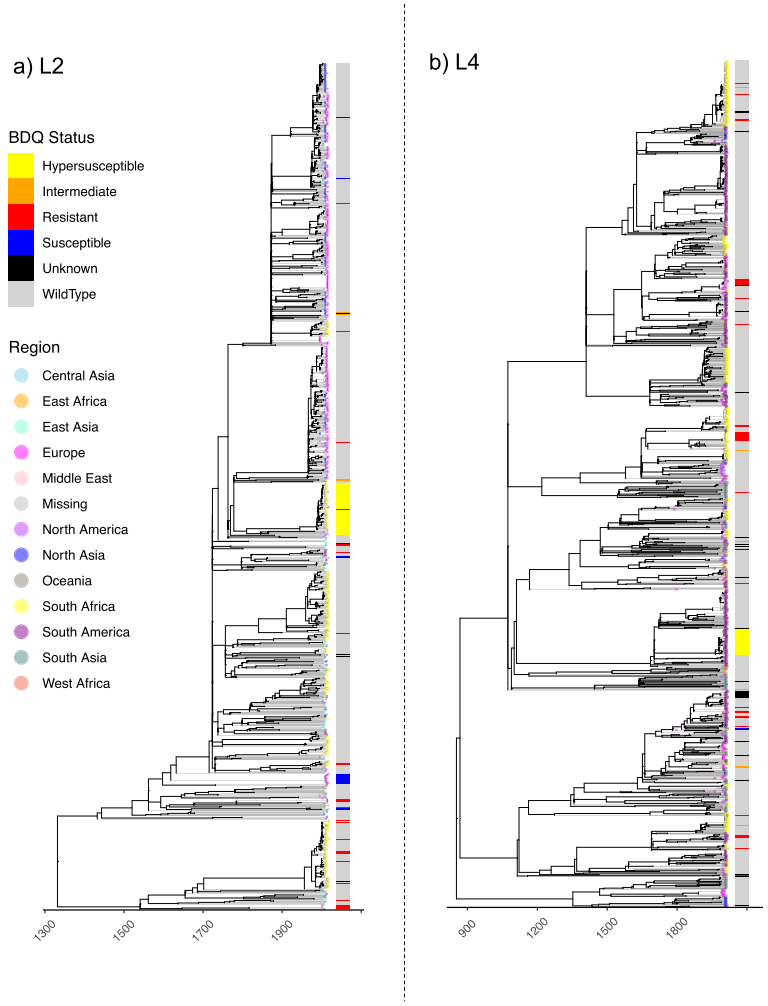

Fig. 2 Global time-calibrated *Mtb* phylogenies. Inferred dated phylogenies (*x*-axis) for the **a** lineage 2 and **b** lineage 4 datasets. Tips are coloured by the geographic region of sampling as given in the legend. The bar provides the assessed phenotype (colour) based on assignment of nonsynonymous/frameshift variants in *mmpR5*

overrepresentation of drug-resistant isolates in public sequencing repositories.

Both the L2 and L4 phylogenetic trees displayed a significant temporal signal following date randomisation (Additional file 2: Fig S11), making them suitable for time-calibrated phylogenetic inference [72]. We estimated the time to the Most Recent Common Ancestor (tMRCA) of both datasets using a Bayesian tip-dating analysis (BEAST2) run on a representative subset of genomes from each dataset (see "Methods", Additional file 2: Fig S12, Table S6). For the final temporal calibration of the L2 dataset, we applied an estimated clock rate of 7.7×10^{-8} ($4.9\times10^{-8}-1.03\times10^{-7}$) substitutions per site per year, obtained from the subsampled BEAST2 [48] analysis, to the global maximum likelihood phylogenetic tree. This resulted in an estimated tMRCA of 1332CE (945CE–1503CE HPD intervals). Using the same approach for the L4 dataset we estimated a clock rate of 7.1×10^{-8} (6.2×10^{-8} – 7.9×10^{-8} HPD intervals) substitutions per site per year resulting in an estimated tMRCA of 853CE (685CE–967CE HPD intervals) (Fig. 2). We observed a slightly higher, yet statistically not significant, clock rate in L2 compared to L4 (Additional file 2: Table S6), with all estimated substitution rates falling largely in line with previously published estimates [80].

Identification of variants in *mmpR5*

Since *atpE* and *pepQ* bedaquiline RAVs are found at low prevalence (1 L2 isolate [0.03%] and 18 L4 isolates [0.49%] [35]), we focused on characterising mutations in *mmpR5*. In total we identified the presence of non-synonymous and promoter *mmpR5* variants in 437 sequences (193 L2 [12.8%], 244 L4 [11.3%]). We classified all identified frameshift, non-synonymous and promoter mutations

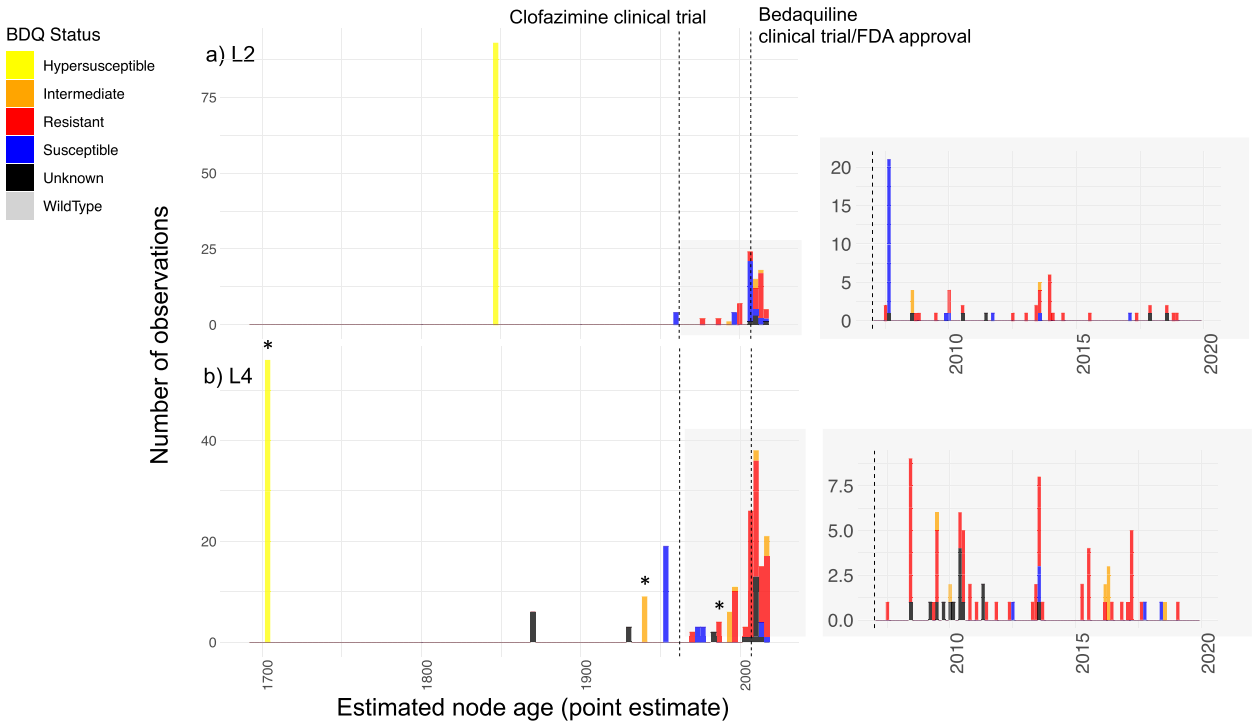

Fig. 3 Estimated age of emergence of *mmpR5* nonsynonymous/frameshift variants. Inferred point estimates for the age of emergence of clades with *mmpR5* variants for the lineage 2 (**a**) and lineage 4 (**b**) datasets, including a zoomed-in reproduction of the period from 2007 to 2020. *Y*-axis provides the absolute number of sequences descending from the identified and dated nodes. The *mmpR5* RAV status is given by the colour as defined in the legend at left. *indicates phenotypic data available for considered isolates that are supportive of MIC classification (see text). The full mutation timelines are provided in Additional file 2: Fig S13–S14 and Additional file 3: Table S7

in *mmpR5*, based on an evaluation of their phenotypic impact through review of published literature and associated *mmpL5* mutations, into six phenotypic categories for bedaquiline susceptibility: wild type, hypersusceptible, susceptible, intermediate, resistant, and unknown (full references available in Additional file 2: Table S3, Fig S3–S7). Across both lineages, 180 sequences were considered as bedaquiline resistant (i.e. classified as intermediate or resistant). The most frequently observed variants are listed in Table 1.

We identified a significant relationship between the presence of *mmpR5* variants and drug resistance status in both the L2 and L4 datasets (Additional file 2: Fig S6–S7), though in both cases we identified otherwise fully phenotypically susceptible isolates carrying *mmpR5* RAVs. Notably, we identified 25 sequenced isolates carrying nonsynonymous/frameshift variants in *mmpR5* and its promoter uploaded with collection dates (or permuted collection dates) prior to the first clinical trials for bedaquiline in 2007. This comprised ten L2 isolates collected before 2007, of which eight harboured variants previously associated to phenotypic bedaquiline resistance (RAVs). For L4 we identified 15 sequences with

Table 1 Frequency of all *mmpR5* variants occurring ≥ 5 times in dataset and their associated resistance classification. Where co-existing *mmpL5* mutations were identified these are indicated

Variant	Associated phenotype	L2	L4	Total
c-11a	Hypersusceptible	93		93
Ile67fs + *mmpL5* Arg202fs	Hypersusceptible		65	65
Asp5Gly	Susceptible	20	3	23
Met146Thr	Resistant	2	20	22
Ile67fs	Resistant	5	17	22
Leu40Val	Susceptible		19	19
Arg90Cys	Intermediate	2	9	11
Glu49fs	Resistant	2	8	10
Val20Ala	Intermediate	1	6	7
Ala59Val	Resistant	7		7
Met1Ala	Resistant	6		6
Gly121Arg	Resistant	5	1	6
Asp141fs	Unknown		6	6
Asn98Asp	Resistant		6	6
Arg96Gly	Resistant		5	5
Arg109Leu + Arg156fs	Resistant		5	5

mmpR5 variants predating 2007, of which six have been previously classified as carrying mutations conferring a bedaquiline resistance phenotype above wild-type ('intermediate') (Fig. 1c–d, Additional file 2: Table S4).

Within the datasets, we identified one L2 isolate (ERR2677436 sampled in Germany in 2016) which already had two *mmpR5* RAVs at low allele frequency — Val7fs (11%) and Val20Phe (20%) — and also contained two low-frequency *atpE* RAVs: Glu61Asp (3.2%) and Alal63Pro (3.7%) [50]. We also identified three isolates obtained in 2007–08 from separate but neighbouring Chinese provinces carrying the *Rv1979c* Val52Gly, which has been suggested to be associated with clofazimine resistance in a study from China [25] but was associated with a normal MIC in another [41], with its role in resistance remaining unclear [31]. Furthermore, frameshift and premature stop mutations in *pepQ* have been previously associated with bedaquiline and clofazimine resistance. In this dataset, we identified 18 frameshift mutations in *pepQ* across 11 patients, one of which also had a *mmpR5* frameshift mutation. In one isolate the *pepQ* frameshift occurred at the Arg271 position previously reported to be associated with bedaquiline resistance [19].

Thirty-four genomes harboured nonsynonymous *mmpR5* variants of unknown phenotypic effect (7 L2, 27 L4), corresponding to 22 unique mutations or combinations of mutations. To assess properties associated to RAVs which may be useful predictors of the phenotypic effect of these unknown variants we employed a machine learning approach, providing a foundation for further exploration of genomic features associated to RAV status (see Additional file 2: Note S1), although this was not used for the categorisation of RAVs in the main analysis.

The time to emergence of *mmpR5* variants

To estimate the age of the emergence of different *mmpR5* non-synonymous variants, we identified all nodes in each of the L2 and L4 global time-calibrated phylogenies delineating clades of isolates carrying a particular *mmpR5* variant (Fig. 3, Additional file 3: Table S7). For the L2 dataset, we identified 49 unique phylogenetic nodes where *mmpR5* mutations emerged. The point estimates for these nodes ranged from March 1845 to November 2018. Eight nonsynonymous/frameshift variants in *mmpR5*, including four bedaquiline RAVs (Met139Ile, Cys46fs, Ala59Val, Asn98fs) and one case expected to lead to an intermediate phenotype (Arg90Cys), were estimated to have emergence dates (point estimates) predating the first bedaquiline clinical trial in 2007 (Additional file 2: Fig S13).

For the L4 dataset, we identified 84 unique nodes where *mmpR5* mutations emerged. The point estimates for these nodes ranged from September 1701 to January 2019 (Fig. 3, Additional file 2: Fig S14). Nineteen *mmpR5* mutations, including five unique bedaquiline RAVs (Gln22Arg, Asn98Asp, Ile67fs × 2, Arg96Gly, Met146Thr) and three predicted to have an intermediate phenotype (Arg90Cys, Val20Ala, Ser53Leu), were estimated to have emerged prior to 2007. We estimate that Arg90Cys emerged between 1930 and 1947, an example of the likely circulation of variants which lead to a response to bedaquiline above wild-type levels pre-existed the first clinical trials for clofazimine in the 1960s. While we identified no nodes with a secondary emergence of *mmpR5* nonsynonymous/frameshift mutations across the L4 dataset, eight nodes were identified in the L2 dataset where a clade already carrying a nonsynonymous/frameshift variant in *mmpR5* subsequently acquired a second nonsynonymous/frameshift mutation.

In the L4 dataset, we noted one large clade of 66 samples, predominantly collected in Peru (henceforth Peruvian clade), which all carry the Ile67fs *mmpR5* mutation, which when observed independently has been linked to bedaquiline resistance [37, 81, 82]. While it is not inconceivable that multiple independent emergences of Ile67fs occurred in this clade, the more parsimonious scenario is a single ancestral emergence. We estimate the time of this emergence to 1702 (1657–1732) (Fig. 3, Additional file 2: Fig S14–S15). Of significance, we identified a frameshift mutation in the adjacent *MmpL5* efflux pump (Arg202fs) in isolates from this Peruvian clade, the protein whose overexpression mediates bedaquiline resistance following loss-of-function of the MmpR5 regulatory protein. This frameshift, which leads to a premature stop codon at amino acid 206, is expected to counteract the otherwise resistance-conferring mutation. This epistatic interaction restoring bedaquiline susceptibility has recently been described elsewhere [23, 41]. The *mmpL5* frameshift mutation was present in all isolates in the Peruvian clade bar one (ERR7339051/LN3756) which had *mmpL5* Arg202Leu. This event of reading-frame restoration is likely explained by a recent secondary duplication of a T downstream of the initial deletion (777876 GGCAT > GGAT, GGAT > GGATT). We considered the phenotype of this strain as unknown. No other *mmpL5* mutations were found in any isolate containing *mmpR5* mutations within this study though we did identify a low prevalence of variants in *mmpL5* and *mmpS5* independent of *mmpR5* mutations across both lineages (Additional file 2: Fig S8–S9).

We also noted a tendency for *mmpR5* mutations to emerge in clades that also displayed genetic markers of rifampicin resistance. This was more common in mutations emerging after 2007 (77.2%) than before 2007 (58.3%). Most of the oldest Ile67fs Peruvian clade was

rifampicin resistant (58/66 samples), with the remaining samples demonstrating only isoniazid resistance.

Phenotypic validation of *mmpR5* variants

Given documented epistasis as a modulator of bedaquiline resistance phenotype, we performed MIC testing on a selection of available isolates and identified further MICs that have been recently published as part of the Cryptic consortium using microtitre plates (Additional file 3: Table S7) [67, 83]. The epidemiological cut-off (ECOFF, defined as MIC of 95–99% of wild-type isolates) for bedaquiline has been proposed to be 0.12 or 0.25 μg/mL depending on the method used, although the final decision was to use an ECOFF of 0.25 μg/mL [83].

We were able to identify 30 L4 isolates from Peru (including members of the aforementioned Peruvian clade) for MIC testing, and a further 9 MICs for L4 that had recently been published by the Cryptic consortium [67]. For the oldest dated *mmpR5* mutation emergence — the L4 Ile67fs mutation in Peruvian isolates with an associated MRCA estimated to 1701—10/11 (90.9%) had a MIC below the lower proposed ECOFF of 0.12 μg/mL, presumably due to the co-existing *mmpL5* loss of function mutation. Hence, we denote isolates from this clade as having a hypersusceptible phenotype. The second oldest predicted resistance mutation (Arg90Cys, dated to 1940) was however associated with MICs ≥ 0.12 μg/mL in 6/7 (85.7%) instances, and in 3/4 (75%) instances for the third oldest predicted resistance-associated mutation for which data were available (Asn98Asp, dated to 1987). These MICs are above the wild-type range, if not formally classified as resistant. Clades with associated MIC confirmation are highlighted in Fig. 3b.

Discussion

Our work establishes that the emergence of variants in *mmpR5*, including bedaquiline RAVs, is not solely driven by bedaquiline use. We identified up to 11 events where RAVs (classified as resistant) emerged prior to the first clinical trials of bedaquiline in 2007 and a further four cases of variants emerging prior to the clinical use of bedaquiline which are expected to give rise to an intermediate phenotype. These are highlighted red and orange respectively in Additional File 3: Table S7, not including the oldest emergence of Ile67fs as its resistant phenotype is negated by the epistatic interaction with *mmpL5* mutations. Phylogenetic inference estimated the oldest clade containing *mmpR5* mutations, composed mostly of samples from Peru carrying *mmpR5* Ile67fs and *mmpL5* Arg202fs, to have emerged around 1702 (1657–1732). We identify two further early emergences of *mmpR5* mutations, estimated to 1871 and 1940 (Asp141fs and Arg90Cys; point estimates), with samples from the latter clade confirmed to have MICs above the wild-type range justifying classification of an intermediate phenotype. Asp141fs has been detected in a bedaquiline susceptible isolate by Rancoita et al. [84] (SRR6479538), accompanied by an I948V *mmpL5* mutation. In the latest WHO catalogue v2, Asp141fs is classified as 'interim association with resistance' to bedaquiline [35]. Thus, the phenotypic implications of Asp141fs remain unclear.

Together our work suggests the likely circulation of variants exhibiting at least borderline resistance even prior to the first clinical trials for clofazimine. Our phylogenetic inference method, which points to multiple emergences of *mmpR5* nonsynonymous/frameshift variants predating the use of bedaquiline, is also confirmed by the direct observation of eight *Mtb* genomes carrying *mmpR5* RAVs sampled prior to 2007 (Additional File 2: Table S4). We also identified, within the aforementioned Peruvian clade, a frameshift mutation in *mmpL5*, which seemed to counteract the otherwise resistance-associated phenotype conferred by *mmpR5* Ile67fs through an epistatic interaction (MIC < 0.12 μg/mL). While Ile67fs is central for bedaquiline resistance in *Mtb*, and this mutation has clearly emerged well prior to the use of bedaquiline and clofazimine in this clade, its phenotypic impact is influenced by the strain genetic background. This observation, together with the uncertainty surrounding Asp141fs, suggests that an extension of resistance prediction frameworks, for example along the lines of the machine learning approach we implement here, that jointly considering mutations in both *mmpR5* and *mmpL5* may help to better ascertain bedaquiline resistance status. Indeed, the WHO now formally recognises the importance of epistasis when interpreting the phenotypic impact of *mmpR5* variants on bedaquiline resistance [35]. Though additional linked genomic and phenotypic (MIC) data will be required to develop a model with satisfactory predictive power.

We identified other localised clusters with *mmpR5* mutations, reinforcing the need for concern even in situations where such mutations are globally rare. This included Met146Thr carrying isolates found in lineage 4 isolates from Eswatini. Met146Thr mutations have been previously associated with a clade that has a rifampicin-resistance conferring mutation located outside of the canonical rifampicin-resistance determining region, and these isolates exhibit elevated bedaquiline MICs [85]. The emergence of the Met146Thr mutation has previously been dated to have emerged in approximately 2003 [23, 41, 85]. This is in reasonable agreement with our analysis on a much larger dataset which inferred an emergence in 2005.6 (95% confidence intervals 2004.8–2006.0). The long-standing presence of variants implicated in resistance and borderline resistance to bedaquiline predating

the use of the drug and at high prevalence in geographically notable cases is of concern, as it suggests that non-synonymous mutations in *mmpR5* exert little fitness cost.

Together, our work suggests the existence of pre-existent reservoirs of bedaquiline-resistant *Mtb*. These may have been selected for through historic clofazimine use, though we inferred at least one case of intermediate resistance to bedaquiline emerging as early as 1930–1947. We note that detected variants in *mmpR5* tend to exist in strains already displaying rifampicin resistance, though they are also found in otherwise fully susceptible strains (Additional File 3: Table S7). Together this suggests the important role of prior drug exposure in selecting for strains with pre-existing (cross-)resistance potential. This reservoir of putatively adaptive variants is expected to expand under drug pressure with the increasing use of bedaquiline and clofazimine in TB treatment. Further, these reservoirs may also pose a threat for other candidate TB agents from different drug classes that are also exported by *mmpS5* and *mmpL5* [19, 22, 64].

The identification of resistance variants occurring before the clinical use of a drug is not limited to *M. tuberculosis* and *mmpR5* alone. To illustrate, within *M. bovis*, there is evidence indicating that the *pncA* H57D mutation, which is associated with resistance to pyrazinamide (PZA), emerged approximately 900 years ago, providing inherent resistance to PZA in the majority of *M. bovis* [86]. Similarly, variations in intrinsic susceptibility to pretomanid have been observed across the MTBC, including *Mtb* lineages, even without prior exposure to nitroimidazoles [87]. It is likely that there are numerous other instances of such loss of function mutations with minimal or no impact on fitness, similar to the case of *mmpR5*. Furthermore, the existence of antimicrobial resistance in different forms has persisted throughout the natural history of various bacteria [88].

Nevertheless, it is crucial to determine the age and diversity of variants that have been implicated in drug resistance to gain a better understanding of the potential for widespread resistance as a contemporary challenge. We identified a large number of different *mmpR5* nonsynonymous/frameshift variants across both of our *Mtb* lineage cohorts; 46 in L2 and 67 in L4. This suggests the mutational target leading to bedaquiline resistance is wider than for most other current TB drugs and raises concerns about the ease with which bedaquiline resistance can emerge during treatment. It is further concerning that resistance to the new class of nitroimidazole drugs, such as pretomanid and delamanid, is also conferred by loss of function mutations in any of at least six genes, suggesting that they may also have a low barrier to resistance [89], though current studies suggest acquired resistance rates are low [39].

While we identified many non-synonymous variants in *mmpR5*, only one (Ile67fs) has been previously definitively linked to resistance. We acknowledge that several of our detected variants have no associated MIC values available in the literature and are thus currently not fully phenotypically validated, and we treat them as 'unknown' in this work. As some of these 'unknown' variants will likely be associated with a phenotype in the future, and possibly confirmed as RAVs, a subset of the early emerging 'unknown' variants may turn out to represent additional instances of bedaquiline resistance predating the use of the drug. Though, determining the phenotypic consequences of *mmpR5* variants remains challenging as reports correlating MICs to genotypes remain scarce. Moreover, at least four different methods are used to determine MICs, some of which do not have associated critical concentrations. Even where critical concentrations have been set, different isolates carrying the same mutations can fall on either side of the breakpoint between wild-type and drug-resistant [36]. The choice of breakpoints has also been called into question [90, 91] underlying the need to validate broth microdilution assays comprehensively [92].

Prediction of phenotypic bedaquiline resistance from genomic data is further complicated by the existence of hypersusceptibility variants. For example, the c-11a variant located in the promoter of MmpR5, which appears to increase susceptibly to bedaquiline [40], was observed to be fixed throughout a large clade within L2. The early emergence of this variant and its geographical concentration in South Africa and Eswatini may suggest the role of non-pharmacological influences on MmpR5 which regulates multiple MmpL efflux systems [20]. Further, analysis of hypersusceptibility is limited by the truncated lower MIC range of the UKMYC microtitre plates, with many isolates giving MICs below the lower end of the measured range. While large-scale genotype/phenotype analyses will likely support the development of rapid molecular diagnostics, targeted or whole genome sequencing, at reasonable depths, may provide the only opportunity to detect all possible *mmpR5* RAVs, and possible co-occurring mutations, in clinical settings.

Bedaquiline resistance can also be conferred by other RAVs including in *pepQ* (bedaquiline and clofazimine), *atpE* (bedaquiline only) [82] and *Rv1979c* (clofazimine only). We only found *atpE* RAVs at low allele frequency in one patient who also had *mmpR5* variants (sample accession ERR2677436), which is in line with other evidence suggesting they rarely occur in clinical isolates, likely due to a high fitness cost. Likewise, we only identified *Rv1979c* RAVs in three patients in China, although there were other variants in *Rv1979c* for which the ability

to cause phenotypic resistance has not been previously assessed. Frameshift *pepQ* mutations that are potentially causative of resistance were identified in 11 cases, in keeping with its possible role as an additional rare resistance mechanism.

Conclusion

Our findings, of reservoirs of *mmpR5* RAVs predating the therapeutic use of bedaquiline, are of high clinical relevance as the presence of *mmpR5* variants during therapy in clinical strains has been associated with substantially worse outcomes in patients treated with drug regimens including bedaquiline [37]. Although it is uncertain what the impact of *mmpR5* RAVs is on outcomes when present prior to treatment [93, 94], it is imperative to monitor and prevent the wider transmission of bedaquiline-resistant clones, particularly in high MDR/XDR-TB settings. Early evaluation of new TB drug candidates entering clinical trials will also be vital given early data suggesting possible cross-resistance for DprE1 inhibitors such as macozinone [22]. The large and disparate set of mutations in *mmpR5* we identified, with differing phenotypes and some having been in circulation historically, adds further urgency to the development of rapid drug susceptibility testing for bedaquiline to inform effective treatment choices and mitigate the further spread of DR-TB.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01289-5.

Additional file 1: Table S1. Full metadata, including estimated date of collection where relevant, and predicted bedaquiline resistance status for accessions included in the Lineage 2 and Lineage 4 datasets (see sheets Lineage 2 and Lineage 4)

Additional file 2: Fig S1. Distribution of difference (in decimal years) of collection date to release date for sequences in the lineage 2 (L2) and lineage 4 (L4) dataset. For lineage 4, having excluded samples from three eighteenth century mummies, an average of 6.9 years (0.6–19.1 years 95% CI) was obtained. For lineage 2, an average lag-time of 4.7 years (0.5–12.6 years 95% CI) was obtained. **Fig S2.** Maximum likelihood phylogenetic tree of the lineage 2 (a) and lineage 4 (b) datasets. Tip colours provide the country of sample collection and outer bars give the resistance to four antimycobacterial drugs: fluroquinolones (FLQ), isoniazid (INH), kanamycin (KAN) and rifampicin (RMP). **Fig S3.** Number (count) of *mmpR5* variants split by predicted phenotype (x-axis) for each geographic region included in each of the lineage 2 (a) and lineage 4 b) datasets. **Fig S4.** Co-occurrence count of *mmpR5* variant phenotype predictions (x-axis) and resistance status a) and genotypic resistances b) for the lineage 2 dataset. **Fig S5.** Co-occurrence count of *mmpR5* variant phenotype predictions (x-axis) and resistance status a) and genotypic resistances b) for the lineage 4 dataset. **Fig S6.** Contingency tables and χ^2 results for the distribution of susceptible, MDR and XDR *Mtb* amongst strains carrying variants in *mmpR5* in the lineage 2 dataset. Plots provide the squared standardised residuals contributing to the rejection of the null hypothesis. The colour intensity and size of the circle is proportion to the contribution with positive displayed in blue and negative in red. **Fig S7.** Contingency tables and χ^2 results for the distribution of susceptible, MDR and XDR *Mtb* amongst strains carrying variants in *mmpR5* in the lineage 4 dataset. Plots provide the squared standardised residuals contributing to the rejection of the null hypothesis. The colour intensity and size of the circle is proportion to the contribution with positive displayed in blue and negative in red. **Fig S8.** Phylogenetic distribution of LOF mutations identified in *mmpL5* and *mmpS5* identified in L2 isolates. Phylogeny is provided with tip colours according to inferred bedaquiline resistance status. Heatmap provides colour for presence of a mutation as ordered by the vertical columns. **Fig S9.** Phylogenetic distribution of LOF mutations identified in *mmpL5* and *mmpS5* identified in L4 isolates. Phylogeny is provided with tip colours according to inferred bedaquiline resistance status. Heatmap provides colour for presence of a mutation as ordered by the vertical columns. **Fig S10.** Summary plot of SHAP values. Each point represents the SHAP value of a single prediction for a particular feature. Points are stacked vertically using density estimation. 'WT', 'mutant' and 'MW' denotes the wild type amino acid, amino acid variant and molecular weight (Da) respectively. 'Property' refers to whether an amino acid was non-polar, polar, positively charged or negatively charged. 'Ligand_binding', 'dna_binding' and 'dimerisation' refer to whether the amino acid residue is involved in ligand binding, DNA binding or dimerisation. 'Position' refers to the integer 5'-3' position of the variant. Positive SHAP values imply an increase in the predicted probability of resistance due to the presence of the feature. **Fig S11.** Linear regressions of root-to-tip distance (y-axis) versus sampling dates (x-axis) for global *Mtb* datasets; lineage 2 (a-b) and lineage 4 (c-d). Regressions are performed both without (a-c) and with (b-d) imputation of missing dates (see Methods). Here, the *p*-value (tiprandomisation test) is calculated by fitting a linear regression to the root to tip distance vs sampling date for 10,000 randomisations and adding the number of randomised fits that present a better regression coefficient than the real data (divided by 10,000). **Fig S12.** Sub-sampled datasets temporal regression. The lineage 2 data covered 24 years of molecular evolution. The lineage 4 dataset comprised 232 years of evolution. **Fig S13.** Full mutational timeline for the estimated date of emergence (x-axis) and confidence intervals of nodes with descendent tips carrying nonsynonymous variants in *mmpR5* in the lineage 2 dataset. All nonsynonymous variants are depicted. Confidence bars are coloured according to the region where the isolate was collected. Symbols provide the point estimates of the age of the node coloured by *mmpR5* predicted phenotype. Symbols are used for all mutations occurring in ≥5 isolates, in this case. Grey dashed lines provide the collection date of all sequenced isolates included in the analysis with *mmpR5* variants. Data available in Supplementary Table S7. **Fig S14.** Full mutational timeline for the estimated date of emergence (x-axis) and confidence intervals of nodes with descendent tips carrying nonsynonymous variants in *mmpR5* in the lineage 4 dataset. All nonsynonymous variants are depicted. Confidence bars are coloured according to the region where the isolate was collected. Symbols provide the point estimates of the age of the node coloured by *mmpR5* predicted phenotype. Symbols are used for all mutations occurring in ≥5 isolates. Grey dashed lines provide the collection date of all sequenced isolates included in the analysis with *mmpR5* variants. Data available in Supplementary Table S7. **Fig S15.** a) Lineage 4 Peruvian *mmpR5* Ile67fs + *mmpL5* Arg202fs carrying clade phylogeny, which has a tMRCA dating to 1702 (1657–1732). Phenotypic resistances for fluoroquinolones (FLQ), isoniazid (INH), kanamycin (KAN) and rifampicin (RMP) are provided as outer coloured rings. Most samples are from Peru (purple), though two samples are from Europe (Sweden and the Netherlands). b) Provides the generalised skyline plot estimate of effective population size through time based on the timed phylogeny of this clade. Grey lines provide the full skyline plot, black lines provide the coalescent intervals. The first clinical use of clofazimine and bedaquiline are provided by the axis at top. **Table S2.** Source of whole genome sequences included in the global lineage 2 and lineage 4 alignments following quality checks and as given in Table S1. The number in brackets designate those with *mmpR5* differing from wild type. **Table S3.** Classification of previously observed resistance associated with *mmpR5* variants identified in this study. **Table S4.** Sequence data identified with *mmpR5* variants predating 2007. Dates flagged with an asterisk (*) indicate those dates which have been permuted using the metadata of all samples (see Est. column of Supplementary Table S1). **Table S5.** Number of estimated emergence (homoplasic) events for major *mmpR5* variants considered. **Table S6.**

Summary of phylogenetic dating approaches applied to the L4 and L2 datasets. BactDating was run in both cases but failed to converge after $1e^7$ or $3e^7$ MCMC iterations for the L4 dataset. The subsampled BEAST2 runs highlighted yellow provided the highest likelihood following path sampling over all strict clock models and were run only on accessions with associated collection dates. These rates were applied to the maximum likelihood phylogenetic tree for temporal estimation of resistance emergence (main text Fig. 2, see Methods). **Table S7 [external excel document].** Inferred age of nodes and preceding nodes for L2 samples with *mmpR5* nonsynonymous variants. Cells are coloured as per phenotype annotations (see main text Fig. 1, 2, 3). Presence of *mmpL5* variants is noted, as are MICs where available. TBProfiler resistance profiles as either "S" for susceptible, "RR" for rifampicinresistant and "preXDR" for fluoroquinolone-resistant. **Table S8 [external excel document].** Predicted probability of bedaquiline resistance based on the amino acid properties of *mmpR5* variants following a machine learning predictive approach (see Methods and Supplementary Note 1). **Table S9.** Precision, recall, F1, AUPRC, sensitivity and specificity scores for gradient-boosted tree classifier. Standard deviation was calculated across the 10 outer loops of the nested crossvalidation protocol. **Supplementary Note 1.** Predicting bedaquline resistance using machine learning techniques.

Additional file 3: Table S7. Inferred age of nodes and preceeding nodes for L2 samples with *mmpR5* nonsynonymous variants. Cells are coloured as per phenotype annotations (see main text Fig. 1, 2, 3).

Additional file 4: Table S8. Predicted probability of bedaquiline resistance based on the amino acid properties of mmpR5 variants following application of a machine learning predictve approach (see Methods and Supplementary text S1).

Authors' contributions

LvD, CN and FB conceived and designed the study. JM, NP, AG, MO, AP, OBB, VE and LG provided sequence data. ATO, JP, MA, CCST and XD performed and advised on computational analyses. LvD, CN and FB wrote the manuscript with input from all co-authors. All authors read and approved the final manuscript.

Funding

CN and JM are supported by the Wellcome Trust (203583/Z/16/Z and 203919/Z/16/Z, respectively). LvD is supported by a UCL Excellence Fellowship. F.B. acknowledges support from the BBSRC (equipment grant BB/R01356X/1). FB additionally acknowledges the National Institute for Health Research University College London Hospitals Biomedical Research Centre. M.A. was supported by a Ph.D. scholarship from University College London. All authors acknowledge the UCL Biosciences Big Data equipment grant from BBSRC (BB/R01356X/1).

Availability of data and materials

Raw sequence data and full metadata for all newly generated isolates are available on NCBI Sequencing Read Archive under BioProject ID: PRJEB39837.

Declarations

Ethics approval and consent to participate

Ethical approval for sample collection and processing of the novel data made available in this work was obtained from the IRB of the Universidad Peruana Cayetano Heredia as part of previously published studies [56, 95]. Institutional approval was obtained from the Peruvian Ministry of Health. We note that individual patient consent was not sought as the data was collected and analysed anonymously. The research confirms to all aspects of the Helsinki declaration.

Consent for publication

Not applicable.

Competing interests

AP is currently employed by Janssen. Dr Pym's involvement with the research described herein precedes his employment at Janssen. The remaining authors declare that they have no competing interests.

Author details

[1]UCL Genetics Institute, University College London, Darwin Building, Gower Street, London, UK. [2]Division of Infection and Immunity, University College London, London, UK. [3]Africa Health Research Institute, Durban, South Africa. [4]Department of Medicine, Imperial College, London, UK. [5]Wellcome Trust Liverpool Glasgow Centre for Global Health Research, Liverpool, UK. [6]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. [7]CAPRISA MRC-HIV-TB Pathogenesis and Treatment Research Unit, Durban, South Africa. [8]TB Centre, London School of Hygiene & Tropical Medicine, London, UK. [9]Department of Medicine & Epidemiology, Columbia University Irving Medical Center, New York, NY, USA. [10]Division of Infectious Diseases and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway. [11]Laboratorio de Investigacion y Enfermedades Infecciosas, Universidad Peruana Cayetano Heredia, Lima, Peru. [12]Department of Infection, Immunity and Inflammation, Institute of Child Health, University College London, London, UK. [13]School of Life Sciences and Department of Statistics, University of Warwick, Coventry, UK.

Received: 6 January 2023 Accepted: 19 January 2024
Published online: 19 February 2024

References

1. World Health Organization. Global tuberculosis report 2022. (WHO, 2022).
2. Cegielski JP, et al. Multidrug-resistant Tuberculosis treatment outcomes in relation to treatment and initial versus acquired second-line drug resistance. Clin Infect Dis. 2016;62:418–30. https://doi.org/10.1093/cid/civ910.
3. World Health Organization. Global tuberculosis report 2019. (2019).
4. Andries K, et al. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science. 2005;307:223–7. https://doi.org/10.1126/science.1106753.
5. Diacon AH, et al. Multidrug-resistant tuberculosis and culture conversion with bedaquiline. N Engl J Med. 2014;371:723–32. https://doi.org/10.1056/NEJMoa1313865.
6. Food and Drug Administration. SIRTURO approval letter. Retrieved Jan 15, 2024, from https://www.accessdata.fda.gov/drugsatfda_docs/appletter/2012/204384orig1s000ltr.pdf.
7. Borisov SE. et al. Effectiveness and safety of bedaquiline-containing regimens in the treatment of MDR- and XDR-TB: a multicentre study. Eur Respir J 2017;49. https://doi.org/10.1183/13993003.00387-2017
8. Guglielmetti L. et al. Long-term outcome and safety of prolonged bedaquiline treatment for multidrug-resistant tuberculosis. Eur Respir J 2017;49. https://doi.org/10.1183/13993003.01799-2016
9. Olayanju O. et al. Long-term bedaquiline-related treatment outcomes in patients with extensively drug-resistant tuberculosis from South Africa. Eur Respir J 2018;51. https://doi.org/10.1183/13993003.00544-2018
10. Ndjeka N. et al. High treatment success rate for multidrug-resistant and extensively drug-resistant tuberculosis using a bedaquiline-containing treatment regimen. Eur Respir J 2018;52. https://doi.org/10.1183/13993003.01528-2018
11. World Health Organization. Module 4: treatment - drug-resistant tuberculosis treatment, 2022 update. (2022).
12. Conradie F, et al. Bedaquiline-Pretomanid-Linezolid regimens for drug-resistant Tuberculosis. N Engl J Med. 2022;387:810–23. https://doi.org/10.1056/NEJMoa2119430.
13. Berry C, et al. TB-PRACTECAL: study protocol for a randomised, controlled, open-label, phase II-III trial to evaluate the safety and efficacy of regimens containing bedaquiline and pretomanid for the treatment of adult patients with pulmonary multidrug-resistant tuberculosis. Trials. 2022;23:484. https://doi.org/10.1186/s13063-022-06331-8.
14. Paton NI, Cousins C, Suresh C. Treatment strategy for rifampin-susceptible tuberculosis. Reply N Engl J Med. 2023;388:2298. https://doi.org/10.1056/NEJMc2304776.
15. Manson AL, et al. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. Nat Genet. 2017;49:395–402. https://doi.org/10.1038/ng.3767.
16. Cohen KA, et al. Evolution of extensively drug-resistant Tuberculosis over four decades: Whole genome sequencing and dating analysis of

Mycobacterium tuberculosis isolates from KwaZulu-Natal. PLoS Med. 2015;12:e1001880. https://doi.org/10.1371/journal.pmed.1001880.
17. Eldholm V, Balloux F. Antimicrobial resistance in *Mycobacterium tuberculosis*: the odd one out. Trends Microbiol. 2016;24:637–48. https://doi.org/10.1016/j.tim.2016.03.007.
18. Huitric E, et al. Rates and mechanisms of resistance development in *Mycobacterium tuberculosis* to a novel diarylquinoline ATP synthase inhibitor. Antimicrob Agents Chemother. 2010;54:1022–8. https://doi.org/10.1128/AAC.01611-09.
19. Almeida D, et al. Mutations in *pepQ* confer low-level resistance to Bedaquiline and Clofazimine in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2016;60:4590–9. https://doi.org/10.1128/AAC.00753-16.
20. Andries K, et al. Acquired resistance of *Mycobacterium tuberculosis* to bedaquiline. PLoS ONE. 2014;9:e102135. https://doi.org/10.1371/journal.pone.0102135.
21. Hartkoorn RC, Uplekar S, Cole ST. Cross-resistance between clofazimine and bedaquiline through upregulation of MmpL5 in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2014;58:2979–81. https://doi.org/10.1128/AAC.00037-14.
22. Poulton NC, Azadian ZA, DeJesus MA, Rock JM. Mutations in *rv0678* confer low-level resistance to Benzothiazinone DprE1 inhibitors in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2022;66:e0090422. https://doi.org/10.1128/aac.00904-22.
23. Vargas R Jr, et al. Role of epistasis in Amikacin, Kanamycin, Bedaquiline, and Clofazimine resistance in *Mycobacterium tuberculosis* Complex. Antimicrob Agents Chemother. 2021;65:e0116421. https://doi.org/10.1128/AAC.01164-21.
24. Bloemberg GV, et al. Acquired resistance to Bedaquiline and Delamanid in therapy for Tuberculosis. N Engl J Med. 2015;373:1986–8. https://doi.org/10.1056/NEJMc1505196.
25. Xu J. et al. Primary Clofazimine and Bedaquiline resistance among isolates from patients with multidrug-resistant tuberculosis. Antimicrob Agents Chemother 2017;61. https://doi.org/10.1128/AAC.00239-17
26. Zimenkov DV, et al. Examination of bedaquiline- and linezolid-resistant *Mycobacterium tuberculosis* isolates from the Moscow region. J Antimicrob Chemother. 2017;72:1901–6. https://doi.org/10.1093/jac/dkx094.
27. de Vos M, et al. Bedaquiline microheteroresistance after cessation of Tuberculosis treatment. N Engl J Med. 2019;380:2178–80. https://doi.org/10.1056/NEJMc1815121.
28. Ghodousi A. et al. Acquisition of cross-resistance to Bedaquiline and Clofazimine following treatment for Tuberculosis in Pakistan. Antimicrob Agents Chemother 2019;63. https://doi.org/10.1128/AAC.00915-19
29. Polsfuss S, et al. Emergence of low-level delamanid and Bedaquiline resistance during extremely drug-resistant tuberculosis treatment. Clin Infect Dis. 2019;69:1229–31. https://doi.org/10.1093/cid/ciz074.
30. Mokrousov I, Akhmedova G, Polev D, Molchanov V, Vyazovaya A. Acquisition of bedaquiline resistance by extensively drug-resistant *Mycobacterium tuberculosis* strain of Central Asian outbreak clade. Clin Microbiol Infect. 2019;25:1295–7. https://doi.org/10.1016/j.cmi.2019.06.014.
31. Kadura S, et al. Systematic review of mutations associated with resistance to the new and repurposed *Mycobacterium tuberculosis* drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. J Antimicrob Chemother. 2020;75:2031–43. https://doi.org/10.1093/jac/dkaa136.
32. Roberts LW. et al. Repeated evolution of bedaquiline resistance in *Mycobacterium tuberculosis* is driven by truncation of mmpR5. bioRxiv, 2022.2012.2008.519610. 2022. https://doi.org/10.1101/2022.12.08.519610
33. Sonnenkalb L, et al. Bedaquiline and clofazimine resistance in *Mycobacterium tuberculosis*: an in-vitro and in-silico data analysis. Lancet Microbe. 2023;4:e358–68. https://doi.org/10.1016/S2666-5247(23)00002-2.
34. Ismail N, et al. Genetic variants and their association with phenotypic resistance to bedaquiline in *Mycobacterium tuberculosis*: a systematic review and individual isolate data analysis. Lancet Microbe. 2021;2:E604–16. https://doi.org/10.1016/S2666-5247(21)00175-0.
35. World Health Organization. Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. 2023. https://iris.who.int/handle/10665/374061. Accessed 31 Jan 2024.
36. World Health Organization. Technical report on critical concentrations for TB drug susceptibility testing of medicines used in the treatment of drug-resistant TB. 2018.
37. Nimmo C. et al. Bedaquiline resistance in drug-resistant tuberculosis HIV co-infected patients. Eur Respir J 2020;55. https://doi.org/10.1183/13993003.02383-2019
38. Martinez E, et al. Mutations associated with *in vitro* resistance to Bedaquiline in *Mycobacterium tuberculosis* isolates in Australia. Tuberculosis (Edinb). 2018;111:31–4. https://doi.org/10.1016/j.tube.2018.04.007.
39. Timm J, et al. Baseline and acquired resistance to bedaquiline, linezolid and pretomanid, and impact on treatment outcomes in four tuberculosis clinical trials containing pretomanid. PLOS Glob Public Health. 2023;3:e0002283. https://doi.org/10.1371/journal.pgph.0002283.
40. Villellas C, et al. Unexpected high prevalence of resistance-associated *Rv0678* variants in MDR-TB patients without documented prior use of clofazimine or bedaquiline. J Antimicrob Chemother. 2017;72:684–90. https://doi.org/10.1093/jac/dkw502.
41. Merker M, et al. Phylogenetically informative mutations in genes implicated in antibiotic resistance in *Mycobacterium tuberculosis* complex. Genome Med. 2020;12:27. https://doi.org/10.1186/s13073-020-00726-5.
42. Coll F, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat Commun. 2014;5:4812. https://doi.org/10.1038/ncomms5812.
43. Sobkowiak B, et al. Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. BMC Genomics. 2018;19:613. https://doi.org/10.1186/s12864-018-4988-z.
44. Brynildsrud OB, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. Sci Adv. 2018;4:eaat5869.
45. Bradley P, den Bakker HC, Rocha EPC, McVean G, Iqbal Z. Ultrafast search of all deposited bacterial and viral genomic data. Nat Biotechnol. 2019;37:152–9. https://doi.org/10.1038/s41587-018-0010-1.
46. Merker M, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. Nat Genet. 2015;47:242–9. https://doi.org/10.1038/ng.3195.
47. Luo T, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. Proc Natl Acad Sci U S A. 2015;112:8136–41. https://doi.org/10.1073/pnas.1424063112.
48. Norheim G, et al. Tuberculosis outbreak in an educational institution in Norway. J Clin Microbiol. 2017;55:1327–33. https://doi.org/10.1128/JCM.01152-16.
49. Kay GL, et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. Nat Commun. 2015;6:6717. https://doi.org/10.1038/ncomms7717.
50. Nimmo C, 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:e165–74. https://doi.org/10.1016/S2666-5247(20)30031-8.
51. Nimmo C, et al. Dynamics of within-host *Mycobacterium tuberculosis* diversity and heteroresistance during treatment. EBioMedicine. 2020;55:102747. https://doi.org/10.1016/j.ebiom.2020.102747.
52. Nimmo C, et al. Whole genome sequencing *Mycobacterium tuberculosis* directly from sputum identifies more genetic diversity than sequencing from culture. BMC Genomics. 2019;20:389. https://doi.org/10.1186/s12864-019-5782-2.
53. Dheda K, et al. Outcomes, infectiousness, and transmission dynamics of patients with extensively drug-resistant tuberculosis and home-discharged patients with programmatically incurable tuberculosis: a prospective cohort study. Lancet Respir Med. 2017;5:269–81. https://doi.org/10.1016/S2213-2600(16)30433-7.
54. Streicher EM, et al. Molecular epidemiological interpretation of the epidemic of extensively drug-resistant Tuberculosis in South Africa. J Clin Microbiol. 2015;53:3650–3. https://doi.org/10.1128/JCM.01414-15.
55. Guerra-Assuncao JA. et al. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. Elife 2015;4. https://doi.org/10.7554/eLife.05166
56. Grandjean L, et al. Transmission of multidrug-resistant and drug-susceptible Tuberculosis within households: a prospective cohort study. PLoS Med. 2015;12:e1001843. https://doi.org/10.1371/journal.pmed.1001843. discussion e1001843.
57. Grandjean L, et al. Convergent evolution and topologically disruptive polymorphisms among multidrug-resistant tuberculosis in Peru. PLoS ONE. 2017;12:e0189838. https://doi.org/10.1371/journal.pone.0189838.

58. Ismail N, Omar SV, Ismail NA, Peters RPH. Collated data of mutation frequencies and associated genetic variants of bedaquiline, clofazimine and linezolid resistance in *Mycobacterium tuberculosis*. Data Brief. 2018;20:1975–83. https://doi.org/10.1016/j.dib.2018.09.057.
59. Ghajavand H. et al. High prevalence of Bedaquiline resistance in treatment-naive tuberculosis patients and Verapamil effectiveness. Antimicrob Agents Chemother 2019;63. https://doi.org/10.1128/AAC.02530-18
60. Fowler PW. 2017 pygsi v.1.0.0: a Python class to interrogate BIGISI. 2018. https://doi.org/10.5281/zenodo.1407085.
61. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013. arXiv:1303.3997
62. Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;11(10):11–111033. https://doi.org/10.1002/0471250953.bi1110s43.
63. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32:292–4. https://doi.org/10.1093/bioinformatics/btv566.
64. Hariguchi N. et al. OPC-167832, a novel carbostyril derivative with potent antituberculosis activity as a DprE1 Inhibitor. Antimicrob Agents Chemother 2020;64. https://doi.org/10.1128/AAC.02020-19
65. Phelan JE, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med. 2019;11:41. https://doi.org/10.1186/s13073-019-0650-x.
66. Coll F, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 2015;7:51. https://doi.org/10.1186/s13073-015-0164-0.
67. The Cryptic Consortium. A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics. PLoS Biol. 2022;20:e3001721. https://doi.org/10.1371/journal.pbio.3001721.
68. Page AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom. 2016;2:e000056. https://doi.org/10.1099/mgen.0.000056.
69. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35:4453–5. https://doi.org/10.1093/bioinformatics/btz305.
70. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018;46:e134. https://doi.org/10.1093/nar/gky783.
71. Menardo F, et al. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. BMC Bioinformatics. 2018;19:164. https://doi.org/10.1186/s12859-018-2164-8.
72. Bouckaert R, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650.
73. Bouckaert RR, Drummond AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol Biol. 2017;17:42. https://doi.org/10.1186/s12862-017-0890-6.
74. Baele G, et al. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol. 2012;29:2157–67. https://doi.org/10.1093/molbev/mss084.
75. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8. https://doi.org/10.1093/bioinformatics/bty633.
76. Yu GC, Smith DK, Zhu HC, Guan Y, Lam TTY. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210x.12628.
77. O'Neill MB, et al. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Mol Ecol. 2019;28:3241–56. https://doi.org/10.1111/mec.15120.
78. Rutaihwa LK, et al. Multiple Introductions of *Mycobacterium tuberculosis* lineage 2-Beijing into Africa over centuries. Front Ecol Evol. 2019;7:ARTN 112. https://doi.org/10.3389/fevo.2019.00112.
79. Bradley P, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and *Mycobacterium tuberculosis*. Nat Commun. 2015;6:10063. https://doi.org/10.1038/ncomms10063.
80. Menardo F, Duchene S, Brites D, Gagneux S. The molecular clock of *Mycobacterium tuberculosis*. PLoS Pathog. 2019;15:e1008067. https://doi.org/10.1371/journal.ppat.1008067.
81. Ismail N, Peters RPH, Ismail NA, Omar SV. Clofazimine exposure *in vitro* selects efflux pump mutants and Bedaquiline resistance. Antimicrob Agents Chemother 2019;63. https://doi.org/10.1128/AAC.02141-18
82. Andres S, et al. Bedaquiline-resistant tuberculosis: dark clouds on the horizon. Am J Respir Crit Care Med. 2020;201:1564–8. https://doi.org/10.1164/rccm.201909-1819LE.
83. 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
84. Rancoita PMV. 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
85. Beckert P, 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.
86. Loiseau C, et al. An African origin for *Mycobacterium bovis*. Evol Med Public Health. 2020;2020:49–59. https://doi.org/10.1093/emph/eoaa005.
87. Bateson A, et al. Ancient and recent differences in the intrinsic susceptibility of *Mycobacterium tuberculosis* complex to pretomanid. J Antimicrob Chemother. 2022;77:1685–93. https://doi.org/10.1093/jac/dkac070.
88. D'Costa VM, et al. Antibiotic resistance is ancient. Nature. 2011;477:457–61. https://doi.org/10.1038/nature10388.
89. Rifat D, et al. Mutations in fbiD (*Rv2983*) as a novel determinant of resistance to pretomanid and delamanid in *Mycobacterium tuberculosis*. Antimicrob Agents Ch. 2021;65:ARTN e01948–20. https://doi.org/10.1128/AAC.01948-20.
90. Koser CU, Maurer FP. Minimum inhibitory concentrations and sequencing data have to be analysed in more detail to set provisional epidemiological cut-off values for *Mycobacterium tuberculosis* complex. Eur Respir J 2023;61. https://doi.org/10.1183/13993003.02397-2022
91. Kahlmeter G, Turnidge J. The determination of epidemiological cut-off values requires a systematic and joint approach based on quality controlled, non-truncated minimum inhibitory concentration series. Eur Respir J 2023;61. https://doi.org/10.1183/13993003.02259-2022
92. World Health Organization. Optimized broth microdilution plate methodology for drug susceptibility testing of *Mycobacterium tuberculosis* complex. 2022. https://iris.who.int/handle/10665/353066.
93. Liu Y, et al. Reduced susceptibility of *Mycobacterium tuberculosis* to Bedaquiline during antituberculosis treatment and its correlation with clinical outcomes in China. Clin Infect Dis. 2021;73:e3391–7. https://doi.org/10.1093/cid/ciaa1002.
94. Pym AS, et al. Bedaquiline in the treatment of multidrug- and extensively drug-resistant tuberculosis. Eur Respir J. 2016;47:564–74. https://doi.org/10.1183/13993003.00724-2015.
95. Grandjean L, et al. The association between *Mycobacterium Tuberculosis* genotype and drug resistance in Peru. PLoS ONE. 2015;10: e0126271. https://doi.org/10.1371/journal.pone.0126271.

Publisher's Note