Epidemics 36 (2021) 100471

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Quantifying transmission fitness costs of multi-drug resistant tuberculosis

Jūlija Pečerska [a,i,*], Denise Kühnert [b], Conor J. Meehan [c], Mireia Coscollá [d], Bouke C. de Jong [e], Sebastien Gagneux [f,g], Tanja Stadler [h,i,*]

[a] School of Life Sciences and Facility Management, ZHAW, Wädenswil, Switzerland
[b] Transmission, Infection, Diversification & Evolution Group, Max Planck Institute for the Science of Human History, Jena, Germany
[c] School of Chemistry and Bioscience, University of Bradford, Bradford, UK
[d] Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, València, Spain
[e] Unit of Mycobacteriology, Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[f] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland
[g] University of Basel, Basel, Switzerland
[h] Department of Biosystems Science and Engineering, ETHZ, Basel, Switzerland
[i] Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

ARTICLE INFO

Keywords:
Antibiotic resistance
Multi-type birth–death model
Phylodynamics
Whole genome *M. tuberculosis*

ABSTRACT

As multi-drug resistant tuberculosis (MDR-TB) continues to spread, investigating the transmission potential of different drug-resistant strains becomes an ever more pressing topic in public health. While phylogenetic and transmission tree inferences provide valuable insight into possible transmission chains, phylodynamic inference combines evolutionary and epidemiological analyses to estimate the parameters of the underlying epidemiological processes, allowing us to describe the overall dynamics of disease spread in the population. In this study, we introduce an approach to *Mycobacterium tuberculosis* (*M. tuberculosis*) phylodynamic analysis employing an existing computationally efficient model to quantify the transmission fitness costs of drug resistance with respect to drug-sensitive strains. To determine the accuracy and precision of our approach, we first perform a simulation study, mimicking the simultaneous spread of drug-sensitive and drug-resistant tuberculosis (TB) strains. We analyse the simulated transmission trees using the phylodynamic multi-type birth–death model (MTBD, (Kühnert et al., 2016)) within the BEAST2 framework and show that this model can estimate the parameters of the epidemic well, despite the simplifying assumptions that MTBD makes compared to the complex TB transmission dynamics used for simulation. We then apply the MTBD model to an *M. tuberculosis* lineage 4 dataset that primarily consists of MDR sequences. Some of the MDR strains additionally exhibit resistance to pyrazinamide — an important first-line anti-tuberculosis drug. Our results support the previously proposed hypothesis that pyrazinamide resistance confers a transmission fitness cost to the bacterium, which we quantify for the given dataset. Importantly, our sensitivity analyses show that the estimates are robust to different prior distributions on the resistance acquisition rate, but are affected by the size of the dataset – i.e. we estimate a higher fitness cost when using fewer sequences for analysis. Overall, we propose that MTBD can be used to quantify the transmission fitness cost for a wide range of pathogens where the strains can be appropriately divided into two or more categories with distinct properties.

1. Introduction

Tuberculosis (TB) continues to be a major problem for global public health. A connected and pressing issue is the continued detection of drug-resistant TB, and especially of multidrug-resistant (MDR-TB) strains, which resist treatment by at least two main first-line drugs, rifampicin and isoniazid. Rifampicin-resistant and MDR-TB made up as much as half a million of the 10.6 million new tuberculosis cases worldwide in 2016 (WHO, 2017). In the same year, an estimated 19% of previously treated TB cases were rifampicin- or multidrug-resistant. While MDR-TB is treatable and curable by second-line drugs, there are only a few second-line treatment options, all of which require regimens that last from 9 months up to 2 years and are expensive and toxic (WHO, 2016). While these treatments are normally successful in curing MDR-TB patients, WHO reports that in 2017 an average of 6.2% of MDR-TB cases resisted treatment by the most effective second-line anti-TB drugs, representing the so-called extensively drug-resistant

* Corresponding authors.
E-mail addresses: julija.pecerska@zhaw.ch (J. Pečerska), tanja.stadler@bsse.ethz.ch (T. Stadler).

https://doi.org/10.1016/j.epidem.2021.100471
Received 1 February 2019; Received in revised form 14 January 2020; Accepted 17 May 2021
Available online 21 May 2021

(XDR-TB) cases. In at least 127 countries worldwide, one or more cases of XDR-TB had been reported by the end of 2017 (WHO, 2018). The continuous detection of such strains in transmission clusters and the lack of new anti-TB drugs highlights the need for preventing further transmission of drug-resistant strains (Kendall et al., 2015).

Any treatment selects for drug-resistant strains and any drug resistance is a burden for the individual patient. A resistant strain with high transmission potential may cause a resistant epidemic and thus poses a serious risk for the general population. Thus, public health measures should aim at preventing the emergence of resistant strains with a high transmission potential. In this study we aim to quantify the transmission fitness cost of drug-resistant strains using *M. tuberculosis* genomic data.

New sequencing technologies allow us to obtain large numbers of *M. tuberculosis* genome sequences (Meehan et al., 2018; Sengstake et al., 2017; Casali et al., 2014). Extensive genetic sequencing allows us to detect some types of drug resistance earlier compared to phenotypic methods (Yakrus et al., 2014; Miotto et al., 2014; Pankhurst et al., 2016; Colijn and Cohen, 2016), and to detect drug resistance in cases when standardized tests are not available (Horne et al., 2012). New technologies allow real-time whole genome sequencing (WGS) of ongoing epidemics (Walker et al., 2018).

Phylogenetic and transmission analyses of WGS data attempt to reconstruct transmission between infected individuals. Tools for phylogenetic and transmission tree reconstruction from TB WGS data are increasingly becoming available, e.g. Didelot et al. (2014, 2017) and Klinkenberg et al. (2017).

Phylodynamic analysis, an approach introduced more than a decade ago by Grenfell et al. (2004), aims at unifying the inference of epidemiological and evolutionary dynamics of pathogens. This approach aims to estimate the parameters of the tree-generating process, e.g. transmission and cure rates in the case of an epidemiological model, jointly with the evolutionary relationships between sampled sequences. There are still a number of challenges to be tackled in phylodynamics (Frost et al., 2015), particularly, since the first generation of phylodynamic tools have been used and validated exclusively on viral sequences. Now that whole genome sequences are readily available for other types of pathogens, rigorous testing needs to be performed to further validate their use (Biek et al., 2015).

Few phylodynamic methodological approaches have been developed specifically for analysing *M. tuberculosis* datasets. One example is work by Merker et al. (2018), where population size estimates were used to approximate the fitness of strains with compensatory mutations. However, to our knowledge, no previous study has directly estimated epidemiological dynamics such as the relative transmission fitness of drug resistant strains for TB.

In this study, we first investigate the appropriateness of a phylodynamic tool developed for viral pathogens to study the epidemiological dynamics of TB. We first simulate epidemics under an epidemiological model specific for TB, including latent (exposed) periods and treatment periods. We then apply this phylodynamic tool to the simulated data and evaluate the fitness costs for drug-resistant TB strains compared to the drug-sensitive TB strains. The tool is called the multi-type birth–death (MTBD) model (Kühnert et al., 2016), which works within the BEAST2 software framework (Bouckaert et al., 2014). Inference under MTBD is based on a multi-type birth–death-sampling process. It assumes that different strain types (such as drug sensitive and drug resistant) circulate simultaneously within an epidemic, and allows estimation of type-dependent transmission rates based on sequencing data. To ensure computational feasibility, our configuration of MTBD effectively ignores the complex TB dynamics such as latency and treatment.

Throughout, we estimate the relative transmission fitness r_λ of drug-resistant strains as the ratio of the drug-resistant strain transmission rate (λ_R) to the drug-sensitive strain transmission rate (λ_S), i.e. $r_\lambda = \lambda_R/\lambda_S$. This definition of relative transmission fitness quantifies the average decrease or increase in the number of new cases per unit of time caused by a patient infected with a resistant strain compared to a patient with a drug-sensitive strain. For a given relative fitness r_λ, the transmission fitness cost is $(1 - r_\lambda) \times 100\%$. This configuration of MTBD has been applied to estimate the relative transmission fitness of drug-resistant mutations for the human immunodeficiency virus (HIV) (Kühnert et al., 2018). However, we present here the first study in which it is applied to a bacterial – and hence much more slowly evolving – pathogen.

Our simulation study shows that MTBD parameter estimates are highly robust when estimating the relative transmission fitness r_λ of drug resistant strains, despite long periods of treatment and latency used in the simulation scenarios.

To illustrate the utility of MTBD for TB epidemiological analysis, we apply it to an *M. tuberculosis* dataset sampled over the course of five years in Kinshasa, the capital of the Democratic Republic of the Congo. The dataset contains sequences from re-treatment cases of TB which mainly exhibit MDR phenotypes. Many of these MDR sequences also carry substitutions that indicate pyrazinamide resistance (Meehan et al., 2018). We set out to test the previously posed hypothesis suggesting that pyrazinamide resistance reduces the ability of a strain to be transmitted from host to host (den Hertog et al., 2015). In the terminology used here this translates to the relative transmission fitness r_λ of drug-resistant strains being below one. To test this hypothesis, we quantify the relative transmission fitness of additional pyrazinamide resistance when compared to pyrazinamide-sensitive MDR strains.

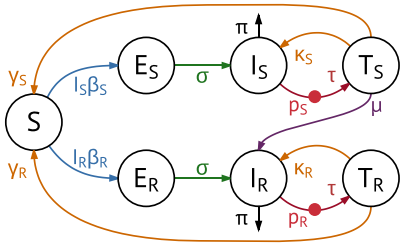
2. Materials and methods

2.1. Simulation study

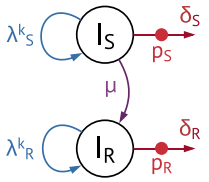
2.1.1. Simulating epidemics

In order to simulate realistic epidemics we designed an epidemiological model that accounts for the most important aspects of TB dynamics. This model builds upon previous models described in the literature (e.g. Gomes et al., 2007; Cohen et al., 2009; Pinho et al., 2015; Dowdy et al., 2013), and was adjusted to more closely represent the spread of drug-sensitive and drug-resistant *M. tuberculosis* strains within the scope of a single epidemic. In our modelling framework, strain spread is described by the Susceptible–Exposed–Infectious–Treated ($SEIT_2$) model, shown in Fig. 1(a). The model is tailored towards *M. tuberculosis* transmission as follows. Upon exposure to the bacterium only 10% of the susceptible population (S compartment) proceed to infection; others will develop neither disease nor infectiousness and can therefore be ignored in the model (Vynnycky and Fine, 1997). As the sampling for the available dataset has been done within the span of five years, we restrict our simulations to modelling short-term epidemics. Previous infections have no clear effect on immunity to consecutive disease (Verver et al., 2005; Chiang and Riley, 2005; Yew and Leung, 2005), and existing methods of vaccination (e.g. the BCG vaccine) seem to have a negligible effect on infectious disease dynamics in adults in endemic settings (Gomes et al., 2004). Therefore, recovered individuals return to being susceptible after successful treatment.

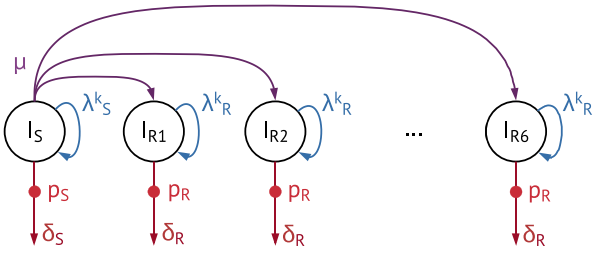
Each simulated epidemic starts with a single patient infected with a drug-sensitive strain (compartment I_S in Fig. 1(a)) and N-1 susceptible individuals (compartment S in Fig. 1(a)), where N is the total population size. A patient enters the latent (exposed) phase (E_S compartment) upon infection with a drug-sensitive strain, and moves (with rate σ) to the active phase of TB infection (I_S compartment). Patients in the active phase can transmit to susceptible individuals from the S compartment with rate β_S, die with rate π, and start treatment (moving to the T_S compartment) with rate τ. As a patient starts treatment, they will be sampled and the *M. tuberculosis* genome sequenced with probability p_S. We assume that successful treatment always leads to recovery (i.e. the individual moves to the S compartment with rate γ_S), whereas dropped treatment or otherwise unsuccessful treatment leads to disease



(a) $SEIT_2$ model with drug-sensitive and drug-resistant strains, subscripts S and R respectively. Rates are marked as follows ($x \in \{S, R\}$): β_x - infection rate, σ - disease progression rate, π - death rate, τ - treatment rate, κ_x - relapse rate, γ_x - recovery rate, μ - resistance evolution rate.



(b) $MTBD_2$ model with drug-sensitive and drug-resistant strains, subscripts S and R respectively. Rates are marked as follows ($x \in \{S, R\}$): λ^k_x - time-dependent birth rate per interval k, δ_x - recovery rate, μ - resistance evolution rate.



(c) The $MTBD_7$ model setup for Kinshasa analysis of clusters with at most 6 different types of pyrazinamide resistance substitutions per cluster. Each of the I_{Rn} compartments, where n is the compartment number, represents a distinctive resistance mutation within a cluster. Rates are marked as follows ($x \in \{S, R\}$): λ^k_x - time-dependent birth rate per interval k, δ_x - recovery rate, μ - resistance evolution rate. Resistance cannot be lost, only acquired, and all resistant strains have the same fitness relative transmission fitness.

Fig. 1. Different models used for simulation and analysis. In all figures, the compartments are marked as follows: Susceptible — S, Exposed — E_x, Infected — I_x, under Treatment — T_x, where x is either drug-sensitive or drug-resistant, S or R respectively. Sampling probability is marked by p_x, where x is S or R.

relapse (i.e. the individual moves to the I_S compartment again, with rate κ_S). Furthermore, we assume that there are no co-infections. As all diagnosed cases in the study area are currently getting treatment, there is little possibility for self-cure. Hence, infected individuals never recover without treatment in our model.

To account for the drug-resistant strains (compartments E_R, I_R, T_R in Fig. 1(a)), we add a rate μ with which individuals in the treated compartment T_S may develop drug resistance and thus move into a resistant infectious compartment I_R. We assume that drug resistance is never lost, i.e. a resistant individual cannot move back to the sensitive class. As it is also possible for drug resistance to be transmitted, we allow new infections which follow the same dynamics as for drug-sensitive strains. Individuals enter the latent phase (E_R compartment) upon infection at a rate β_R and progress to active disease (I_R compartment) at a rate σ. Again, infectious individuals can be treated (e.g. using second-line treatment in the case of MDR-TB) and thus move to compartment T_R, where they can recover or relapse. A patient entering the T_R class will be sampled and the *M. tuberculosis* genome sequenced with probability p_R. We call this model an $SEIT_2$ model, and in general an $SEIT_m$ model with $m-1$ resistant classes.

The simulations were performed using the Bayesian inference framework BEAST2 (Bouckaert et al., 2014). The BEAST2 package MASTER (Moments and Stochastic Trees from Event Reactions) (Vaughan and Drummond, 2013) was used to simulate stochastic realizations of epidemic histories. The simulation model was specified as chemical master equations (CMEs) describing the transitions between different $SEIT_m$ model states happening with predefined rates. The stochastic simulator produced a random outcome of the epidemic in the form of a transmission tree. The simulations with initial population sizes N = 200,000 or N = 1000 stopped when 300 or 150 cases were sampled, respectively, which is close to the sample number available in the empirical dataset from Kinshasa. The large population size results in exponentially growing epidemics. The smaller population size of N = 1000 resulted in epidemics where the infected population size saturated (see supplementary Figure 9).

To match real life sampling and to ensure identifiability, we only keep trees with a minimum sample of 30 patients with a drug-resistant infection and 30 patients with a drug-sensitive infection, and restart the simulation otherwise. We also restart the simulation in cases when the epidemic died out before reaching the desired number of sampled cases. While this constraint is biasing the tree sample for surviving epidemics, we want to mimic real world data, and we would not be able to obtain accurate estimates for empirical datasets that have fewer sequences.

To mimic real life situations we used a number of different configurations for our model parameters β_x, σ, τ, γ_x, κ_x, p_x, π, and μ, where $x \in \{S, R\}$. We specify the basic reproductive number $R_{0,S}$ of the sensitive strain type, defined as the expected number of secondary infections caused by a single drug-sensitive infected individual at the start of the simulation prior to changing resistance status. Thus, in case a drug-sensitive individual evolves drug resistance, only the secondary infections caused prior to drug resistance contribute to the $R_{0,S}$. Analogously, the respective basic reproductive number of resistant strains is $R_{0,R}$.

In the simulation setup, we specify $R_{0,x}$ instead of β_x, $x \in \{S, R\}$. This re-parametrization of β_x given the parameters $R_{0,x}$, σ, τ, γ_x, κ_x, p_x, π, μ, $x \in \{S, R\}$, is described in the Supplement. We use combinations of values for $R_{0,S}$ and $R_{0,R}$ for which the drug-resistant strain causes either the same or a slightly lower number of secondary infections than the drug-sensitive strain. We also included a case in which the basic reproductive number of the resistant strain $R_{0,R}$ is below the epidemic threshold 1. The $R_{0,x}$ parameter combinations for simulation were as follows: $(R_{0,S}, R_{0,R}) = (1.3, 1.1)$, (1.2, 1.1), (1.1, 1.1), and (1.2, 0.9). Note that all R_0 values are around 1, since TB is endemic in many countries (Stadler, 2011; Ma et al., 2018).

The time spent in the exposed and infectious compartments together (prior to first treatment or death) is fixed to one time unit, which we set to one year in our simulations. The proportion of time spent in the exposed and infectious compartments, respectively, is varied. The effect of a higher proportion of time spent in the exposed compartment on the tree shape is that bifurcations (new infections) will start to

appear on branches later after the start of the branch, since individuals reach the infectious state later. average time spent in the exposed compartment in the simulations ranges from $t_E = 0, 0.2, 0.4, 0.5, 0.6, 0.8$ to 0.9 year, which is specified by the rate $\sigma = [\infty, 5, 2.5, 2, 5/3, 5/4, 10/9]$ year^{-1}. Hence, the average times spent in the infectious compartment are $t_I = 1 - t_E = [1, 0.8, 0.6, 0.5, 0.4, 0.2, 0.1]$ year, specified by $\tau + \pi = [1, 5/4, 5/3, 2, 2.5, 5, 10]$ year^{-1}. Thus, $1/\sigma + 1/(\tau+\pi) = 1$, which keeps the total time of infection before first treatment or death at one year. This way we allow the time a person spends while exposed and not yet infectious to be up to nine times longer than the infectious time. Individuals infected with either sensitive or resistant strains are removed from the infectious pool due to fatal outcomes with probability 0.1 and proceed to treatment with probability 0.9, (i.e. $\tau = 0.9 \times 1/t_I$). The following rates are used for the recovery rate as consequence of treatment: $\gamma_S = 1.0$ year^{-1}, $\gamma_R = 0.5$ year^{-1}, which takes into account the fact that resistant strains need longer treatment given currently recommended treatment regimens. Mean rates of relapse are set to $\kappa_S = 0.1$ year^{-1}, $\kappa_R = 0.075$ year^{-1}, corresponding to a lower chance of relapse for the resistant strains with appropriate treatment. Drug resistance mutations were acquired at a rate of $\mu = 0.04$ year^{-1}. We assume that *M. tuberculosis* drug resistance reversal does not occur (Andersson and Hughes, 2010; Casali et al., 2012; Allen et al., 2017). Unfortunately, relapse and drug resistance acquisition rates have not yet been quantified conclusively, so these rates were set in an ad hoc way. We also simulated trees without exposure or possible relapse by setting $\sigma = \infty$ and $\gamma_x = \infty$ (essentially setting the time spent in the E_x and T_x compartments to 0, $x \in \{S, R\}$), referred to as the SIS_2 model.

Due to drug resistant strains being of greater clinical interest, a higher proportion of them is sampled compared to sensitive strains. Hence, the sampling probabilities were set to $p_S = 0.1$ and $p_R = 0.3$ in the simulations.

To account for the stochasticity of the epidemiological process for each of the different parameter configurations we ran a hundred separate instances.

2.1.2. Analysis of the simulated epidemics

Employing a full $SEIT_m$ model for phylodynamic inference would be very demanding computationally (see supplementary section MTBD vs. $SEIT_m$). Instead, we analyse the simulated trees using the BEAST2 package MTDB (Kühnert et al., 2016). We configure it to fit a simpler epidemiological model (Fig. 1(b)), and estimate the posterior distribution of the model parameters given each tree simulated using $SEIT_2$. MTBD allows us to estimate a time-dependent transmission rate λ_x^k, which accounts for possible changes in transmission rates due to e.g. susceptible depletion or a newly introduced effective vaccination strategy. Transmission rates are modelled as piecewise constant rates, i.e. the transmission rate is constant within a user-specified time interval k (λ_x^k), after which it may change to another constant rate (λ_x^{k+1}), and so on. The number of time intervals is user-specified. The model estimates a removal (become uninfectious) rate δ. We use a MTBD setup in which all infected individuals are infectious, i.e. the latent and treatment phases are ignored. We further assume that δ is constant through time as the treatment strategies stay the same during the time spanned by our phylogenetic tree. Sensitive and resistant strain have the same δ as any difference in time until treatment is likely negligible. Similarly, MTBD estimates the sampling proportions p_S and p_R and resistance acquisition rate μ, which are directly comparable between MTBD and $SEIT_2$. We fix the sampling probability in the simulation analysis to the true values (see supplementary section Parameter definitions). We use the MTBD setup which disallows so-called sampled ancestors.

In the classic MTBD setup, one would estimate a separate λ_S^x per time interval k and per strain $x \in \{S, R\}$. However we argue that the relative transmission fitness λ_R^k/λ_S^k in real epidemics is independent of the speed of spread (e.g. due to the varying number of susceptible individuals) at a particular point in time, thus the disadvantage a strain has after developing drug resistance stays constant. This means that we assume $r_\lambda = \lambda_R^k/\lambda_S^k$ is constant for all time intervals k. We implemented this assumption in MTBD by estimating r_λ, δ_S, δ_R, μ and $R_{e,S} = \lambda_S^k/(\delta_S+\mu)$. The latter is called the effective reproductive number which we estimate for the sensitive strain. The effective reproductive number $R_{e,S}$ of the sensitive (resp. $R_{e,R}$ of the resistant) strain type is defined as the expected number of secondary infections caused by a single drug-sensitive (resp. drug-resistant) infected individual at time t, before leaving the class x, $x \in \{S, R\}$. Thus, analogous to the basic reproductive number, in case a drug-sensitive individual evolves drug resistance, we only count the secondary infections caused prior to evolving the drug resistance. Thus, $R_{e,S}^k = \lambda_S^k/(\delta_S+\mu)$ and $R_{e,R}^k = \lambda_R^k/\delta_R$ for each time interval k.

We evaluate the performance of MTBD for analysing $SEIT_2$ model simulations by comparing the estimated relative transmission rate r_λ to the true ratio β_R/β_S in $SEIT_2$. The prior distributions used for $R_{e,S}$, r_λ, δ_x, p, and μ are provided in Table 1. We perform all analyses assuming both (i) a constant transmission rate and (ii) a piecewise constant transmission rate over three time intervals. For (ii) we split the time covered by the simulated tree such that each interval has the same number of branching events approximately. For all analyses of the simulation runs the Markov chain Monte Carlo (MCMC) reached an ESS for all parameters of at least 200. For some of the simulation configurations with $t_E = 0.9$ and $t_I = 0.1$ the simulations failed to run due to the epidemics dying out almost instantly, thus they were excluded from the results.

2.2. The Kinshasa dataset

The Kinshasa *M. tuberculosis* dataset consists of 324 sequences sampled from re-treatment patients, most of which were identified as Lineage 4 (309). The sequences were sampled over the course of 5 years and the sampling calendar dates were recorded. The sequence alignment is 6567 nucleotides long, not including any known resistance mutations.

Of the 309 sequences, 170 were clustered MDR-TB strains. The Lineage 4 sequences are not all part of one single transmission cluster, but form a number of smaller clusters identified previously (Meehan et al., 2018). These transmission clusters are based on a 12 single-nucleotide polymorphism (SNP) cut-off. This cut-off excluded the mutations known to cause resistance to reduce false clustering due to similar drug resistance profile (any mutations defined by Feuerriegel et al., 2015). For the purpose of this paper we also removed non-MDR sequences, i.e. sequences that lack one or both isoniazid or rifampicin resistance, and sequences which have different MDR profiles within a cluster.

102 of the clustered strains additionally exhibited pyrazinamide resistance (Meehan et al., 2018). Pyrazinamide was used in Kinshasa as an anti-tuberculosis drug in the form of a fixed combination tablet. Pyrazinamide is an antimycobacterial pro-drug that is activated by the enzyme pyrazinamidase, which is encoded by the non-essential *pncA* gene in *M. tuberculosis* (Yadon et al., 2017). As pyrazinamide action depends on the activity of the enzyme encoded by *pncA* (Njire et al., 2016), multiple different single point mutations in *pncA* may cause resistance to pyrazinamide by disrupting the enzyme. Little convergent evolution has been detected on that gene (Miotto et al., 2014), and it appears that resistance reversal mutations are extremely unlikely (Andersson and Hughes, 2010). The dataset contains sequences that have 59 different *pncA* gene mutations and we assume that the relative transmission fitness for each of the different mutations is the same.

The resulting dataset consisted of 33 transmission clusters, sized 2 to 30 strains. Each cluster contained at most 6 different *pncA* substitutions, meaning that in each cluster we have up to 6 resistant compartments. In the resulting $MTBD_7$ model, each resistant compartment informs the same transmission rate. Drug resistance substitutions are not used to infer phylogenies, but only to categorize strains into different resistance compartments. The division of strains into different compartments does

Table 1
Prior distributions for the MTBD parameters.

Analysis	$R_{e,S}$	r_λ	δ_x	μ	p_x
Simulations	Lognormal(0, 1.25)	Lognormal(0, 0.5)	Lognormal(0, 0.5)	Exp(1.0)	Fixed
Kinshasa	Lognormal(0, 1.25)	Lognormal(0, 0.5)	Lognormal(0, 0.5)	Exp(1.0) Exp(0.2) Exp(50)	Beta(23, 977)

not enforce their clustering on the tree. The same *pncA* substitution can evolve multiple times along the tree if the tree structure estimated from the genetic sequences favours *de novo* resistance rather than transmitted resistance (high μ, low $R_{e,R}$). We however disallow resistance reversal.

The model setup used for analysis is shown in Fig. 1(c). We performed MTBD analyses assuming a constant $R_{e,S}$ over the whole time period (we report in the Results section that, based on our simulation study, r_λ is estimated reliably without assuming time-variation for $R_{e,S}$). We set the priors for the MTBD parameters as specified in Table 1 and the prior for the substitution rate was set to a Log-normal distribution with the mean of $1.5e^{-7}$ and standard deviation of 1.0 (Meehan et al., 2018). While the distribution is Log-normal, the mean is specified not in log, but in real space, such that it translates directly to estimated substitution rates. We set the sampling proportion to be equal for both strain types as all strains were sampled regardless of pyrazinamide resistance status. In particular, we set a narrow prior on the sampling proportion, centring the mean sampling proportion at 2.3%, as estimated by Meehan et al. (2018). The sampling proportion is a lot lower than in the simulations, however the difference in sampling is accounted for in the prior and should not bias the estimates of other parameters, as MTBD accounts for the sampling proportions in the likelihood (see Kühnert et al., 2016). Then, we estimate phylogenetic trees for each of the 33 clusters; the MTBD and evolutionary parameters are shared across all clusters.

The empirical analysis was done based on several small clusters, while the simulation study was done on one large cluster (and revealed reliable results for that scenario, see the Results section). To validate our empirical analysis approach involving several clusters, we have performed analyses only using one or a few of the Kinshasa clusters. Starting with the largest cluster in the Kinshasa data, we analysed increasing numbers of clusters, sequentially including the smaller cluster sizes. We similarly removed the largest clusters, thereby reducing the size of the dataset.

The relative fitness r_λ and the de-novo resistance evolution rate μ should be inversely correlated, as new infectious and *de novo* resistance acquisitions are the only routes to a drug-resistant infection. We investigated the robustness of the pyrazinamide resistance acquisition rate μ and r_λ estimates by changing the prior on the mutation rate from Exp(1.0), translating to a mean rate of 1, to Exp(0.2), translating to a mean rate of 5. Such priors set the mean time until resistance acquisition to 1 and 0.2 years respectively. We additionally ran the full cluster analysis under a very restrictive prior on μ (Exp(50), translating to a very low mean rate of 0.02) to see whether this will greatly influence the results. In all of the analyses the MCMC reached an ESS for all parameters of at least 175.[1]

3. Results

3.1. Simulation study

We simulate *M. tuberculosis* phylogenetic trees under an $SEIT_2$ model (shown in Fig. 1(a)), performing 100 simulations for each chosen parameter combination. We then estimate the epidemiological parameters based on the simulated phylogenetic trees using the MTBD package (model shown in Fig. 1(b)) within BEAST v2.0, resulting in a sample of the posterior distribution for each model parameter. We summarize each posterior distribution by computing the median and the 95% highest posterior density (HPD) interval. To evaluate the estimates for all 100 simulated trees together, we report the median of the set of parameter medians.

Here we report the relative transmission fitness r_λ, which is defined as the ratio of transmission rates of the two strains: λ_R/λ_S. If the r_λ is greater than 1, we conclude that the MDR *M. tuberculosis* strain is fitter than the drug-sensitive strain. Fig. 2 and supplementary Figures 1 to 8 show the resulting estimates from all simulations, the simplest ignoring the E and T compartment and the most complex including a long latent period (large t_E) and treatment.

First, we discuss the Figures showing the simulated epidemic on a fairly small population size, N = 1000 (Fig. 2 and supplementary Figures 1 to 4). The transmission rates and consequently the effective reproductive number R_e decrease through time due to a depletion of susceptibles. As explained in the Materials and Methods section, we assume R_e to be either (i) constant or (ii) allow three piecewise constant intervals for R_e in the inference. First, we observe that violating the MTBD assumptions by adding the E and T compartments neither affects the estimate of the relative transmission fitness r_λ nor the estimates of $R_{e,S}$ and $R_{e,R}$. When assuming (i) that R_e is constant through time, the r_λ is estimated very well (in 91%–99% of the simulations for each configuration the true value falls within the estimated HPD). Not only the coverage is high, accuracy is good as well. For example, when the relative fitness is relatively high (e.g. $r_\lambda \approx 0.72$), in at least 72% of the analyses the HPD interval excludes 1 when it should be excluded due to the fitness cost of resistance. As expected, the estimated R_e is lower than the true R_0 when a constant R_e is used as it averages over the whole time period which includes times with low susceptible counts.

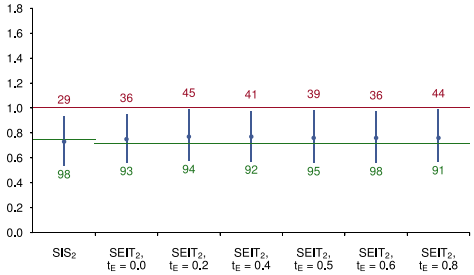
For (ii), the $R_{e,S}$ and $R_{e,R}$ estimates in the first interval of the epidemic correspond to the true R_0 and the second and third interval show a drop in the estimated R_e due to a decreasing susceptible population size. The r_λ is estimated well, as in 86%–98% of simulations for each configuration the true value is within the estimated HPD. Here, in the cases when the relative fitness cost is high (e.g. $r_\lambda \approx 0.72$), in at least 55% of the analyses, the corresponding HPD interval excludes one.

Given a large population size (supplementary Figures 5 to 8), we estimate all parameters reliably, both when using one and three intervals for R_e estimation. Additionally, given more data the estimates of r_λ become more precise, as for $r_\lambda \approx 0.72$ in at least 81% of the analyses the HPD interval excludes 1 in both single and three interval estimates.

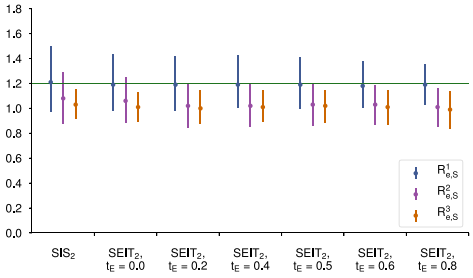
3.2. The Kinshasa dataset

We ran the analyses on the complete dataset using the package MTBD configured as shown in Fig. 1(c) under two different priors on the resistance acquisition rate μ, Exp(1) and Exp(0.2). In particular we assumed a constant $R_{e,S}$ through time as the simulations revealed that this assumption still produces reliable r_λ estimates. The two analyses estimate a median relative transmission fitness of approximately 0.64: $r_\lambda = 0.6417$ (median, 95% HPD: [0.5477, 0.7378]) for Exp(1) and $r_\lambda = 0.6403$ (median, 95% HPD: [0.5454, 0.7394]) for Exp(0.2) (Fig. 3). We translate the estimated relative fitness to a pyrazinamide resistance transmission fitness cost of approximately 36%. To test robustness of the results when the acquisition rate is much slower, we additionally

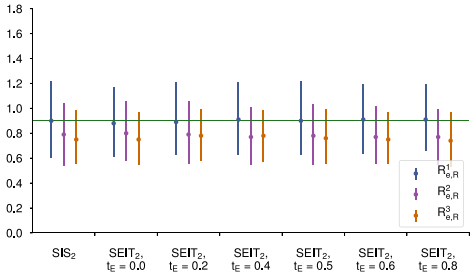
[1] With the exception of two runs on a small number of clusters, where the prior did not mix, however as the runs on the incomplete dataset were done to verify the approach and every other run mixed, we disregard those.



(a) r_λ estimates for three intervals. The green line shows the true r_λ value from the simulations and the number in green indicates the percentage of runs whose 95% HPD includes the truth. The red line indicates the value 1 and the number in red indicates the percentage of runs that produce estimates whose 95% HPD include 1. If included in the HPD, the value 1 means that the method can not distinguish between the two strains with different transmission rates.



(b) $R_{e,S}$ estimates for three intervals. Blue represents the estimate in the time interval starting at the root of the tree, purple the middle time interval, and red the most recent time interval. Note that due to a decreasing number of susceptible individuals, we only expect the blue bars to appropriately estimate the true basic reproductive number shown as a green horizontal line.



(c) $R_{e,R}$ estimates for three intervals. Blue represents the estimate in the time interval starting at the root of the tree, purple the middle time interval, and red the most recent time interval. Note that due to a decreasing number of susceptible individuals, we only expect the blue bars to appropriately estimate the true basic reproductive number shown as a green horizontal line.

Fig. 2. r_λ, $R_{e,S}$ and $R_{e,R}$ estimates plotted in relation to the different simulation models, for $r_\lambda \approx 0.72$, $R_{0,S} = 1.2$ and $R_{0,R} = 0.9$, 1000 individuals in the population, 150 samples and 3 intervals for R_e estimates. Each plot shows the median parameter estimates for 100 simulation runs for each configuration. The points on the vertical lines indicate the median of estimate medians per 100 runs. Similarly, the upper and lower bounds show the median values of the 95% HPD interval limits per 100 runs.

used a prior of Exp(50) on μ. This analysis estimates a slightly higher relative transmission fitness with largely overlapping confidence intervals: $r_\lambda = 0.7194$ (median, 95% HPD: $[0.6261, 0.8063]$). The parameter estimates for μ for the Exp(0.2), Exp(1) and Exp(50) priors overlap by a large margin:
0.0774 (median, 95% HPD: $[0.0498, 0.1116]$), 0.0752 (median, 95% HPD: $[0.0466, 0.1071]$) and 0.0414 (median, 95% HPD: $[0.0277, 0.0572]$) (see Supplementary Figure 10).

In order to test robustness of our results to dividing sequences into clusters, we perform incremental analyses, where we first analyse only the biggest cluster, then the two biggest clusters, etc. The results for the analyses under an Exp(1) prior on μ are shown in supplementary Figure 11a. Second, we perform a decremental cluster analysis, where we start with the full dataset and then remove the largest cluster from the analysed dataset, then the two largest clusters, etc. The corresponding results are shown in supplementary Figure 11b. The incremental cluster analyses show a consistent fitness cost and an effect of the dataset size on the fitness cost estimate. As more information is added, the between-host pyrazinamide resistance transmissibility increases (from $r_\lambda \approx 0.5$ to $r_\lambda \approx 0.64$), and stabilizes after a certain dataset size is reached (see supplementary Figure 11a). Furthermore, the prior on μ does not affect these results (supplementary Figure 12). Similarly, in the decremental cluster analyses, the cost rises as more sequences are removed from the analyses.

We further performed the incremental cluster analyses without sequence data, while still specifying the cluster sizes, sampling times, and drug resistance types, i.e. information on the composition of prevalence data. For this, all sequence data was replaced by a single unknown nucleotide character in the configuration files. The posterior estimates are shown in supplementary Figure 13. Importantly, r_λ for the full dataset is estimated around 0.42. Thus, the same method using prevalence-related data only predicts a relative transmission fitness of 0.42, while adding sequences predicts a relative fitness of around 0.64.

Additionally, we investigated whether the estimated relative transmission fitness in the analyses without genomic data reflects simple data properties such as the proportion of pyrazinamide-resistant strains to all strains, or the diversity of the pyrazinamide-resistant strains. However, we find no clear trend (supplementary Figure 13 vs. supplementary Figure 14). This suggests that the fitness cost is at least in part informed by the sampling times and drug resistance statuses, even when disregarding evolutionary relationships between samples.

4. Discussion

We have shown that the BEAST2 MTBD package can be used to estimate relative transmission fitness for drug-resistant TB strains compared to drug-sensitive strains. Even though our MTBD configuration ignores latency and treatment during infection, it reliably estimates the transmission dynamics for simulations with long latent and treatment phases. This insight is useful beyond the analysis of *M. tuberculosis*. Many infections can be treated but may relapse. Many phylodynamic tools do not directly model exposure, treatment, or relapse for computational reasons. Here, we show that a phylodynamic tool with such a setup can still robustly estimate relevant epidemiological quantities of epidemics with treatment, latency, relapse, and other dynamics. Latency should be accounted for whenever possible, however, our results suggest that we can gain valuable insights on pathogen transmission dynamics even when latency could not be explicitly incorporated into the analysis due to computational or statistical reasons. As in every simulation study, only a range of parameters can be investigated. Here we picked parameter values that are realistic for MDR-TB epidemics,

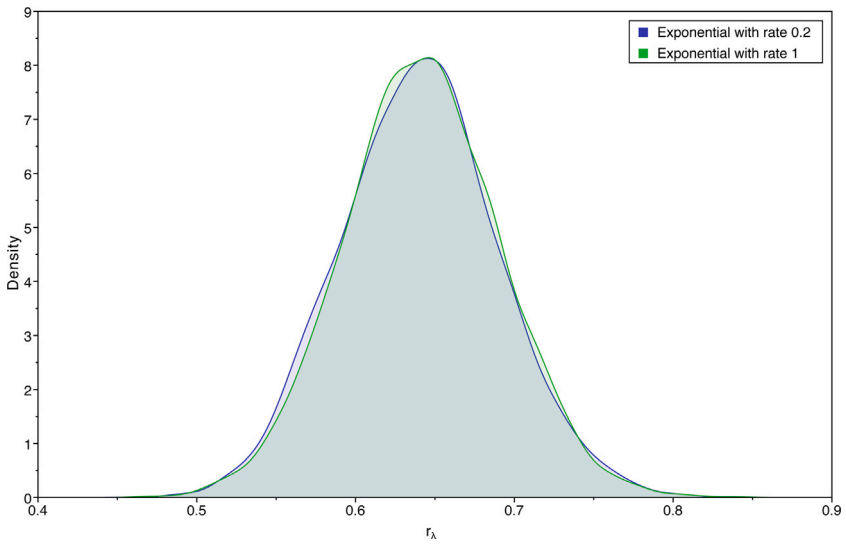


Fig. 3. The posterior probability density of r_λ for the full dataset analysis of the Kinshasa sequences under 2 different priors on the resistance acquisition rate μ.

and researchers interested in different pathogens are highly encouraged to use our BEAST2 configuration files (in supplement) with modified parameter settings to explore the appropriateness of MTBD for their study system.

Previous works considered fitness cost *in vitro* for drug resistant *M. tuberculosis* strains. For example, Gagneux et al. (2006) and others (Mariam et al., 2004; Davies et al., 2000) have assessed fitness costs of drug resistance in *M. tuberculosis* using competition assays in different media. They measure fitness costs via differences in cell growth, also referred to as in vitro fitness cost. However, in vitro fitness cost is not equivalent to transmission fitness cost. Once transmission fitness cost for a range of drug resistant strains is quantified, one can assess the correlation between *in vitro* and observational transmission fitness costs. At present, no estimates of pyrazinamide resistance transmission fitness costs *in vitro* are available, so a direct comparison between different estimates is currently impossible.

Kendall et al. (2015) have used incidence data to estimate relative transmissibility of MDR-TB, showing that a predominant number (median 95.9%; 95% uncertainty range [68.0, 99.6]) of MDR-TB cases are due to transmission rather than de-novo resistance acquisition. On the other hand, Burgos et al. (2003) have used *M. tuberculosis* drug resistance and genotype data to cluster isolates from San Francisco and estimate drug resistance fitness costs in comparison to drug sensitive strains in the form of a proxy for the ratio of reproductive numbers. Their analyses show a high overall fitness cost of drug resistance, showing that on average the estimated reproductive number ratio of any drug-resistant to drug-susceptible TB strains is 0.51 (95% confidence interval [0.37, 0.69], resistant to isoniazid, streptomycin, or both). Moreover, their dataset shows no secondary MDR-TB cases, whereas our studied dataset shows signal for transmitted MDR-TB cases. Luciani et al. (2009) estimate that the relative fitness of drug-resistant strains varies from 0.3 in Venezuela to 1.0 in Cuba and Estonia, showing that depending on the country in question the fitness costs can vary drastically. Our estimate of around 0.64 for the relative fitness of pyrazinamide resistant strains in Kinshasa falls inside the confidence intervals both in Burgos et al. (2003) and Luciani et al. (2009).

In summary, Kendall et al. (2015) have no genomic information available, while Burgos et al. (2003) only make use of genotyping data to cluster isolates by similarity, rather than by inferring evolutionary relationships. In our analyses, we observed higher estimates of transmission fitness costs when ignoring the genomic data and hence the evolutionary relationship among samples (see Supplementary Figure 12). Luciani et al. (2009) estimate fitness costs from genetic data using approximate Bayesian computation, bundling multiple different resistant strain types together, which leads to averaging of possible costs in any different resistance type. While in this study we bundle substitutions on a single gene that confer resistance to a single drug, we make sure that the MDR substitutions are identical within a cluster.

Phylogenetic and transmission analyses have previously been performed on WGS *M. tuberculosis* data. These analyses have mainly focused on inferring the timing of epidemics and on inferring transmission networks with direction of transmission. Works such as Didelot et al. (2014, 2017) have used phylogenetic trees inferred by BEAST to infer transmission trees. On the other hand Klinkenberg et al. (2017) have implemented simultaneous transmission and phylogenetic tree estimation, which allows for more precise estimates of transmission event times, as unobserved events are unconstrained by the previously estimated phylogenetic trees. However, this latter tool was tested on densely sampled populations, and is thus not applicable to datasets as the one used here. Moreover, none of these tools allow us to infer parameters defining the dynamics of epidemic spread, such as the transmission rates and the relative fitness of distinct strains.

Based on phylogenetic and transmission analyses, one can attempt to make qualitative conclusions on specific strain fitness based on the clustering of the samples on phylogenetic and transmission trees. One phylogenetic analysis aiming at assessing transmission fitness cost based on WGS data is Casali et al. (2014), where the authors used whole genome sequences and their reconstructed phylogenetic relationships to investigate the transmission fitness costs of drug resistance. They used the clustering of isolates as an indicator of transmission fitness, where closely clustering isolates with an inferred common ancestor indicate transmitted resistance, whereas single isolates indicate acquired resistance. The authors speculate that the most prevalent substitution p.Ile6Leu in the pyrazinamide resistance gene *pncA*, which does not confer resistance *in vitro*, does confer clinical resistance with no reduced transmissibility. Unfortunately, this specific substitution is not present in the dataset from Kinshasa so it was impossible to check this hypothesis.

Our approach allows us to look at specific resistances and estimate their transmission success in relation to other strains circulating in the same epidemic. We quantify the transmission fitness cost of pyrazinamide resistant MDR *M. tuberculosis* to be around 36% relative to pyrazinamide-sensitive MDR *M. tuberculosis* strains. While it would be most interesting to investigate the transmission fitness costs for each specific resistance mutation, rather than assuming that the cost is the same for all mutations causing pyrazinamide resistance, we need to have a significant amount of sequences with identical *pncA* mutations to be able to estimate their specific transmission costs. Our

dataset, however, does not contain enough sequences exhibiting the same mutation in order to quantify its transmission fitness cost. Our model does not allow for co-infection, which could have occurred in the Kinshasa patients but is impossible to detect. The long culturing period of the strains required before sequencing (at least 6 weeks) results in outcompeting of any mixed infections and the sequencing of a single dominant clone. We lack information on possible confounding factors such as HIV co-infection status which could potentially influence our estimates. However, the possible association between pyrazinamide resistance and HIV status has been previously explored and no significant association was found Budzik et al. (2014).

The BDMM model explicitly accounts for incomplete sampling in the computation of the phylodynamic likelihood. This in turn means that difference in sampling strategies for the different types of analysed strains will not bias the results as long as this prior information on the sampling strategy is included using informative priors on the sampling proportions. It is important to also set the substitution rates depending on the time scales on which the sequences are available. If the data is available in a similar format to what is analysed here, e.g. in 12 SNP clusters covering a relatively short time period, estimates of the substitution rates for shorter time scales are more reasonable than estimates from the whole evolutionary history of *M. tuberculosis*. Additionally, it would be beneficial to include information on the rates of resistance acquisition.

In the future, we would like to estimate the relative transmission fitness of isoniazid and rifampicin resistances, the two resistances defining MDR-TB, compared to drug sensitive strains. Such analyses require datasets containing large numbers of both drug-sensitive and drug-resistant strains. Unfortunately, drug-sensitive *M. tuberculosis* strains are often of lower clinical interest and are therefore rarely sequenced. Indeed, mainly MDR strains were available for Kinshasa, thus such analyses were impossible. To our knowledge, no reasonably sized datasets of linked cases containing both sensitive and resistant strain sequences are available at the moment. Upon availability of such data sets, our approach could be employed to compare between-host fitness e.g. between strains with and without compensatory mutations, between HIV negative and positive patients, and between prison and non-prison-associated TB cases.

Overall, we show through simulation that we can use the modified MTBD method to analyse pathogens where the strains can be appropriately divided into two or more categories with distinct properties. Importantly, since we employ phylogenetic trees as a model for evolutionary histories, these pathogens may not recombine drastically on an epidemiological scale. This is fulfilled for TB and many viral pathogens where parts of the genome never recombine. Many other bacteria may not be appropriately analysed with this method due to e.g. frequent plasmid exchange. In this work we show that we can estimate the transmission fitness effects of additional resistance in MDR-TB. We expect that our method will more generally be useful to quantify the epidemic spread of drug resistances in a range of pathogens, and in that way may shed light on optimal treatment strategies aiming to avoid the selection for highly transmissible drug resistances causing epidemic outbreaks.

CRediT authorship contribution statement

Jūlija Pečerska: Conceptualization, Methodology, Software, Validation, Writing - Original Draft, Writing - review & editing, Visualization. **Denise Kühnert:** Conceptualization, Methodology, Software, Writing - review & editing. **Conor J. Meehan:** Data curation, Writing - review & editing. **Mireia Coscollá:** Writing - review & editing. **Bouke C. de Jong:** Data curation, Writing - review & editing. **Sebastien Gagneux:** Conceptualization, Writing - review & editing, Supervision. **Tanja Stadler:** Conceptualization, Methodology, Validation, Writing - original draft, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by SystemsX.ch.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.epidem.2021.100471.

The supplementary materials contain formulae for simulation parameter definition as well as additional figures illustrating parameter estimation.

References

Allen, R.C., Engelstadter, J., Bonhoeffer, S., McDonald, B.A., Hall, A.R., 2017. Reversing resistance: different routes and common themes across pathogens. Proc. Biol. Sci. 284 (1863), http://dx.doi.org/10.1098/rspb.2017.1619, URL https://www.ncbi.nlm.nih.gov/pubmed/28954914.

Andersson, D.I., Hughes, D., 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nat. Rev. Microbiol. 8 (4), 260–271.

Biek, R., Pybus, O.G., Lloyd-Smith, J.O., Didelot, X., 2015. Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. 30 (6), 306–313. http://dx.doi.org/10.1016/j.tree.2015.03.009, URL https://www.ncbi.nlm.nih.gov/pubmed/25887947.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4), e1003537.

Budzik, J.M., Jarlsberg, L.G., Higashi, J., Grinsdale, J., Hopewell, P.C., Kato-Maeda, M., Nahid, P., 2014. Pyrazinamide resistance, Mycobacterium tuberculosis lineage and treatment outcomes in San Francisco, California. PLoS One 9 (4), e95645. http://dx.doi.org/10.1371/journal.pone.0095645, URL https://www.ncbi.nlm.nih.gov/pubmed/24759760.

Burgos, M., DeRiemer, K., Small, P.M., Hopewell, P.C., Daley, C.L., 2003. Effect of drug resistance on the generation of secondary cases of tuberculosis. J. Infect. Dis. 188 (12), 1878–1884.

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S.R., Ignatyeva, O., Kontsevaya, I., Corander, J., Bryant, J., Parkhill, J., Nejentsev, S., Horstmann, R.D., Brown, T., Drobniewski, F., 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat. Genet. 46 (3), 279–286.

Casali, N., Nikolayevskyy, V., Balabanova, Y., Ignatyeva, O., Kontsevaya, I., Harris, S.R., Bentley, S.D., Parkhill, J., Nejentsev, S., Hoffner, S.E., Horstmann, R.D., Brown, T., Drobniewski, F., 2012. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res. 22 (4), 735–745. http://dx.doi.org/10.1101/gr.128678.111, URL https://www.ncbi.nlm.nih.gov/pubmed/22294518.

Chiang, C.-Y., Riley, L.W., 2005. Exogenous reinfection in tuberculosis. Lancet Infect. Dis. 5 (10), 629–636.

Cohen, T., Dye, C., Colijn, C., Williams, B., Murray, M., 2009. Mathematical models of the epidemiology and control of drug-resistant TB. Expert Rev. Respir. Med. 3 (1), 67–79.

Colijn, C., Cohen, T., 2016. Whole-genome sequencing of Mycobacterium tuberculosis for rapid diagnostics and beyond. Lancet Respir. Med. 4 (1), 6–8. http://dx.doi.org/10.1016/s2213-2600(15)00510-x.

Davies, A.P., Billington, O.J., Bannister, B.A., Weir, W.R., McHugh, T.D., Gillespie, S.H., 2000. Comparison of fitness of two isolates of Mycobacterium tuberculosis, one of which had developed multi-drug resistance during the course of treatment. J. Infect. 41 (2), 184–187. http://dx.doi.org/10.1053/jinf.2000.0711, URL https://www.ncbi.nlm.nih.gov/pubmed/11023769.

Didelot, X., Fraser, C., Gardy, J., Colijn, C., 2017. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. Mol. Biol. Evol. 34 (4), 997–1007. http://dx.doi.org/10.1093/molbev/msw275, URL https://www.ncbi.nlm.nih.gov/pubmed/28100788.

Didelot, X., Gardy, J., Colijn, C., 2014. Bayesian inference of infectious disease transmission from whole-genome sequence data. Mol. Biol. Evol. 31 (7), 1869–1879. http://dx.doi.org/10.1093/molbev/msu121, URL http://www.ncbi.nlm.nih.gov/pubmed/24714079.

Dowdy, D.W., Dye, C., Cohen, T., 2013. Data needs for evidence-based decisions: a tuberculosis modeler's 'wish list'. Int. J. Tuberc. Lung Dis. 17 (7), 866–877.

Feuerriegel, S., Schleusener, V., Beckert, P., Kohl, T.A., Miotto, P., Cirillo, D.M., Cabibbe, A.M., Niemann, S., Fellenberg, K., 2015. PhyResSE: a Web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data. J. Clin. Microbiol. 53 (6), 1908–1914. http://dx.doi.org/10.1128/JCM.00025-15, URL https://www.ncbi.nlm.nih.gov/pubmed/25854485.

Frost, S.D., Pybus, O.G., Gog, J.R., Viboud, C., Bonhoeffer, S., Bedford, T., 2015. Eight challenges in phylodynamic inference. Epidemics 10, 88–92. http://dx.doi.org/10.1016/j.epidem.2014.09.001, URL https://www.ncbi.nlm.nih.gov/pubmed/25843391.

Gagneux, S., Long, C.D., Small, P.M., Van, T., Schoolnik, G.K., Bohannan, B.J., 2006. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science 312 (5782), 1944–1946.

Gomes, M.G.M., Franco, A.O., Gomes, M.C., Medley, G.F., 2004. The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy. Proc. Biol. Sci. 271 (1539), 617–623.

Gomes, M.G.M., Rodrigues, P., Hilker, F.M., Mantilla-Beniers, N.B., Muehlen, M., Cristina Paulo, A., Medley, G.F., 2007. Implications of partial immunity on the prospects for tuberculosis control by post-exposure interventions. J. Theoret. Biol. 248 (4), 608–617.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A., Holmes, E.C., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303 (5656), 327–332. http://dx.doi.org/10.1126/science.1090727, URL https://www.ncbi.nlm.nih.gov/pubmed/14726583.

den Hertog, A.L., Sengstake, S., Anthony, R.M., 2015. Pyrazinamide resistance in Mycobacterium tuberculosis fails to bite? Pathog. Dis. 73 (6), ftv037.

Horne, D.J., Pinto, L.M., Arentz, M., Lin, S.Y.G., Desmond, E., Flores, L.L., Steingart, K.R., Minion, J., 2012. Diagnostic accuracy and reproducibility of WHO-endorsed phenotypic drug susceptibility testing methods for first-line and second-line antituberculosis drugs. J. Clin. Microbiol. 51 (2), 393–401. http://dx.doi.org/10.1128/jcm.02724-12.

Kendall, E.A., Fofana, M.O., Dowdy, D.W., 2015. Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis. Lancet Respir. Med. 3 (12), 963–972. http://dx.doi.org/10.1016/s2213-2600(15)00458-0.

Klinkenberg, D., Backer, J.A., Didelot, X., Colijn, C., Wallinga, J., 2017. Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. PLoS Comput. Biol. 13 (5), e1005495. http://dx.doi.org/10.1371/journal.pcbi.1005495, URL https://www.ncbi.nlm.nih.gov/pubmed/28545083.

Kühnert, D., Kouyos, R., Shirreff, G., Pečerska, J., Scherrer, A.U., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Günthard, H.F., Stadler, T., Bonhoeffer, S., Study, S.H.C., 2018. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. PLOS Pathogens.

Kühnert, D., Stadler, T., Vaughan, T.G., Drummond, A.J., 2016. Phylodynamics with migration: A computational framework to quantify population structure from genomic data. Mol. Biol. Evol. 33 (8), 2102–2116.

Luciani, F., Sisson, S.A., Jiang, H., Francis, A.R., Tanaka, M.M., 2009. The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. USA 106 (34), 14711–14715. http://dx.doi.org/10.1073/pnas.0902437106, URL https://www.ncbi.nlm.nih.gov/pubmed/19706556.

Ma, Y., Horsburgh, C., White, L.F., Jenkins, H.E., 2018. Quantifying TB transmission: a systematic review of reproduction number and serial interval estimates for tuberculosis. Epidemiol. Infect. 146 (12), 1478–1494. http://dx.doi.org/10.1017/S0950268818001760, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092233/.

Mariam, D.H., Mengistu, Y., Hoffner, S.E., Andersson, D.I., 2004. Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. Antimicrob. Agents Chemother. 48 (4), 1289–1294. http://dx.doi.org/10.1128/aac.48.4.1289-1294.2004.

Meehan, C.J., Moris, P., Kohl, T.A., Pečerska, J., Akter, S., Merker, M., Gehre, F., Lempens, P., Stadler, T., Kaswa, M.K., Kühnert, D., Niemann, S., de Jong, B.C., 2018. The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology 37, (ISSN: 23523964) pp. 410–416. http://dx.doi.org/10.1016/j.ebiom.2018.10.013, https://linkinghub.elsevier.com/retrieve/pii/S2352396418304249,

Merker, M., Barbier, M., Cox, H., Rasigade, J.P., Feuerriegel, S., Kohl, T.A., Diel, R., Borrell, S., Gagneux, S., Nikolayevskyy, V., Andres, S., Nubel, U., Supply, P., Wirth, T., Niemann, S., 2018. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. Elife 7, http://dx.doi.org/10.7554/eLife.38200, URL https://www.ncbi.nlm.nih.gov/pubmed/30373719.

Miotto, P., Cabibbe, A.M., Feuerriegel, S., Casali, N., Drobniewski, F., Rodionova, Y., Bakonyte, D., Stakenas, P., Pimkina, E., Augustynowicz-Kopec, E., Degano, M., Ambrosi, A., Hoffner, S., Mansjo, M., Werngren, J., Rusch-Gerdes, S., Niemann, S., Cirillo, D.M., 2014. Mycobacterium tuberculosis pyrazinamide resistance determinants: a multicenter study. MBio 5 (5), e01819–14.

Njire, M., Tan, Y., Mugweru, J., Wang, C., Guo, J., Yew, W., Tan, S., Zhang, T., 2016. Pyrazinamide resistance in Mycobacterium tuberculosis: Review and update. Adv. Med. Sci. 61 (1), 63–71. http://dx.doi.org/10.1016/j.advms.2015.09.007, URL https://www.ncbi.nlm.nih.gov/pubmed/26521205.

Pankhurst, L.J., del Ojo Elias, C., Votintseva, A.A., Walker, T.M., Cole, K., Davies, J., Fermont, J.M., Gascoyne-Binzi, D.M., Kohl, T.A., Kong, C., Lemaitre, N., Niemann, S., Paul, J., Rogers, T.R., Roycroft, E., Smith, E.G., Supply, P., Tang, P., Wilcox, M.H., Wordsworth, S., Wyllie, D., Xu, L., Crook, D.W., 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir. Med. 4 (1), 49–58. http://dx.doi.org/10.1016/s2213-2600(15)00466-x.

Pinho, S.T., Rodrigues, P., Andrade, R.F., Serra, H., Lopes, J.S., Gomes, M.G., 2015. Impact of tuberculosis treatment length and adherence under different transmission intensities. Theor. Popul. Biol.

Sengstake, S., Bergval, I.L., Schuitema, A.R., de Beer, J.L., Phelan, J., de Zwaan, R., Clark, T.G., van Soolingen, D., Anthony, R.M., 2017. Pyrazinamide resistance-conferring mutations in pncA and the transmission of multidrug resistant TB in Georgia. BMC Infect. Dis. 17 (1), 491.

Stadler, T., 2011. Inferring epidemiological parameters on the basis of allele frequencies. Genetics 188 (3), 663–672. http://dx.doi.org/10.1534/genetics.111.126466, URL https://www.genetics.org/content/188/3/663.

Vaughan, T.G., Drummond, A.J., 2013. A stochastic simulator of birth-death master equations with application to phylodynamics. Mol. Biol. Evol. 30 (6), 1480–1493.

Verver, S., Warren, R.M., Beyers, N., Richardson, M., van der Spuy, G.D., Borgdorff, M.W., Enarson, D.A., Behr, M.A., van Helden, P.D., 2005. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am. J. Respir. Crit. Care Med. 171 (12), 1430–1435.

Vynnycky, E., Fine, P.E., 1997. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. Epidemiol. Infect. 119 (2), 183–201.

Walker, T.M., Merker, M., Knoblauch, A.M., Helbling, P., Schoch, O.D., van der Werf, M.J., Kranzer, K., Fiebig, L., Kröger, S., Haas, W., Hoffmann, H., Indra, A., Egli, A., Cirillo, D.M., Robert, J., Rogers, T.R., Groenheit, R., Mengshoel, A.T., Mathys, V., Haanperä, M., Soolingen, D.v., Niemann, S., Böttger, E.C., Keller, P.M., Avsar, K., Bauer, C., Bernasconi, E., Borroni, E., Brusin, S., Coscollá Dévis, M., Crook, D.W., Dedicoat, M., Fitzgibbon, M., Gagneux, S., Geiger, F., Guthmann, J.-P., Hendrickx, D., Hoffmann-Thiel, S., van Ingen, J., Jackson, S., Jaton, K., Lange, C., Mazza Stalder, J., O'Donnell, J., Opota, O., Peto, T.E.A., Preiswerk, B., Roycroft, E., Sato, M., Schacher, R., Schulthess, B., Smith, E.G., Soini, H., Sougakoff, W., Tagliani, E., Utpatel, C., Veziris, N., Wagner-Wiening, C., Witschi, M., 2018. A cluster of multidrug-resistant Mycobacterium tuberculosis among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. Lancet Infect. Dis. 18 (4), 431–440. http://dx.doi.org/10.1016/s1473-3099(18)30004-5.

WHO, 2016. WHO Treatment Guidelines for Drug-Resistant Tuberculosis, 2016 Update. Report, WHO, URL http://www.who.int/tb/areas-of-work/drug-resistant-tb/treatment/resources/en/.

WHO, 2017. Global Tuberculosis Report 2017. Report, WHO.

WHO, 2018. Global Tuberculosis Report 2018. Report, WHO.

Yadon, A.N., Maharaj, K., Adamson, J.H., Lai, Y.P., Sacchettini, J.C., Ioerger, T.R., Rubin, E.J., Pym, A.S., 2017. A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. Nature Commun. 8 (1), 588. http://dx.doi.org/10.1038/s41467-017-00721-2, URL https://www.ncbi.nlm.nih.gov/pubmed/28928454.

Yakrus, M.A., Driscoll, J., Lentz, A.J., Sikes, D., Hartline, D., Metchock, B., Starks, A.M., 2014. Concordance between molecular and phenotypic testing of Mycobacterium tuberculosis complex isolates for resistance to rifampin and isoniazid in the United States. J. Clin. Microbiol. 52 (6), 1932–1937. http://dx.doi.org/10.1128/JCM.00417-14, URL https://www.ncbi.nlm.nih.gov/pubmed/24648563.

Yew, W.W., Leung, C.C., 2005. Are some people not safer after successful treatment of tuberculosis? Am. J. Respir. Crit. Care Med. 171 (12), 1324–1325.