nature genetics

Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis

Christopher B Ford[1], Rupal R Shah[1], Midori Kato Maeda[2], Sebastien Gagneux[3,4], Megan B Murray[5], Ted Cohen[5–7], James C Johnston[8], Jennifer Gardy[8,9], Marc Lipsitch[1,7] & Sarah M Fortune[1,10,11]

© 2013 Nature America, Inc. All rights reserved.

A key question in tuberculosis control is why some strains of *M. tuberculosis* are preferentially associated with resistance to multiple drugs. We demonstrate that *M. tuberculosis* strains from lineage 2 (East Asian lineage and Beijing sublineage) acquire drug resistances *in vitro* more rapidly than *M. tuberculosis* strains from lineage 4 (Euro-American lineage) and that this higher rate can be attributed to a higher mutation rate. Moreover, the *in vitro* mutation rate correlates well with the bacterial mutation rate in humans as determined by whole-genome sequencing of clinical isolates. Finally, using a stochastic mathematical model, we demonstrate that the observed differences in mutation rate predict a substantially higher probability that patients infected with a drug-susceptible lineage 2 strain will harbor multidrug-resistant bacteria at the time of diagnosis. These data suggest that interventions to prevent the emergence of drug-resistant tuberculosis should target bacterial as well as treatment-related risk factors.

Recently, strains of *M. tuberculosis* have emerged that are resistant to most or all effective antibiotics[1–4]. Given the low mutation rate of *M. tuberculosis*[5,6] and its slow replication rate, it is unclear how the bacterium acquires resistance to multiple antibiotics, especially in the face of concurrent multiple-drug treatment. The most commonly cited risk factors for treatment failure due to antibiotic resistance are patient noncompliance[1–4,7–9], inappropriate drug regimens and dosing[2,5,6,9,10], and primary infection with drug-resistant strains[11,12]. The relative contributions of bacterial determinants to treatment failure have been unclear. Recently, whole-genome sequencing of *M. tuberculosis* isolates showed the key role of newly arising mutation in the emergence of drug resistance[13–15]. Sequencing of *M. tuberculosis* from patients in whom antibiotic therapy was unsuccessful showed that new mutations conferring antibiotic resistance can arise multiple times within a given individual[15]. Moreover, several studies suggest that certain strains of *M. tuberculosis* may be more frequently associated with multidrug resistance (MDR)[16,17]. Given that all drug resistance in *M. tuberculosis* occurs through chromosomal mutation, these data suggest that the mutational capacity of the bacterium may be a key determinant of the likelihood of developing drug resistance.

M. tuberculosis forms phylogeographic lineages associated with particular human populations[18–21]. Although *M. tuberculosis* is less genetically diverse than many other pathogens, there is both experimental and clinical evidence that *M. tuberculosis* strains from different lineages vary in their capacity to cause disease[20–24] and acquire drug resistance[11,12,16,20,25–28]. Specifically, *M. tuberculosis* strains in lineage 2 (the East Asian lineage, which includes the Beijing family of strains) have been epidemiologically associated with increased risk of drug resistance in several cross-sectional studies in diverse locales, although not in all[29]. Strains from this lineage have polymorphisms in genes involved in DNA replication, recombination and repair relative to strains from lineage 4 (the Euro-American lineage), raising the possibility that they are more mutable than other *M. tuberculosis* strains[30]. However, these epidemiological observations might also reflect social and programmatic factors (such as noncompliance and inappropriate dosing) correlating with the phylogeography of the lineage 2 strains. Indeed, *in vitro* studies comparing the rate of acquisition or the frequency of drug resistance in lineage 2 and lineage 4 strains have produced differing results[31,32]. Here we sought to determine the rate at which *M. tuberculosis* strains of different lineages acquire drug resistance and the effect of strain-based differences in mutation rate on the predicted *de novo* generation of MDR in individuals with tuberculosis.

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA. [2]Curry International Tuberculosis Center, Division of Pulmonary and Critical Care Medicine, University of California, San Francisco, San Francisco, California, USA. [3]Swiss Tropical & Public Health Institute, Basel, Switzerland. [4]University of Basel, Basel, Switzerland. [5]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [6]Division of Global Health Equity, Brigham and Women's Hospital, Boston, Massachusetts, USA. [7]Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [8]Communicable Disease Prevention and Control Services, British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada. [9]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada. [10]Ragon Institute of MGH, MIT and Harvard, Boston, Massachusetts, USA. [11]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. Correspondence should be addressed to S.M.F. (sfortune@hsph.harvard.edu).

Received 12 September 2012; accepted 6 May 2013; published online 9 June 2013; doi:10.1038/ng.2656

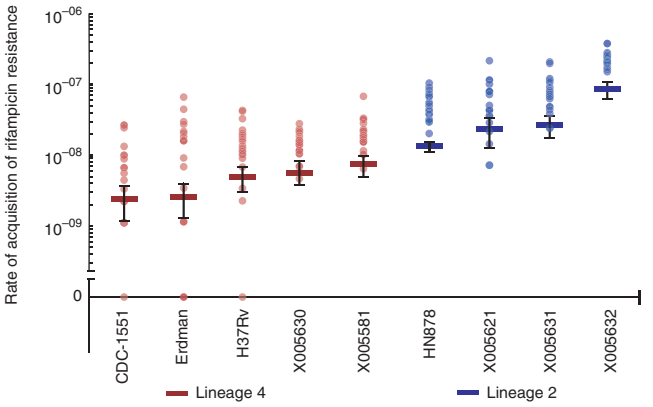


Figure 1 Lineage 2 strains more rapidly acquire rifampicin resistance. Fluctuation analysis was used to determine the rates at which clinical and laboratory strains from both lineage 2 and lineage 4 acquired resistance to rifampicin (2 μg/ml). Circles represent mutant frequency (number of mutants per cell plated in a single culture), where darker circles represent multiple cultures with the same frequency. Bars represent the estimated mutation rates (mutations conferring rifampicin resistance per generation), with error bars representing the 95% confidence intervals. Strains are shown on the *x* axis, and the acquisition rate of rifampicin resistance is shown on the *y* axis in log scale. Fluctuation analysis parameters are listed in **Supplementary Table 1**.

RESULTS

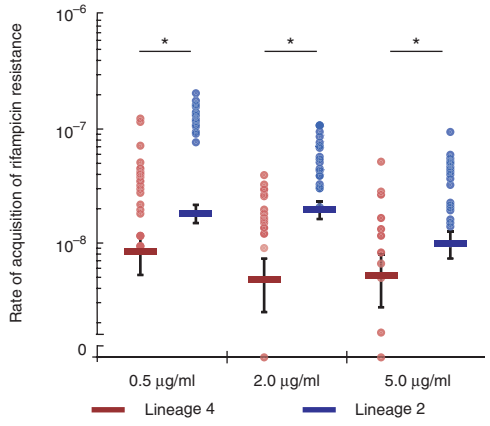
Effect of mutation and genetic background on drug resistance

To measure the rate at which drug resistance is acquired in strains from different *M. tuberculosis* lineages, we performed Luria-Delbrück fluctuation analysis[33,34] on a panel of laboratory and clinical isolates from both lineage 2 and lineage 4. All strains were fully drug susceptible, with minimum inhibitory concentrations (MICs) at least 100-fold lower than the drug concentrations at which we assessed the acquisition of drug resistance (MIC (rifampicin) < 0.015 μg/ml; MIC (isoniazid) < 0.007 μg/ml; **Supplementary Table 1**). Within both lineages, there was some strain-to-strain variation in the rate at which rifampicin resistance was acquired (**Fig. 1**). However, every strain from lineage 2 acquired resistance to rifampicin (2 μg/ml) at a significantly higher rate than every lineage 4 strain, with a nearly tenfold difference between the means of the two groups (**Fig. 1** and **Supplementary Table 1**).

The higher rate of acquisition of rifampicin resistance in lineage 2 strains could reflect three possible mechanisms: (i) differences in ability to survive and mutate after exposure to antibiotic, (ii) inherent differences in the number of *rpoB* mutations conferring rifampicin resistance (target size) or (iii) differences in the basal mutation rate in the absence of selection. To test these hypotheses, we chose well-characterized representatives from lineage 2 and lineage 4—HN878 and CDC-1551, respectively—for further study.

Differences in response to antibiotic

We first sought to determine whether lineage 2 and lineage 4 strains differed in their ability to acquire drug resistance after exposure to rifampicin. Fluctuation analysis assumes that all mutations occur before selection and that all mutants replicate as well as wild type[33]; however, if a strain is capable of surviving and mutating in the presence of a drug, it will produce a greater number of drug-resistant mutants. Building on similar studies in *Saccharomyces cerevisiae*[35], we reasoned that, if mutations occur in the presence of a drug, then lowering the drug concentration might allow strains from both lineages to grow and acquire mutations following exposure. Conversely, increasing the drug concentration might abrogate the ability of both strains to survive and acquire mutations in the presence of the drug. However, we found that statistically significant increases in the acquisition rate of rifampicin resistance for the lineage 2 strain HN878 relative to the lineage 4 strain CDC-1551 were maintained over tenfold variation in drug concentration (0.5–5 μg/ml) (**Fig. 2** and **Supplementary Table 1**).



To extend these findings, the distribution of mutants observed in each fluctuation assay can be analyzed using previously developed tools[35]. This analysis takes advantage of the fact that mutations occurring in culture, before antibiotic exposure, result in a Luria-Delbrück distribution of mutants (number of mutants per culture). Although mutations arise according to a Poisson distribution, the subsequent outgrowth of mutants in broth culture generates a Luria-Delbrück distribution. In contrast, mutations occurring after plating with antibiotic occur according to a Poisson distribution, without the expansion in broth culture that creates a Luria-Delbrück distribution[33,36,37]. Thus, the appearance of any additional mutants resulting from the acquisition of resistance after antibiotic exposure will generate a mixed distribution of mutants, containing mutants with a Poisson distribution arising after plating with drug and mutants with a Luria-Delbrück distribution arising before drug exposure.

We therefore used a curve-fitting approach to determine whether the distribution of mutants (number of mutants per culture) in the two lineages was better fit using a one-parameter Luria-Delbrück model or a two-parameter Luria-Delbrück and Poisson mixture model (**Fig. 3a–f**). The algorithm first fits the data to a Luria-Delbrück model alone and then optimizes the fit with the addition of a Poisson model. A purely Poisson model was also fit to serve as a reference. We used the Akaike information criterion with correction for sample size (AIC_C) to identify instances where the Luria-Delbrück model alone provided optimal fit compared to either the Poisson model or the two-parameter mixture model[38,39]. AIC_C quantifies the evidence in favor of choosing a more complex model (here, the two-parameter mixture model) over a simpler model

Figure 2 Altering drug concentration does not alter the observation that lineage 2 strains more rapidly acquire rifampicin resistance. Fluctuation analysis was used to determine the rates at which resistance was acquired to rifampicin (0.5 and 2, 5 μg/ml) in representative strains from both lineage 2 and lineage 4. Circles represent mutant frequency (number of mutants per cell plated in a single culture), where darker circles represent multiple cultures with the same frequency. Bars represent the estimated mutation rates, with error bars representing the 95% confidence intervals. Significance was determined by comparing strain pairs using the Wilcoxon rank-sum test; $*P < 0.05$. Strains are shown on the *x* axis, and the rate at which rifampicin resistance is acquired is shown on the *y* axis in log scale. Fluctuation analysis parameters are listed in **Supplementary Table 1**.

© 2013 Nature America, Inc. All rights reserved.

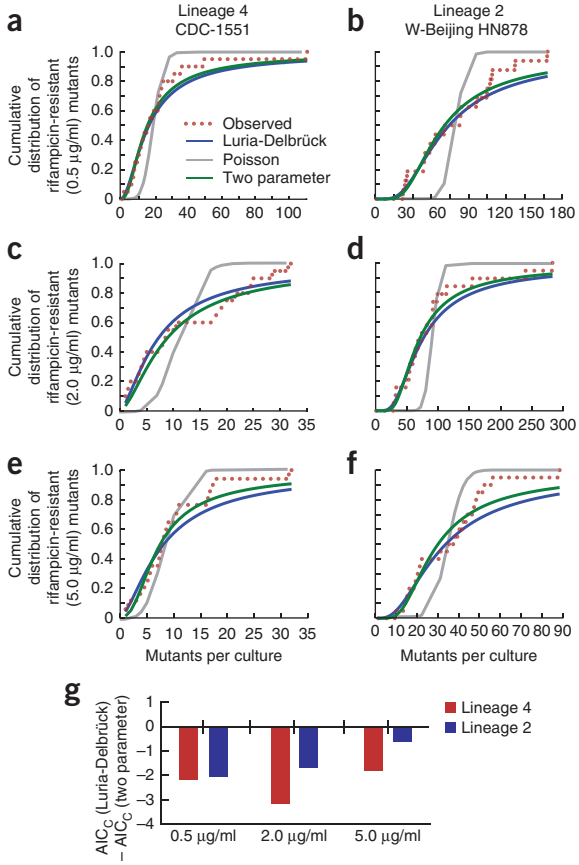
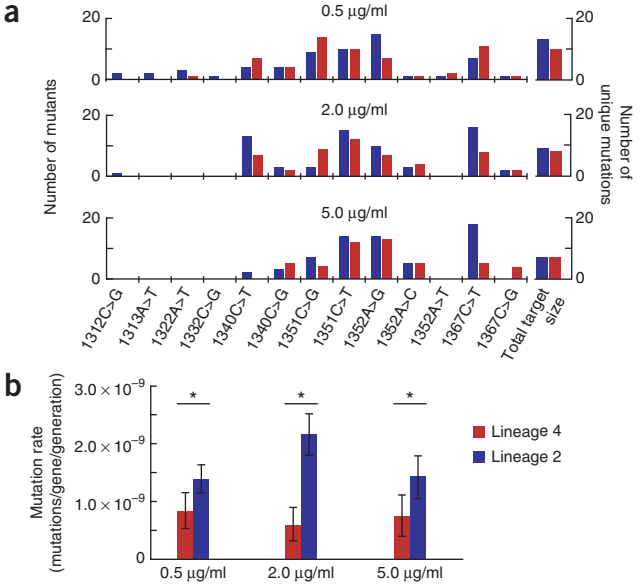


Figure 3 The cumulative distribution of drug-resistant mutants from lineage 2 and lineage 4 indicates that mutations do not occur after exposure to antibiotic. (**a**–**f**) Curve-fitting analysis was performed to determine whether the cumulative distribution of the fluctuation analysis data for strains plated in antibiotic better fit a one-parameter Luria-Delbrück model, a one-parameter Poisson model or a two-parameter Luria-Delbrück and Poisson mixture model. The number of mutants per culture is shown on the *x* axis, and the probability of observing *x* or fewer mutants per culture is shown on the *y* axis. Results are shown for lineage 4 strain CDC-1551 plated on rifampicin (0.5 µg/ml) (**a**), lineage 2 strain HN878 plated on rifampicin (0.5 µg/ml) (**b**), lineage 4 strain CDC-1551 plated on rifampicin (2 µg/ml) (**c**), lineage 2 strain HN878 plated on rifampicin (2 µg/ml) (**d**), lineage 4 strain CDC-1551 plated on rifampicin (5 µg/ml) (**e**) and lineage 2 strain HN878 plated on rifampicin (5 µg/ml) (**f**). (**g**) To determine which model best fit each data set, we determined AIC_C values. A smaller AIC_C represents a better fit given a penalty for more parameters in a model. If AIC_C (Luria-Delbrück model) is smaller than AIC_C (two-parameter model), then the resulting value will be negative, reflecting a better fit for the Luria-Delbrück model (**Supplementary Table 2**).

(here, the Luria-Delbrück model), appropriately penalizing the fit of the more complex model by the increase in model complexity. As expected, ΔAICc (AIC_C (Luria-Delbrück model) – AIC_C (Poisson model)) was highly negative under all conditions, confirming that the Luria-Delbrück model fit the distributions markedly better than the Poisson model alone (**Supplementary Table 2**). More revealing, ΔAICc (AIC_C (Luria-Delbrück model) – AIC_C (two-parameter model)) was also negative, indicating that there was insufficient evidence to support the inclusion of an additional Poisson component in the Luria-Delbrück distributions (**Fig. 3g** and **Supplementary Table 2**). This finding suggests that mutation acquired after exposure to antibiotic is not responsible for the higher rate with which the lineage 2 strain HN878 acquires rifampicin resistance.

This analytic approach also suggests that the difference in the rates at which rifampicin resistance occurs is not due to strain-based differences in the fitness effects of the mutations conferring drug resistance[33,37]. If the drug-resistant mutants in either strain were subject to a high fitness cost, the outgrowth of mutants in culture before selection would have been slower than that of drug-susceptible cells, driving the Luria-Delbrück distribution back toward the underlying Poisson distribution of mutation. However, our data suggest that, in both strains, drug-resistant mutants occur primarily according to a Luria-Delbrück distribution.



Differences in target size

Rifampicin resistance is encoded by multiple mutations in the rifampicin resistance–determining region (RRDR) of *rpoB*. Strains in which there are a greater number of potential mutations in the RRDR that produce drug resistance would more rapidly acquire resistance to rifampicin. Therefore, we sought to determine whether *M. tuberculosis* strains of different lineages differ in the total number of mutations that confer rifampicin resistance, accounting for differences in the rates at which rifampicin resistance occurs. We sequenced the RRDR of *rpoB* in 600 independent mutants (100 from each fluctuation assay in **Fig. 2**) to determine the number of mutations conferring rifampicin resistance in both strains under each condition tested (**Fig. 4a** and **Supplementary Table 3**). Although there is potentially a discrepancy between the mutations conferring drug resistance that

Figure 4 Small differences in target size and differences in basal mutation rate are responsible for the observed differences in the rate of acquisition of drug resistance. (**a**) The target size of each strain under each condition was determined by sequencing the RRDRs of 100 isolates from each strain under each condition. Each mutation is shown on the *x* axis, with coordinates representing positions within *rpoB* (Rv0667). The number of mutants per strain uniquely formed within a culture is shown on the *y* axis. The lineage 4 strain CDC-1551 is shown in red, and the lineage 2 strain HN878 is shown in blue. Target size (the number of unique mutations conferring rifampicin resistance) is shown on the right. (**b**) The per-base mutation rate is determined by normalizing the rate of acquisition of rifampicin resistance by target size. Rifampicin concentration is shown on the *x* axis, and mutation rate per base pair is shown on a linear scale on the *y* axis. Significance was determined by comparing strain pairs using the Wilcoxon rank-sum test; $*P < 0.05$; error bars represent 95% confidence intervals. Values are listed in **Supplementary Table 4**.

© 2013 Nature America, Inc. All rights reserved.

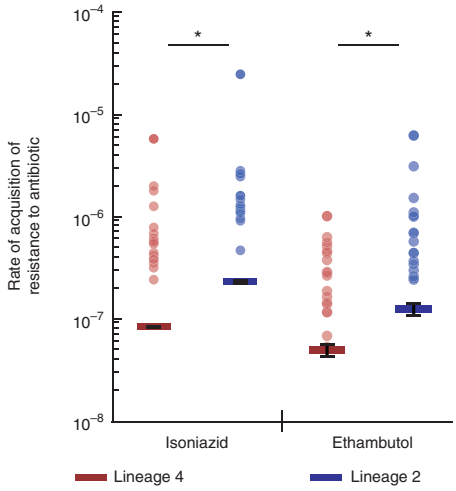


Figure 5 A representative lineage 2 strain acquires isoniazid and ethambutol resistance at a higher rate than a strain from lineage 4. Fluctuation analysis was used to determine the rates at which the lineage 4 strain CDC-1551 and the lineage 2 strain HN878 acquired resistance to isoniazid (1 μg/ml) and ethambutol (5 μg/ml). Circles represent mutant frequency (number of mutants per cell in a single culture), where darker circles represent multiple cultures with the same frequency. Bars represent the estimated mutation rates, with error bars representing the 95% confidence intervals. Significant differences were defined by non-overlapping 95% confidence intervals; $^*P < 0.05$ in comparison of mutant frequencies by Wilcoxon rank-sum test. Values are listed in **Supplementary Table 1**.

are found *in vitro* and *in vivo*[40], all of the mutations that we identified corresponded to mutations seen clinically[41]. For both strains, the target size became smaller as drug concentration increased, indicating that some *rpoB* mutations generate lower level rifampicin resistance. There were small differences in target size between the two strains at two of the three drug concentrations tested, such that the numbers of mutations conferring resistance at both 0.5 and 2 μg/ml were higher in the lineage 2 strain. These data suggest that target size differences between strains contribute to strain-based differences in the acquisition of rifampicin resistance. However, correcting for target size to determine the per-base mutation rate, we found that the lineage 2 strain HN878 still had a significantly higher rate at which resistance was required compared to the lineage 4 strain CDC-1551 at each of the drug concentrations tested (**Fig. 4b** and **Supplementary Table 4**). Therefore, it is likely that differences in basal mutation rate lead to differences in the acquisition of rifampicin resistance.

Resistance to other antibiotics

If the mutation rate of HN878 is higher than that of CDC-1551, then the lineage 2 *M. tuberculosis* strain should also acquire resistances to other antibiotics at a higher rate. We therefore assessed the rates at which HN878 and CDC-1551 acquired resistance to ethambutol (5 μg/ml) and isoniazid (1 μg/ml). For both antibiotics, the rate of resistance was nearly 3-fold (2.51-fold and 2.75-fold, respectively) higher in the lineage 2 strain HN878, consistent with the increased rate at which this strain acquired rifampicin resistance (**Fig. 5** and **Supplementary Table 1**). Taken together, these results suggest that *M. tuberculosis* strains from lineage 2 have a higher basal mutation rate than strains from lineage 4.

In vitro mutation rates to predict *in vivo* resistance

We then sought to understand how these *in vitro* measures of mutation translate to the *in vivo* environment. In our previous work, we determined that, in nonhuman primates, *M. tuberculosis* mutates at a relatively fixed rate over time, and this *in vivo* per-day mutation rate is well approximated by the *in vitro* per-day mutation rate as measured by fluctuation analysis and adjusted for target size[6]. To determine whether the *in vitro* mutation rate is similarly concordant with the mutation rate of *M. tuberculosis* during human infection, we analyzed the whole-genome sequences of *M. tuberculosis* isolates derived from an outbreak of a lineage 4 strain in British Columbia, Canada[42]. We determined the number of SNPs in each strain relative to a historical isolate, identifying SNPs according to parameters that were experimentally validated through Sanger resequencing in our previous work (**Fig. 6a**). By reconstructing the phylogeny of these strains through Bayesian Markov chain Monte Carlo (MCMC) analysis[43,44] (**Supplementary Fig. 1**) and informing the phylogeny with dates for each isolate, we have estimated the base substitution rate (equivalent to the mutation rate under a neutral model of evolution[45]) in this outbreak. Notably, we found that the British Columbia strains acquired mutations at approximately the same rate over time as previously shown for *M. tuberculosis* strains isolated from macaques with active and latent disease, irrespective of disease course (**Fig. 6b** and **Supplementary Table 5**). In addition, the rate at which these strains acquired mutations *in vivo* was well approximated by the *in vitro* per-day mutation rate of the lineage 4 strain (Erdman) used in the macaque infections as defined by fluctuation analysis. Finally, consistent with both our previous work[6] and the work of others[46], these data indicate that a molecular clock of 0.3–0.5 mutations per genome per year may be applicable to the analysis of *M. tuberculosis* genetic diversity over the time scales assessed here (approximately 10 years).

A model of drug resistance predicts MDR before treatment

Given these data, we developed a stochastic simulation model of the evolution of drug resistance within an infected human host to assess the potential clinical impact of the observed differences in mutation rate between lineage 2 and lineage 4 strains. Our model of drug resistance uses a stochastic mutation parameter in which mutation occurs at a constant rate over time, and we informed this parameter with the *in vitro* mutation rates for CDC-1551 and HN878 as proxies

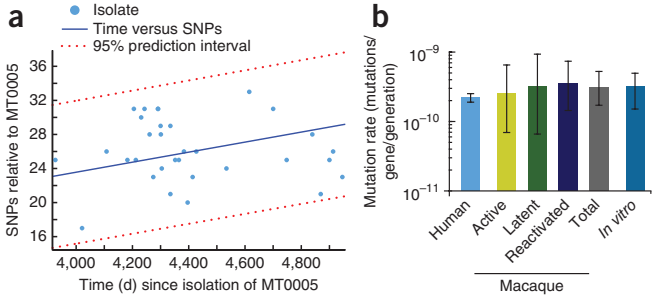


Figure 6 Bayesian MCMC analysis shows a mutation rate in humans similar to that estimated in strains from the macaque model and *in vitro*. (**a**) The number of SNPs and the number of days separating clinical isolates and MT0005 are plotted. SNPs located in repeat regions (PE_PGRSs, PPEs and transposable elements) were excluded, consistent with our previous analysis[6]. The data are fit to a first-order polynomial to show the trend (dark blue line). (**b**) Estimates of mutation rate in human isolates were derived by reconstructing the phylogeny of the isolates in **a**. Mutation rate is shown on the *y* axis in log scale. Estimates of mutation rate from the macaque model and the infecting strain Erdman (*in vitro*) were determined previously[6]. Error bars represent 95% confidence intervals.

© 2013 Nature America, Inc. All rights reserved.

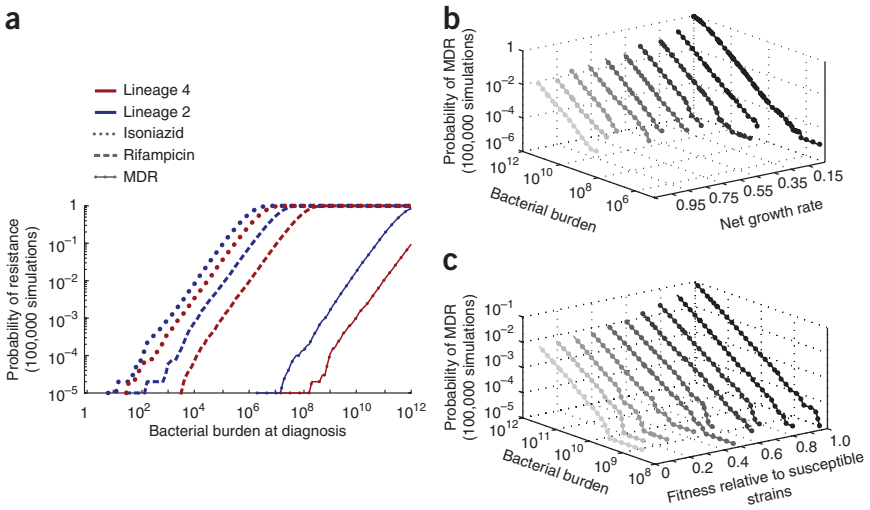


Figure 7 A stochastic simulation mathematical model predicts the emergence of multidrug-resistant tuberculosis before the onset of treatment. (**a**) Estimates of the probability of observing MDR in a population were derived using a stochastic mathematical model of resistance in 200,000 simulations, 100,000 for each for the strains from lineage 2 and lineage 4. Model parameters are listed in **Supplementary Table 6**. Bacterial burden at diagnosis is shown on the *x* axis, and the probability of observing resistance is shown on the *y* axis in log scale. (**b**,**c**) To determine the sensitivity of our model to variations in growth rate (**b**) and fitness (**c**), we varied each parameter (**Supplementary Table 6**) and determined the probability of observing resistance (*z* axis, log scale) at any given bacterial burden (*y* axis, log scale) for a specified parameter set (*x* axis).

for their mutation rates in the human host (**Supplementary Fig. 2a** and **Supplementary Table 6**). We used this model to simulate the emergence of MDR (defined here as resistance to both rifampicin and isoniazid) in an infected individual before diagnosis and treatment (**Supplementary Fig. 2b**).

In the model, as a result of the differences in mutation rate, individuals infected with the lineage 2 strain HN878 were at a notably increased risk of acquiring MDR before treatment compared to individuals infected with the lineage 4 strain CDC-1551 (**Fig. 7a**). When all other parameters (bacterial birth and death rates, fitness and bacterial burden at the time of diagnosis) were kept equal, the difference in the probability of MDR before diagnosis and treatment was approximately 22-fold. We found similar results when using an alternative model of drug resistance in which mutation is replication rather than time dependent (**Supplementary Fig. 2c**)[47]. We assessed the sensitivity of our model to fluctuations in both growth rate and fitness (**Fig. 7b**,**c**). Varying these parameters did not alter our principal conclusion that individuals infected with lineage 2 strains of *M. tuberculosis* are at notably higher risk of the *de novo* acquisition of MDR, reflecting the multiplicative effects of the increased risks associated with acquiring resistance to individual drugs due to a higher basal mutation rate.

DISCUSSION

Here we demonstrate that strains from lineage 2 of *M. tuberculosis* (the East Asian lineage, which includes the Beijing family of strains) acquire resistance to drugs *in vitro* more rapidly than strains from lineage 4 (the Euro-American lineage). This difference is likely not the result of enhanced ability of lineage 2 strains to survive and mutate in the presence of drug, and we find no evidence of strong fitness effects that would explain the observed differences in the rate at which drug resistance is acquired. Notably, we do find evidence that the genetic context of a given *M. tuberculosis* strain can affect the range of observed mutations conferring resistance to a single drug. In our analysis, the lineage 2 strain HN878 was permissive for a broader range of *rpoB* mutations compared to the lineage 4 strain CDC-1551. However, this difference in target size is not sufficient to explain the observed difference in the rates at which rifampicin resistance occurs, suggesting that a basal difference in mutation before selection drives the accelerated rate at which drug resistance is acquired in the HN878 strain. In support of this notion, we find that HN878 more rapidly acquires resistance to not only rifampicin but also to isoniazid and ethambutol.

We also find significant variation in the mutation rates of the strains within each lineage. Previous analyses suggested that there is substantial genetic and phenotypic diversity among isolates from the various *M. tuberculosis* lineages[48]. We expect that differences in mutation rate and in target size both contribute to the 2- to 35-fold differences in the rates at which rifampicin resistance is acquired that we have measured in these other strains. Further work will be required to establish the relative contribution of these factors to the rate of acquisition of drug resistance for each strain and to determine to what extent these findings can be generalized to strains from other *M. tuberculosis* lineages.

To establish the *in vivo* relevance of these findings, we sought to assess the concordance between mutation rates measured *in vitro* and *in vivo*. Notably, we found that the mutation rate of *M. tuberculosis* *in vitro* is very close to the mutation rate—assessed as mutation per unit time—in isolates from a human transmission chain. Thus, *M. tuberculosis* acquires mutations at a similar rate over time in people as it does in an actively growing culture *in vitro*. This observation is consistent with our previous finding that the *in vitro* mutation rate over time was similar to the rate of mutation over time in *M. tuberculosis* isolated from macaques with latent and active disease. Taken together, these findings define a molecular clock for *M. tuberculosis* that may be used in future evolutionary and epidemiological studies, although care must be taken to identify and compensate for potential sources of variation in mutation rate[49,50].

We propose two possible explanations for the finding that *M. tuberculosis* acquires mutations at the same rate over time *in vitro*, in macaques and across a human transmission chain. First, there may be a population of *M. tuberculosis* replicating *in vivo* at a rate similar to the replication rate *in vitro*. These bacteria may be over-represented in culturable clinical isolates, suggesting that they may be more likely to cause disease and be transmitted. Alternatively, it is possible that the mutation rate of *M. tuberculosis* is driven by a time-dependent rather than a replication-dependent factor. For example, the replicative error rate in *M. tuberculosis* could be very low with respect to time, and mutations may occur both *in vitro* and *in vivo* largely through DNA damage from endogenous metabolic processes or exogenous stressors.

We expect that these models may be resolved in part by elucidating the molecular basis of strain-based differences in mutation rate. The differences in mutation rate among the clinical *M. tuberculosis* strains measured here are more modest than the differences that distinguish

© 2013 Nature America, Inc. All rights reserved.

clinical isolates of other bacteria such as *Pseudomonas aeruginosa* and *Escherichia coli*[51–53]. Clinical isolates of these pathogens may become orders of magnitude more mutable than wild-type strains through the loss of mismatch repair[51]. However, mycobacteria, like all other actinomycetes, lack mismatch repair entirely[54,55], and the molecular basis of replicative fidelity in mycobacteria remains unclear. Although mutations in DNA replication and repair genes are enriched in some *M. tuberculosis* strains from lineage 2 (ref. 30), no single point mutation has been found to accelerate the basal mutation rate of *M. tuberculosis* in isogenic strains. Notably, lineage 2 strains also differ from lineage 4 strains in important metabolic pathways[56,57]. Thus, it is possible that genetic differences outside of those affecting DNA replication and repair contribute to the differences in mutation rate that we have measured.

We have used the observation that *M. tuberculosis* mutates at a constant rate per unit time to develop a predictive model of the evolution of MDR *in vivo*. Our model demonstrates that it is possible to observe the evolution of MDR before the onset of treatment, consistent with the predictions of a previous replication-dependent model[47]. Moreover, differences in mutation rate have approximately multiplicative effects for each mutation acquired, leading to stark differences in the *de novo* emergence of MDR. Indeed, we parameterized our model with data from the HN878 strain, which has only modestly elevated acquisition rates of rifampicin and isoniazid resistance. Strain X005632 has a 35-fold higher rate for the development of rifampicin resistance compared to CDC-1551. With X005632 having a similarly elevated rate for isoniazid resistance, the risk of MDR in an individual infected with X005632 would be nearly three orders of magnitude higher than for an individual infected with CDC-1551. Predicted differences in the occurrence of at least one multidrug-resistant bacterium within infected individuals may not be directly proportional to the clinical risk of drug-resistant tuberculosis because individual bacteria may not have equal capacity to cause disease. However, the magnitude of these differences suggests that strain-based variation in the acquisition of drug resistance is a key risk factor in the development of drug-resistant disease. Indeed, just as strains with higher mutation rates have increased risk of acquiring mutations that confer drug resistance and mutations that facilitate the acquisition of drug resistance (for example, those affecting efflux pumps[58–60]), these strains have increased capacity to acquire compensatory mutations[61] that may enhance their fitness *in vivo*.

Although our model focused on the evolution of MDR (rifampicin and isoniazid resistance), the findings are applicable to the evolution of resistance to any antibiotic. New antibiotics and novel regimens of new and current antibiotics are being developed, combining drugs for which resistance is not yet prevalent. The expectation is that, with proper implementation, a novel regimen will treat patients infected with multidrug-resistant *M. tuberculosis* and prevent the emergence of resistance to new drugs[62,63]. However, both clinical and high-resolution sequencing findings suggest that patients in whom therapy is unsuccessful are infected with strains that are resistant to only a subset of the antibiotics administered[14,15,64]. Thus, even in the context of a novel regimen, resistance to these new antibiotics may be difficult to avoid, especially in the context of infection by strains from lineage 2.

Consistent with epidemiological data suggesting that severe disease at diagnosis is associated with the acquisition of MDR in new cases[65,66], our model also predicts that bacterial burden is a key determinant of the probability of drug resistance. In the face of substantial capacity for mutation and resistance, early and active case detection with novel, sensitive point-of-care diagnostics remains our best hope of curbing the drug resistance epidemic. Smear microscopy is the most common primary diagnostic for *M. tuberculosis* around the world but is orders of magnitude less sensitive than both culture and molecular diagnostics[67,68]. If, as our model suggests, higher bacterial burden at diagnosis results in increased risk for the evolution of multidrug-resistant *M. tuberculosis*, then improving diagnostic sensitivity will not only curtail ongoing transmission of disease, as previously suggested[69,70], but will also limit the *de novo* emergence of drug resistance. The risk of drug resistance seems to be even greater in the setting of infection with lineage 2 strains of *M. tuberculosis*. Taken together, these data emphasize that *M. tuberculosis* strains differ in their propensity for acquiring drug resistance and suggest that these biological factors should be considered in efforts to limit the emergence of new resistance to both existing antibiotics and new treatment regimens.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. Whole-genome sequencing of human isolates was performed previously[42], and reads were deposited in the NCBI Sequence Read Archive (SRA) under accession SRA020129.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by a New Innovator's Award (DP2 0D001378) from the Director's Office of the US National Institutes of Health (S.M.F.), by a subcontract from the National Institute of Allergy and Infectious Diseases (U19 AI076217 to S.M.F. and M.B.M.), by the US National Institutes of Health Models of Infectious Disease Agent Study program, through cooperative agreement 1 U54 GM088558 (M.L.), by a Howard Hughes Medical Institute Physician Scientist Early Career Award (S.M.F.), by a Merit Fellowship from the Harvard University Graduate School of Arts and Sciences (C.B.F.) and by a Doris Duke Charitable Foundation Clinical Scientist Development Award (S.M.F.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences, National Institute of Allergy and Infectious Diseases or the US National Institutes of Health.

AUTHOR CONTRIBUTIONS

C.B.F. designed the study, performed experimental and molecular studies, developed the mathematical model and conducted analyses, prepared the figures and drafted the manuscript. R.R.S. performed experimental studies. M.K.M. and S.G. isolated clinical strains. M.B.M. and T.C. advised development of the mathematical model. J.C.J. and J.G. contributed to analyses. M.L. advised design of the study and data analysis, including development of the mathematical model. S.M.F. designed the study, supervised experimental and molecular studies, and drafted the manuscript. All authors edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Velayati, A.A. *et al.* Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* **136**, 420–425 (2009).
2. Udwadia, Z.F., Amale, R.A., Ajbani, K.K. & Rodrigues, C. Totally drug-resistant tuberculosis in India. *Clin. Infect. Dis.* **54**, 579–581 (2012).
3. Migliori, G.B., De Iaco, G., Besozzi, G., Centis, R. & Cirillo, D.M. First tuberculosis cases in Italy resistant to all tested drugs. *Euro Surveill.* **12**, E070517.1 (2007).
4. Gandhi, N.R. *et al.* Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* **375**, 1830–1843 (2010).
5. David, H.L. Probability distribution of drug-resistant mutants in unselected populations of *Mycobacterium tuberculosis*. *Appl. Microbiol.* **20**, 810–814 (1970).
6. Ford, C.B. *et al.* Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* 10.1038/ng.811 (2011).
7. Bradford, W.Z. *et al.* The changing epidemiology of acquired drug-resistant tuberculosis in San Francisco, USA. *Lancet* **348**, 928–931 (1996).

© 2013 Nature America, Inc. All rights reserved.

8. Goble, M. *et al.* Treatment of 171 patients with pulmonary tuberculosis resistant to isoniazid and rifampin. *N. Engl. J. Med.* **328**, 527–532 (1993).
9. Pablos-Méndez, A., Knirsch, C.A., Barr, R.G., Lerner, B.H. & Frieden, T.R. Nonadherence in tuberculosis treatment: predictors and consequences in New York City. *Am. J. Med.* **102**, 164–170 (1997).
10. Udwadia, Z.F., Pinto, L.M. & Uplekar, M.W. Tuberculosis management by private practitioners in Mumbai, India: has anything changed in two decades? *PLoS ONE* **5**, e12023 (2010).
11. Johnson, R. *et al.* Drug-resistant tuberculosis epidemic in the Western Cape driven by a virulent Beijing genotype strain. *Int. J. Tuberc. Lung Dis.* **14**, 119–121 (2010).
12. Streicher, E.M. *et al.* Genotypic and phenotypic characterization of drug-resistant *Mycobacterium tuberculosis* isolates from rural districts of the Western Cape Province of South Africa. *J. Clin. Microbiol.* **42**, 891–894 (2004).
13. Casali, N. *et al.* Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Res.* **22**, 735–745 (2012).
14. Ioerger, T.R. *et al.* The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. *BMC Genomics* **11**, 670 (2010).
15. Sun, G. *et al.* Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J. Infect. Dis.* 10.1093/infdis/jis601 (2012).
16. European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. *Emerg. Infect. Dis.* **12**, 736–743 (2006).
17. Borrell, S. & Gagneux, S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis*. *Int. J. Tuberc. Lung Dis.* **13**, 1456–1466 (2009).
18. Hershberg, R. *et al.* High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* **6**, e311 (2008).
19. Comas, I. *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* **42**, 498–503 (2010).
20. Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K. & Van Soolingen, D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg. Infect. Dis.* **8**, 843–849 (2002).
21. Kato-Maeda, M. *et al.* Beijing sublineages of *Mycobacterium tuberculosis* differ in pathogenicity in the guinea pig. *Clin. Vaccine Immunol.* **19**, 1227–1237 (2012).
22. Drobniewski, F. *et al.* Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *J. Am. Med. Assoc.* **293**, 2726–2731 (2005).
23. Coscolla, M. & Gagneux, S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov. Today Dis. Mech.* **7**, e43–e59 (2010).
24. de Jong, B.C. *et al.* Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *J. Infect. Dis.* **198**, 1037–1043 (2008).
25. Pang, Y. *et al.* Spoligotyping and drug resistance analysis of *Mycobacterium tuberculosis* strains from national survey in China. *PLoS ONE* **7**, e32976 (2012).
26. Vadwai, V., Shetty, A., Supply, P. & Rodrigues, C. Evaluation of 24-locus MIRU-VNTR in extrapulmonary specimens: study from a tertiary centre in Mumbai. *Tuberculosis* (*Edinb.*) **92**, 264–272 (2012).
27. Huang, H.Y. *et al.* Mixed infection with Beijing and non-Beijing strains and drug resistance pattern of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **48**, 4474–4480 (2010).
28. Anh, D.D. *et al. Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg. Infect. Dis.* **6**, 302–305 (2000).
29. Taype, C.A. *et al.* Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. *Infect. Genet. Evol.* **12**, 577–585 (2012).
30. Mestre, O. *et al.* Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PLoS ONE* **6**, e16020 (2011).
31. de Steenwinkel, J.E.M. *et al.* Drug susceptibility of *Mycobacterium tuberculosis* Beijing genotype and association with MDR TB. *Emerg. Infect. Dis.* **18**, 660–663 (2012).
32. Werngren, J. Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J. Clin. Microbiol.* **41**, 1520–1524 (2003).
33. Rosche, W.A. & Foster, P.L. Determining mutation rates in bacterial populations. *Methods* **20**, 4–17 (2000).
34. Luria, S.E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
35. Lang, G.I. & Murray, A.W. Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* **178**, 67–82 (2008).
36. Stewart, F.M. Fluctuation tests: how reliable are the estimates of mutation rates? *Genetics* **137**, 1139–1146 (1994).
37. Stewart, F.M., Gordon, D.M. & Levin, B.R. Fluctuation analysis: the probability distribution of the number of mutants under different conditions. *Genetics* **124**, 175–185 (1990).
38. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
39. Burnham, K.P. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
40. Bergval, I.L., Schuitema, A.R.J., Klatser, P.R. & Anthony, R.M. Resistant mutants of *Mycobacterium tuberculosis* selected *in vitro* do not reflect the *in vivo* mechanism of isoniazid resistance. *J. Antimicrob. Chemother.* **64**, 515–523 (2009).
41. Sandgren, A. *et al.* Tuberculosis drug resistance mutation database. *PLoS Med.* **6**, e2 (2009).
42. Gardy, J.L. *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* **364**, 730–739 (2011).
43. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
44. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
45. Kimura, M. & Ota, T. On the rate of molecular evolution. *J. Mol. Evol.* **1**, 1–17 (1971).
46. Walker, T.M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* **13**, 137–146 (2013).
47. Colijn, C., Cohen, T., Ganesh, A. & Murray, M.B. Spontaneous emergence of multiple drug resistance in tuberculosis before and during therapy. *PLoS ONE* **6**, e18327 (2011).
48. Portevin, D., Gagneux, S., Comas, I. & Young, D.B. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog.* **7**, e1001307 (2011).
49. Bromham, L. & Penny, D. The modern molecular clock. *Nat. Rev. Genet.* **4**, 216–224 (2003).
50. Rocha, E.P.C. *et al.* Comparisons of d*N*/d*S* are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239**, 226–235 (2006).
51. Hall, L.M.C. & Henderson-Begg, S.K. Hypermutable bacteria isolated from humans—a critical analysis. *Microbiology* **152**, 2505–2514 (2006).
52. Blázquez, J. Hypermutation as a factor contributing to the acquisition of antimicrobial resistance. *Clin. Infect. Dis.* **37**, 1201–1209 (2003).
53. Oliver, A., Cantón, R., Campo, P., Baquero, F. & Blázquez, J. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**, 1251–1254 (2000).
54. Mizrahi, V. & Andersen, S.J. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol. Microbiol.* **29**, 1331–1339 (1998).
55. Springer, B. *et al.* Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol. Microbiol.* **53**, 1601–1609 (2004).
56. Fallow, A., Domenech, P. & Reed, M.B. Strains of the East Asian (W/Beijing) lineage of *Mycobacterium tuberculosis* are DosS/DosT-DosR two-component regulatory system natural mutants. *J. Bacteriol.* **192**, 2228–2238 (2010).
57. Huet, G. *et al.* A lipid profile typifies the Beijing strains of *Mycobacterium tuberculosis*: identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J. Biol. Chem.* **284**, 27101–27113 (2009).
58. Schmalstieg, A.M. *et al.* The antibiotic resistance arrow of time: efflux pump induction is a general first step in the evolution of mycobacterial drug resistance. *Antimicrob. Agents Chemother.* **56**, 4806–4815 (2012).
59. Louw, G.E. *et al.* Rifampicin reduces susceptibility to ofloxacin in rifampicin resistant *Mycobacterium tuberculosis* through efflux. *Am. J. Respir. Crit. Care Med.* **184**, 269–276 (2011).
60. Adams, K.N. *et al.* Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* **145**, 39–53 (2011).
61. Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2012).
62. Williams, K. *et al.* Sterilizing activities of novel combinations lacking first- and second-line drugs in a murine model of tuberculosis. *Antimicrob. Agents Chemother.* **56**, 3114–3120 (2012).
63. Lienhardt, C. *et al.* New drugs for the treatment of tuberculosis: needs, challenges, promise, and prospects for the future. *J. Infect. Dis.* **205**, S241–S249 (2012).
64. Menzies, D. *et al.* Standardized treatment of active tuberculosis in patients with previous treatment and/or with mono-resistance to isoniazid: a systematic review and meta-analysis. *PLoS Med.* **6**, e1000150 (2009).
65. Gelmanova, I.Y. *et al.* Barriers to successful tuberculosis treatment in Tomsk, Russian Federation: non-adherence, default and the acquisition of multidrug resistance. *Bull. World Health Organ.* **85**, 703–711 (2007).
66. Seung, K.J. *et al.* The effect of initial drug resistance on treatment response and acquired drug resistance during standardized short-course chemotherapy for tuberculosis. *Clin. Infect. Dis.* **39**, 1321–1328 (2004).
67. Helb, D. *et al.* Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. *J. Clin. Microbiol.* **48**, 229–237 (2010).
68. Weyer, K. *Laboratory Services in Tuberculosis Control. Part II: Microscopy* (World Health Organization, Geneva, 1998).
69. Dowdy, D.W., Basu, S. & Andrews, J.R. Is passive diagnosis enough? The impact of subclinical disease on diagnostic strategies for tuberculosis. *Am. J. Respir. Crit. Care Med.* **187**, 543–551 (2013).
70. Shaw, J.B. & Wynn-Williams, N. Infectivity of pulmonary tuberculosis in relation to sputum status. *Am. Rev. Tuberc.* **69**, 724–732 (1954).

© 2013 Nature America, Inc. All rights reserved.

ONLINE METHODS

Culture and MIC determination of clinical isolates. Clinical strains were identified as previously described[71]. Strains were grown in broth culture of Middlebrook 7H9 supplemented with 10% Middlebrook oleic acid, albumin, dextrose and catalase solution (OADC), 0.0005% Tween-80 and 0.005% glycerol. To determine MIC, strains were grown to log phase (optical density (OD) of 0.5–0.8) and diluted to an OD of 0.006. Wells were inoculated with 50 µl of culture and then supplemented with 50 µl of medium containing the appropriate concentration of antibiotic. All concentrations were tested in triplicate, and two sets of triplicate control wells were established for each strain. After 6 d, 10 µl of alamar blue (Life Technologies) was added to each well of one triplicate set of controls, where control wells were not supplemented with antibiotic. If control wells changed color from blue to pink after 24 h, 10 µl of alamar blue was added to each triplicate set of experimental wells as well as to a new set of control wells. All wells were examined after 24 h. Any wells that remained blue indicated inhibition of growth. The following concentrations were tested: 0.5, 0.25, 0.125, 0.06125, 0.03063 and 0.01531 µg/ml rifampicin (Sigma, R3501) and 0.25, 0.125, 0.06125, 0.03063, 0.01531 and 0.00766 µg/ml isoniazid (Sigma, I3377). MIC was defined as the lowest concentration of antibiotic that prevented color change from blue to pink.

Fluctuation analysis. Fluctuation analysis was performed as previously described[6]. Briefly, for a single strain of *M. tuberculosis*, starter cultures were inoculated from freezer stocks of culture at OD = 1.0. Once the culture reached an OD of between 0.7 and 1.1, approximately 300,000 cells were used to inoculate 120 ml of 7H9 supplemented with 10% OADC, 0.0005% Tween-80 and 0.005% glycerol, giving a total cell count of 10,000 cells per 4-ml culture. This volume was immediately divided to start 24 cultures of 4 ml each in 30-ml square PETG culture bottles (Nalgene). Cultures were grown at 37 °C with shaking for 11 to 14 d, until reaching an OD of 1.0. Once at an OD of 1.0, 20 cultures were transferred to 15-ml conical tubes and spun at 3,220*g* for 10 min at 4 °C. Cultures were then resuspended in 250–500 µl of 7H9 supplemented with OADC, Tween-80 and glycerol and spotted onto plates with the same medium supplemented with 0.5, 2 or 5 µg/ml rifampicin, 1 µg/ml isoniazid or 5 µg/ml ethambutol (MP Biomedicals, 157949). Once spread using sterile glass beads (4 mm in diameter), plates were allowed to dry and were subsequently incubated at 37 °C for 28 d. Cell counts were determined by serial dilution of four cultures for each strain. The rate at which drug resistance was acquired was determined by calculating m (the estimated number of mutations per culture) based on the number of mutants (r) observed on each plate using the Ma, Sarkar, Sandri (MSS) method as previously described[33,35]. Dividing m by N_t, the number of cells plated for each culture, gives an estimated rate at which drug resistance emerged. We estimated 95% confidence intervals using equations (24) and (25) as described[33,36]. For comparing pairs of fluctuation analysis data (**Fig. 2**), the nonparametric two-sided Wilcoxon rank-sum test (also known as the Mann-Whitney U test) was performed using the ranksum command in Matlab with α set to 0.05, comparing the frequency of drug-resistant mutants in each culture.

Analysis of data from fluctuation analysis. To estimate the extent to which the data met the assumptions of Luria-Delbrück fluctuation analysis, we performed a curve-fitting analysis as described[35]. Briefly, data were fit using either a one-parameter model consistent with the Luria-Delbrück model or a two-parameter model containing an additional parameter describing a Poisson distribution. The fit of each model was assessed using least-squares methodology[35], with AIC_C calculated as described previously[39]. Lower AIC_C reflects better fit given a penalty for increasing the number of parameters, and negative ΔAIC_C (AIC_C (one-parameter model) – AIC_C (two-parameter model)) indicates that the one-parameter model is a better approximation of the data.

Determination of target size. The numbers of *rpoB* mutations conferring resistance to 0.5, 2 and 5 µg/ml rifampicin were determined by transferring 100 colonies (5 from each fluctuation analysis culture) into 100 µl of 7H9 supplemented with 10% OADC, 0.0005% Tween-80 and 0.005% glycerol. Cultures were grown overnight at 37 °C and were then heat inactivated at 85 °C for 2 h. Heat-inactivated cultures were used as templates for PCR and sequencing with previously described primers[72]. Sequences were analyzed for mutation relative to the reference sequence H37Rv, and total numbers of mutations were determined. For each culture, duplicate mutations were only counted once. The absolute number of unique mutations observed across cultures for a given condition was used to determine target size for each strain under each condition.

Estimation of mutation rate from human isolates. To determine the per-base per-day mutation rate in human isolates, phylogenies were created using the concatenated SNP sequences reported from a clonal outbreak of a lineage 4 strain in British Columbia, Canada[42], employing BEAST v.1.7.2 (refs. 43,44) to perform Bayesian MCMC analysis. Before phylogenetic analysis, SNPs located in repeat regions (PE_PGRSs, PPEs and transposable elements) were excluded, consistent with our previous experimentally validated analysis of SNPs from whole-genome sequencing to estimate mutation rate[6]. Concatenated SNP sequences were compiled and prepared using BEAUti v1.7.2 to select analysis parameters and construct the XML input file. Concatenated sequences were converted to NEXUS format and loaded into BEAUti, where time was noted for each isolate. Time was defined in days based on the time elapsed from isolation of the historical isolate MT0005 (in 1995) to symptom onset from the isolate. A generalized time-reversible (GTR) substitution model was used with empirically determined base frequencies. Default priors were used for 10 million chains. Output was analyzed in Tracer v1.5, and all parameters produced an effective sample size of 200 or greater. Phylogenetic tree construction was completed using TreeAnnotator v1.7.2 with a posterior probability limit of 0.5 and a burn-in of 1,000 trees, leaving 9,001 potential trees for construction. Tree visualization was completed using FigTree v1.3.1, and the tree was rooted on MT0005; the most likely tree is depicted in **Supplementary Figure 1**.

Mathematical simulation of drug resistance. We developed a compartmental, partially stochastic mathematical model of the evolution of drug resistance within an individual according to the following set of equations:

$$N_S(t) = (N_S(t-1) \times (b - d_A)) - m_{R\bullet S} - m_{H\bullet S} \quad (1)$$

$$N_R(t) = (N_R(t-1) \times (b \times (1 - cr_R) - d_A)) + m_{R\bullet S} - m_{H\bullet R} \quad (2)$$

$$N_H(t) = (N_H(t-1) \times (b \times (1 - cr_H) - d_A)] + m_{H\bullet S} - m_{R\bullet H} \quad (3)$$

$$N_{MDR}(t) = (N_{MDR}(t-1) \times (b \times (1 - cr_{MDR}) - d_A)) + m_{R\bullet H} + m_{H\bullet R} \quad (4)$$

where

$$m_{R\bullet S} \sim \text{Poisson}(\mu_R \times N_S(t-1)) \quad (5)$$

$$m_{H\bullet S} \sim \text{Poisson}(\mu_H \times N_S(t-1)) \quad (6)$$

$$m_{H\bullet R} \sim \text{Poisson}(\mu_H \times N_R(t-1)) \quad (7)$$

$$m_{R\bullet H} \sim \text{Poisson}(\mu_R \times N_H(t-1)) \quad (8)$$

Here, $N_S(t)$, $N_R(t)$, $N_H(t)$ and $N_{MDR}(t)$ are the population sizes of the susceptible, rifampicin-resistant, isoniazid-resistant and multidrug-resistant populations at time t, respectively. The parameters b and d_A represent the birth and death rates of bacteria in each population, and the parameter cr represents the fitness cost associated with the subscripted resistance, where R represents rifampicin resistance and H represents isoniazid resistance. Birth and death rates represent the rate at which new bacteria are added to the population through growth and the rate at which bacteria are removed from the population through immune killing or senescence, respectively. The subscripted parameters in $m_{X\bullet Y}$ are the number of bacteria transitioning from population N_Y to population N_X. m is a stochastically determined parameter derived from a random Poisson variable, where the Poisson parameter λ is determined by the experimentally estimated rate of acquisition of drug resistance multiplied by the relevant population size and the parameter μ represents the relevant rate at which drug resistance is acquired for each modeled condition. These equations were parameterized with the values shown in **Supplementary Table 6**. All simulations were run in Matlab. For all simulations and for both mutation parameter sets ($\mu_{H\text{-}W}$ and $\mu_{R\text{-}W}$; $\mu_{H\text{-}CDC}$ and $\mu_{R\text{-}CDC}$, where W represents the W-Beijing strain HN878 and CDC represents the strain CDC-1551), simulations of the evolution of drug resistance were run 100,000 times to determine

© 2013 Nature America, Inc. All rights reserved.

the probability of observing drug resistance with a given set of parameters. To determine the effect of varying birth rate, 10 simulations of 200,000 simulated infected individuals (100,000 per simulated strain) were each run with b = (0.20:1.10 in increments of 0.10), giving a net birth rate of 0.05:0.95. To determine the effect of varying the fitness of drug-resistant mutants, 10 simulations of 200,000 infected individuals each (100,000 per simulated strain) were run with $cr_H = cr_R$ = (0.0: 0.90 in increments of 0.10). For all simulations, bacterial burden was allowed to increase to 1×10^{12} bacteria in an infected individual, and the probabilities of observing rifampicin resistance, isoniazid resistance and MDR were determined by dividing the number of simulated individuals with at least one resistant bacterium by the total number of simulated individuals. To determine the probability of observing resistance in a model where mutation is determined by replication dynamics, we used equation (2) from ref. 47 with the parameters listed in **Supplementary Figure 2c**.

71. Tsolaki, A.G. *et al.* Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci. USA* **101**, 4865–4870 (2004).
72. Gagneux, S. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**, 1944–1946 (2006).

© 2013 Nature America, Inc. All rights reserved.