RESEARCH ARTICLE

The molecular clock of *Mycobacterium tuberculosis*

Fabrizio Menardo[1,2]*, **Sebastian Duchêne**[3], **Daniela Brites**[1,2], **Sebastien Gagneux**[1,2]

1 Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland, **2** University of Basel, Basel, Switzerland, **3** Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia

* fabrizio.menardo@swisstph.ch

Abstract

The molecular clock and its phylogenetic applications to genomic data have changed how we study and understand one of the major human pathogens, *Mycobacterium tuberculosis* (MTB), the etiologic agent of tuberculosis. Genome sequences of MTB strains sampled at different times are increasingly used to infer when a particular outbreak begun, when a drug-resistant clone appeared and expanded, or when a strain was introduced into a specific region. Despite the growing importance of the molecular clock in tuberculosis research, there is a lack of consensus as to whether MTB displays a clocklike behavior and about its rate of evolution. Here we performed a systematic study of the molecular clock of MTB on a large genomic data set (6,285 strains), covering different epidemiological settings and most of the known global diversity. We found that sampling times below 15–20 years were often insufficient to calibrate the clock of MTB. For data sets where such calibration was possible, we obtained a clock rate between 1×10^{-8} and 5×10^{-7} nucleotide changes per-site-per-year (0.04–2.2 SNPs per-genome-per-year), with substantial differences between clades. These estimates were not strongly dependent on the time of the calibration points as they changed only marginally when we used epidemiological isolates (sampled in the last 40 years) or three ancient DNA samples (about 1,000 years old) to calibrate the tree. Additionally, the uncertainty and the discrepancies in the results of different methods were sometimes large, highlighting the importance of using different methods, and of considering carefully their assumptions and limitations.

Author summary

One of the major recent advancements in evolutionary biology is the development of statistical methods to infer the past evolutionary history of species and populations with genomic data. In the last five years, many researchers have used the molecular clock (i.e. the observation that genomes accumulate mutations at an approximately constant pace) to study the evolution and epidemiology of *Mycobacterium tuberculosis*, a bacterial pathogen that causes tuberculosis and is responsible for 1.6 million human deaths ever year. Applications of the molecular clock are used to understand when tuberculosis emerged as

a pathogen, the evolution of drug resistance, how different strains transmit and spread across the world and how MTB populations are affected by control programs. Here, we performed a systematic analysis of the molecular clock of MTB, analyzing several whole genome sequence data sets with the same set of methodologies. We characterized the rate of molecular evolution (the pace of the clock), and its variation between different MTB populations and lineages. Our results provide an important guideline for future analyses of tuberculosis and other organisms.

Introduction

In 1962, Zuckerland and Pauling used the number of amino-acid differences among hemoglobin sequences to infer the divergence time between human and gorilla, in what was the first application of the molecular clock [1]. Although many at the time found it "crazy" [2], soon the molecular clock was incorporated in Kimura's neutral theory of molecular evolution [3], and found its place in the foundations of evolutionary biology. Thanks to the improvements of sequencing technologies and statistical techniques, it is now possible to use sequences sampled at different times to calibrate the molecular clock and study the temporal dimension of evolutionary processes in so called measurably evolving populations [4]. These advancements have been most relevant for ancient DNA (aDNA), and to study the evolutionary dynamics of pathogen populations, including one of the deadliest human pathogens: *Mycobacterium tuberculosis*.

In 1994, Kapur and colleagues pioneered molecular clock analyses in MTB: they assumed a clock rate derived from other bacteria and used genetic polymorphisms to infer the age of divergence of different MTB strains [5]. Since the publication of the MTB reference genome [6], whole genome sequence (WGS) data of MTB strains is becoming available at increasing speed, and especially in the last five years, studies using large WGS data sets allowed for precise estimates of the MTB genetic diversity and of the molecular clock rate. Phylogenetic analyses with a molecular clock have been used to estimate the timing of the introduction of MTB clades to particular geographic regions, the divergence time of the MTB lineages, and the age of the most recent common ancestor (MRCA) of the MTB complex [7–13]. Clock models, together with phylodynamic models in a Bayesian setting have been used to characterize tuberculosis epidemics by determining the time at which outbreaks began and ended [14–18], establishing the time of origin and spread of drug resistant clades [11, 14, 19–20], and correlating population dynamics with historical events [9, 12, 20, 21–22]. One example of the potential of molecular clock analyses is the study of Eldholm and colleagues [20], where the collapse of the Soviet Union and of its health system was linked to the increased emergence of drug-resistant strains in former Soviet Republics, thus providing insights into the evolutionary processes promoting drug resistance.

A key aspect about estimating evolutionary rates and timescales in microbial pathogens is assessing their clocklike structure. All molecular clock analyses require some form of calibration. In many organisms, this consists in constraining internal nodes of phylogenetic trees to known divergence times (for example, assuming co-divergence with the host, or the fossil record), but in rapidly evolving pathogens and studies involving ancient (a)DNA, it is also possible to use sampling times for calibrations [23]. In the latter approach, the ages of the tips of a tree, rather than those of internal nodes are constrained to their collection times. Clearly, the sampling time should capture a sufficient number of nucleotide changes to estimate the evolutionary rate, which will depend on the evolutionary rate of the organism and the extent of rate

variation among lineages. Some popular methods to assess such clocklike structure are the root-to-tip regression and the date randomization test (DRT).

While many of the studies inferring evolutionary rates for MTB reported support for a molecular clock [10–11, 13–14, 16, 18, 20, 22], some found a lack of clocklike structure [7, 17–18], and others assumed a molecular clock without testing whether the data had a temporal structure [9, 12, 15, 19, 21]. In all studies where the calibration was based on the sampling time (tip-dating), the clock rate estimates spanned roughly an order of magnitude around 10^{-7} nucleotide changes per site per year. This was in contrast with the results of Comas et al. 2013 [7], where the clock was calibrated assuming co-divergence between MTB lineages and human mitochondrial haplotypes (i.e. internal node calibrations), and was estimated to be around 10^{-9} nucleotide changes per site years. Additionally, the age estimate of the MRCA of MTB obtained by Comas et al. 2013 (~70,000 years) is similar to the one of Wirth et al. 2008 [24] (~ 40,000 years), which was based on the molecular clock analysis of 24 variable tandem repeat loci, but about ten times older than the one of study that used WGS of aDNA samples and tip-dating to calibrate the tree (~ 6,000 years) [8]. The diverging results obtained by Comas et al. 2013 [7], Wirth et al. 2008 [8] and studies based on tip-dating are therefore most likely caused by the adoption of different assumptions and methodologies.

Studies based on WGS and tip-dating found some differences in the clock rate of different lineages: some lineage 2 (L2) data sets [20] were found to have a faster clock rate compared to lineage 4 (L4) data sets [11, 14, 16, 21], while others showed lower clock rates, comparable with L4 [13, 18]. Studies based on aDNA produced slightly lower clock rate estimates [8, 10] compared to studies based on modern strains, thus suggesting support for the phenomenon of time dependency of clock rates in MTB [25–26]. All these results indicate that different MTB lineages and populations might have different clock rates, and that the age of the calibration points could influence the results of the analyses. Comparing the results of different studies has however a main limitation: the observed differences could be due to different rates of molecular evolution among MTB populations, to methodological discrepancies among studies, or a combination of both.

Here we assembled a large genomic data set of sequences belonging to epidemiological contemporary strains (sampled in the last 40 years), including sequences from all major lineages of MTB (6,285 strains in total, belonging to six human adapted lineages, L1-L6, and one lineage predominantly infecting cattle, *M. bovis*). We then applied the same set of methodologies to the whole data set, to individual lineages and sub-lineages, and to selected local outbreaks, thus disentangling the temporal signal of multiple nested data sets with different structures and complexity, and ensuring the comparability of the results among different bacterial clades and epidemiological settings. Finally, in a separate analysis we investigated the time-dependency hypothesis by including three aDNA sequences about 1,000 years old [8].

With this systematic approach, we addressed the following questions:

1. Is there a molecular clock in MTB and how do we detect it?

2. What is the clock rate of MTB, and what is its variation among lineages, sub-lineages and individual outbreaks?

3. Are clock rate estimates dependent on the age of the calibration points in MTB?

Results and discussion

Is there a molecular clock in MTB?

Finding evidence of temporal structure is the first step when performing molecular clock analyses [27]. If there is not enough genetic variation between samples collected at different times,

these cannot be used to calibrate the molecular clock, i.e. the population is not measurably evolving. To test the temporal structure of MTB data sets, we identified 6,285 contemporary strains (sampled in the last 40 years) that passed our quality filters (average coverage > 15X, they were not identified as mixed infection etc. see Methods for details), and for which the date of isolation was known (S1 Table).

We used root-to-tip regression to evaluate the temporal structure of the whole MTB complex and of the individual lineages (L1-L6 and *M. bovis*) [28]. The root-to-tip regression is a regression of the root-to-tip distances as a function of sampling times of phylogenetic trees with branch lengths in units of nucleotide changes per site, where the slope corresponds to the rate. Under a perfect clock-like behavior, the distance between the root of the phylogenetic tree and the tips is a linear function of the tip's sampling year: recently sampled strains are further away from the root than older ones, such that the R^2 is the degree of clocklike behavior [29]. We obtained very low values of R^2 for all lineages (maximum 0.1 for *M. bovis*), indicating a lack of strong clock-like behavior (S1 Fig). Additionally, we found a weak negative slope for L1, L5 and L6, normally interpreted as evidence for a lack of temporal structure, or overdispersion in the lineage-specific clock rates [28] (S1 Fig, S2 Table). A negative slope of the regression line can be caused by an incorrect placement of the root [30]. To address this potential problem, we repeated these analyses rooting the trees with an outgroup. We found a negative slope for L1 and L6 and a positive slope for L5, although with an extremely low value of R^2 (< 0.01). These results indicate that the negative slope of L1 and L6 and the low R^2 values of the three data sets are not due to an incorrect placement of the root (S2 Fig).

Since root-to-tip regression can be used only for exploratory analyses and not for formal hypothesis testing [28], we performed a date randomization test (DRT). The DRT consists in repeatedly reshuffling the date of sampling among taxa and then comparing the clock rate estimates among the observed and reshuffled data sets [27]. If the estimate obtained from the observed data does not overlap with the estimates obtained from the randomized data sets, we can conclude that the observed data has a stronger temporal signal than expected by chance, such that there is statistically significant clocklike structure [27]. Usually the DRT is implemented in a Bayesian phylogenetic setting, however, considering the size and the number of data sets included in this study, an excessive amount of computation would be required. To overcome this problem, we estimated the clock rate with the least-squared dating method (LSD) [31]. The advantage of this method is that it is orders of magnitude faster than fully Bayesian approaches, and can therefore be used on data sets with thousands of taxa and with more randomizations compared to the 10–20 typically used in a Bayesian setting [32]. A limitation of least squares dating is that it typically assumes a single tree topology and vector of branch lengths, and a strict clock (i.e. all branches have the same clock rate). However, a simulation study showed that maximum likelihood trees produced similar estimates compared to the true topology, and that it is robust to uncorrelated variation of the clock rate among branches in the phylogeny [31–33].

For each data set (the MTB complex and the seven analyzed lineages), we reshuffled the year of sampling among tips 100 times and estimated the clock rate of observed and randomized data sets with LSD. All eight data sets except L5 and L6 passed the DRT (Methods, S1 Fig, S2 Table). L5 and L6 are the two lineages with the lowest sample size, 117 and 33 strains, respectively. Moreover most strains were sampled in a short temporal period compared to the other lineages (S3–S8 Figs). It is likely that with additional strains sampled across a larger time period, L5 and L6 will also show evidence for a molecular clock.

We complemented the analysis described above with a Bayesian phylogenetic analysis in BEAST2 [34]. Since this is computationally expensive, we reduced the large data sets (MTBC, L1, L2, L4 and *M. bovis*) to 300 randomly selected strains. For each data set, we selected the

best fitting nucleotide substitution model identified with jModelTest 2 [35]. For this first analysis, we assumed a coalescent constant population size prior, used a relaxed clock model, and a $1/x$ prior for the clock rate, constrained between 10^{-10} and 10^{-5} nucleotide changes per site per year. This interval spans the range of clock rates proposed for *M. tuberculosis* and for most other bacteria [20, 36]. We observed that for all data sets the posterior was much more precise (with a narrow distribution) than the prior, thus indicating that the data was informative [37]. Again, the only exceptions were L5 and L6, where the posterior distribution was flat, ranging between 10^{-10} and 10^{-7} nucleotide changes per site per year, confirming the lack temporal structure of these two data sets (S1 Fig).

We repeated these analyses on 23 sub-lineages and 7 outbreaks and local populations to test whether we could detect a temporal structure also in smaller, less diverse data sets. With this sub-sampling scheme, we could compare the results among different clades, among outbreaks with different epidemiological characteristics, and among local outbreaks and global data sets (see Methods).

We found that 11 sub-lineages and 5 local populations passed the DRT (S2 Table, S3–S6 and S9–S11 Figs). All the data sets that failed the DRT had less than 350 genomes, or were composed of strains sampled in a temporal range of 20 years or less. Additionally, only two of the ten data sets sampled across less than 15 years, and three of the twelve data sets with less than 100 strains passed the DRT (Fig 1; S2 Table), indicating that large sample sizes and wide temporal sampling windows are necessary to obtain reliable estimates of evolutionary rates and timescales in MTB. Conversely, the number of polymorphic positions and the genetic diversity measured with Watterson's estimator did not correlate with the outcome of the DRT (S12 Fig).

Among the three methods generally used to study the temporal structure of a data set, the root-to-tip regression resulted in a negative slope, and therefore failed to detect the temporal structure of some of the data sets that passed the DRT (i.e. L1, L4.1.2 and L1.1.1). Nevertheless, root-to-tip regression can be useful to identify data sets where the temporal signal comes from a single strain, or a few strains (see below). Comparing prior and posterior distributions of the clock rates was also useful to detect the presence of temporal structure, although this was not always in agreement with the results of the DRT: some of the data sets that did not pass the DRT (e.g. L2.2.1_nc2, Trewby 2016 [38]) had a posterior distribution of the clock rate more distinct from the prior than some of the data sets that passed the DRT (e.g. L1.1.1, L1.2.1 and L1.2.2) (S3, S5 and S6 Figs, S2 Table). A possible reason for this could be that LSD and BEAST have different statistical power with different data sets. Additionally, in some cases the deviation of the posterior distribution of the clock rate from the prior could be an artifact caused by tree prior misspecification, and not the result of genuine temporal structure [39].

Sensitivity of the clock rate estimates to the model assumptions

In Bayesian analyses, different models and priors are based on different assumptions about the evolutionary processes, and can thus influence the results [40]. Often different sets of assumptions are tested with the Bayes factor (the ratio of the marginal probabilities of two competing models), and the most likely model is then chosen to estimate the parameters of interest [40]. Given the size and number of the data sets considered in this study, it was not possible to assess the relative fit of many competing models for all data sets. However, model misspecification can result in biased estimates. It was therefore important to investigate the robustness of the results to different models and priors.

We repeated the Bayesian analysis using a uniform prior instead of the $1/x$ prior on the clock rate. We ran a BEAST analysis sampling from the priors and found that the uniform
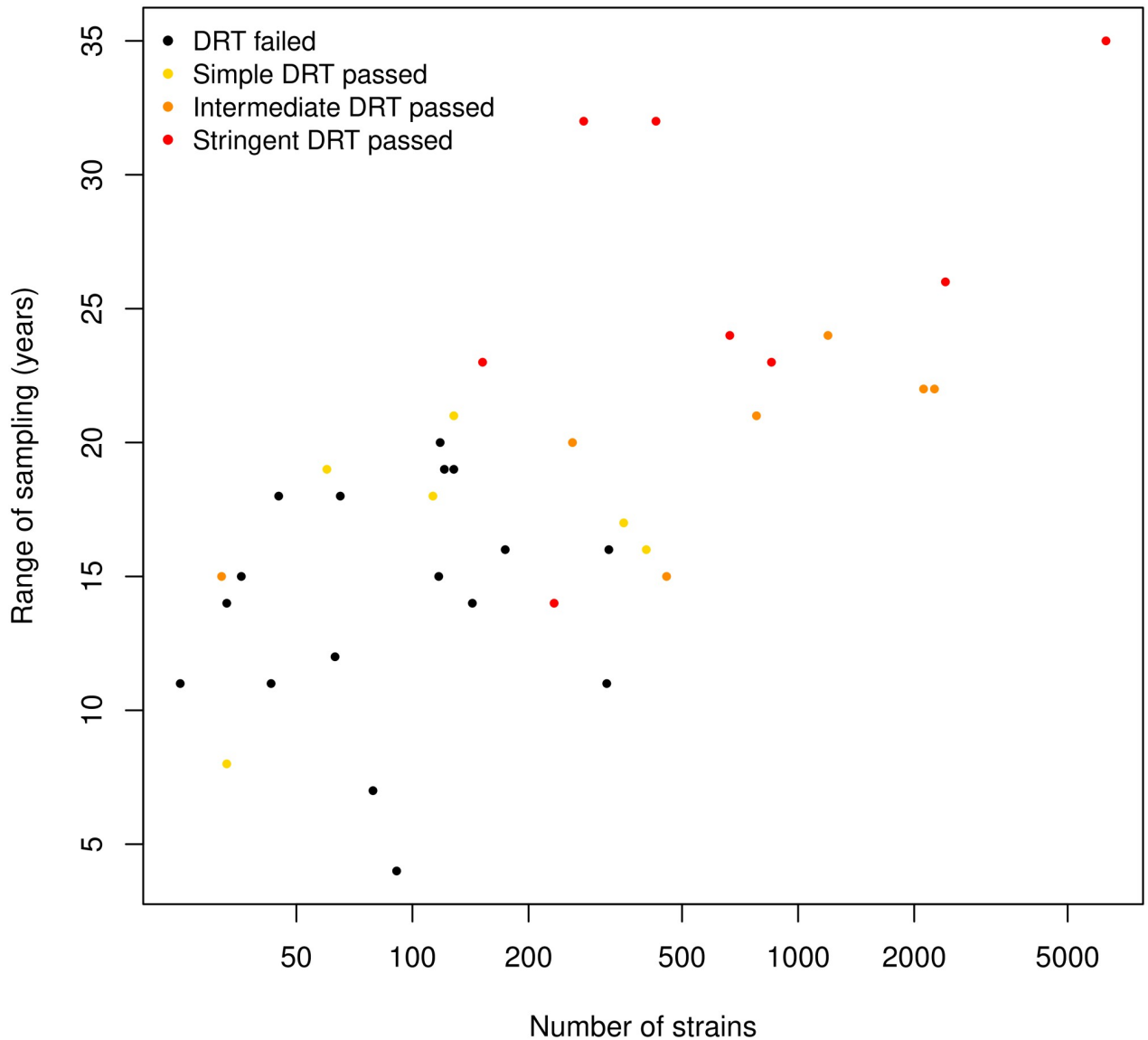
Fig 1. Results of the DRT for all data sets ordered by size and temporal range. Data sets with fewer strains sampled in a shorter period of time tended to fail the DRT.

prior was biased towards high clock rates and put most weight on rates between 10^{-6} and 10^{-5} nucleotide changes per site per year (S13 Fig). For all data sets, we compared the posterior distribution of the clock rate obtained with the two different priors (S14–S16 Figs, S2 Table).

Some data sets showed hardly any difference (e.g. MTBC, L1, L2, L3, L4 etc.), indicating that the data was informative and that the data set had a strong temporal structure. However, this did not always correlate with the results of the DRT. For example, the subset of 300 strains of L2 and the data set Trewby 2016 [38] did not pass the DRT but showed a distinct posterior distribution that was not sensitive to the prior choice. Other data sets, including three that passed the DRT by a small margin (L1.1.1, L1.2.1 and L1.2.2), were more sensitive to the prior choice and resulted in two distinct posterior distributions, indicating a weaker temporal structure (S6 Fig).

An additional assumption of the phylogenetic model that can influence the results of molecular clock analyses is the tree prior (also known as demographic model). We tested the sensitivity to the tree prior by repeating the analysis with an exponential population growth (or shrinkage) prior instead of the constant population size. For this analysis, we used the 1/x prior on the clock rate and we considered only the data sets that passed the DRT (21 data sets). The constant population model is a specific case of the exponential growth model (when the growth rate is equal to zero). Therefore, if the 95% Highest Posterior Density interval (HPD) of the growth rate does not include zero, we can conclude that the data reject a demographic model with constant population size. We found that 14 data sets rejected the constant population size model, and that all of them had positive growth rates (S2 Table). The three data sets that were found to be sensitive to the prior on the clock rate were also sensitive to the tree prior, confirming their low temporal structure and information content, while the results for all other data sets were only moderately influenced by the tree prior (S17 and S18 Figs, S2 Table).

Overall, we found that, except for three data sets (L1.1.1, L1.2.1 and L1.2.2), the clock rate estimates were robust to different priors of the clock rate and to different demographic models. To compare the clock rates of different data sets, we report the analysis with the 1/x prior on the clock rate because the uniform prior can bias the estimates upward. For data sets that showed evidence against the constant population size model (95% HPD of the growth rate not including zero), we report the results of the analysis with the exponential population growth, and for the others, we report the results of the analysis with constant population size.

What is the clock rate of MTB, and what is its variation among lineages, sub-lineages and outbreaks?

We found that the point estimates of all data sets where we detected temporal structure ranged between 2.86×10^{-8} (L3, BEAST analysis) and 4.82×10^{-7} (Eldholm 2016 [20], BEAST analysis) nucleotide changes per site per year. While some data sets had a low range of the 95% confidence interval (CI), reaching the hard limit imposed by LSD of 10^{-10}, most of the CI and 95% highest posterior density intervals (HPD) were included between 10^{-8} and 5×10^{-7} (Fig 2, S2 Table). This range encompasses previous estimates obtained with epidemiological samples and aDNA, and is among the lowest in bacteria, thus supporting our conclusion from above: tip-dating with MTB requires samples collected over a long period of time because of the slow clock rate.

There was one notable exceptions to the pattern described above: the data sets L4_nc which showed a much higher clock rate estimate compared to all other data sets included in this study ($\sim 10^{-6}$; S2 Table). However, this is most likely an artifact: 1) L4_nc is the smallest among all considered data sets, with 32 strains. 2) Most strains are identical or nearly so, collected in the same year, and form a monophyletic clade (S7 and S19 Figs). It is known that data sets with a high degree of temporal and phylogenetic clustering can pass the DRT also when they do not have temporal structure [41]. 3) The root-to-tip regression suggests that the temporal signal comes from one single strain in L4_nc (S5 Fig). We therefore excluded the L4_nc data set from further analyses.

Our results suggest that different lineages of MTB have different clock rates. For example most L1 data sets had point estimates higher than most L4 data sets, although the CI and HPD were often overlapping. The point estimates indicate that the clock rate of L1 is more than double the clock rate of L4: two average L1 strains are expected to differ by 12 SNPs after ten years of divergence, while two average L4 strains will differ by 5 SNPs after the same period of time. This was supported by the results of both LSD, where the 95% CI of L1 and L4 did not overlap,
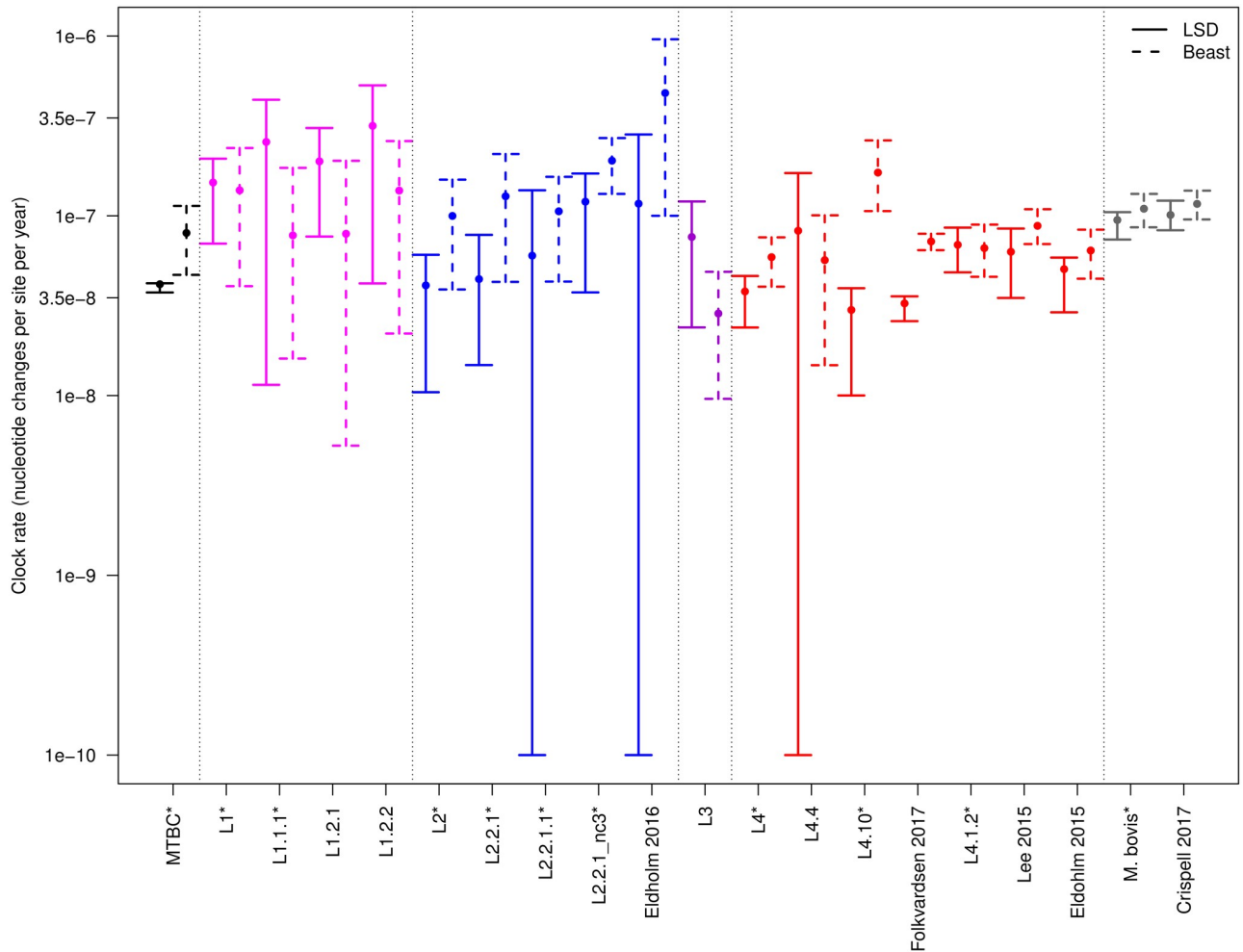
Fig 2. Estimated clock rates of all lineages, sub-lineages and local data sets that passed the DRT. Solid lines represent the 95% confidence interval estimated with LSD, dashed lines represent the 95% highest posterior density (HPD) estimated by BEAST (the larger dot is the median of the posterior distribution). We show here the results of the BEAST analysis with the 1/x prior on the clock rate. For data sets that rejected the constant population size, we show the result obtained with the exponential population growth prior, for the other data sets we show the results obtained with the constant population size prior. Data sets marked with * have been reduced to 300 randomly picked strains for the BEAST analysis.

and BEAST, where the 95% HPD overlapped partially, but the two posterior distributions showed distinct peaks (Fig 2, S2 Table, S20 Fig). A practical implication of these results pertains to the widespread use of SNP distances to identify ongoing transmission in MTB epidemiological studies. Usually, recent transmission is postulated when two or more strains differ by a number of SNPs below a certain threshold [42]. However, this approach will result in systematically lower levels of transmission for clades with faster rates of molecular evolution. For example, a recent study reported low transmission rates of L1 compared to L2 and L4 in Vietnam [43], which could partially be explained by a faster clock rate of L1, as opposed to reduced ongoing transmission.

When considering the results of BEAST, L2 had a higher clock rate compared to L4, and all data sets included in the sub-lineage L2.2.1 showed a faster clock rate compared to the complete L2 data set (Fig 2). The sub-lineage L2.2.1 includes the so called "modern Beijing" family, which was shown to be epidemiologically associated with increased transmission, virulence and drug [43–48], and to have a higher mutation rate compared to L4 strains [49]. However,

the LSD estimates for L2.2.1 and for its sub-lineages, despite showing the same trend of BEAST, support a lower clock rate compared to BEAST, and have large confidence intervals, overlapping with the results of L2 and L4 (Fig 2).

Further evidence of among-lineage variation is provided by the results of the Bayesian analyses, where for most data sets, we obtained coefficients of variation (COV) with a median of 0.2–0.3, and not abutting zero (S2 Table), thus rejecting the strict clock [37].

Taken together, these results indicate that there is a moderate variability among the current rate of molecular evolution of different MTB lineages, which could be caused by different mutation rates as it was reported for L2 and L4 [49], and support the idea that the inference of transmission in MTB should move away from the use of SNP distances to methods that incorporate information about the molecular clock [50].

In our analysis, we included two outbreaks caused by strains belonging to the same sub-lineage (L4.1.2; Eldholm 2015 [14], Lee 2015 [15]). This gave us the opportunity to compare the molecular clock of clades with a similar genetic background in different epidemiological settings. The Eldholm 2015 data set is a sample of an outbreak in Argentina, in which resistance to multiple antibiotics evolved several times independently [14]. The Lee 2015 data set represents an outbreak of drug-susceptible strains in Inuit villages in Québec (Canada) [15]. The clock rates of these two data sets were highly similar (95% CI and HPD ranging between 5.07×10^{-8} and 8.88×10^{-8} for all analyses; Fig 2, S2 Table), thus suggesting that for these two data-sets, different epidemiological characteristics, including the evolution of antibiotic resistance, did not have a large impact on the rate of molecular evolution of MTB.

Overall, our results are comparable with previously published tip-dating studies of MTB that used WGS data, indicating that in practice, using different sequencing and data analysis pipelines is unlikely to lead to drastically different results. However, further studies are needed to reveal in detail the effect that some steps of the bioinformatic pipeline, such as the exclusion of low confidence SNP calls, repetitive genomic regions, or mixed infections might have on the results of molecular clock analyses.

A faster clock for the ancestor of *M. bovis*?

We showed that in the last 40 years, the clock rates of different MTB data sets were moderately divergent. A different question is whether the clock rate was constant during the evolutionary history of the MTB complex. When looking at the phylogenetic tree of the MTB complex, rooted with the genome sequence of *M. canettii*, one notices that strains belonging to different lineages, despite being all sampled in the last 40 years, have different distances from the root (Fig 3). For example, since their divergence from the MRCA of the MTB complex, the two *M. africanum* lineages (L5 and L6) and especially *M. bovis*, accumulated more nucleotide changes than the lineages belonging to MTB *sensu stricto* (L1-L4; Fig 3). Additionally, all methods (root-to-tip regression, LSD and BEAST) if used without an outgroup, placed the root on the branch between *M. bovis* and all other lineages, while rooting the tree with the outgroup *M. canettii* placed the root on the branch connecting MTB *sensu stricto* with *M. africanum* (L5 and L6) and *M.bovis*. The different root placement affects the clock rate estimation only moderately (4.16×10^{-8} LSD analysis without outgroup, 5.59×10^{-8} LSD analysis with outgroup; S2 Table), but it is a further indication of the variation of the rate of molecular evolution during the evolutionary history of the MTB complex. The observation that all *M. bovis* strains, despite having a clock rate similar to all other data sets, have a larger distance from the root of the MTB complex tree compared to other lineages is intriguing, and could be explained by a faster rate of molecular evolution of the ancestors of *M. bovis* (Figs 2 and 3). It is believed that *M. bovis* switched host (from human to cattle) [51–53], and it is possible that during the
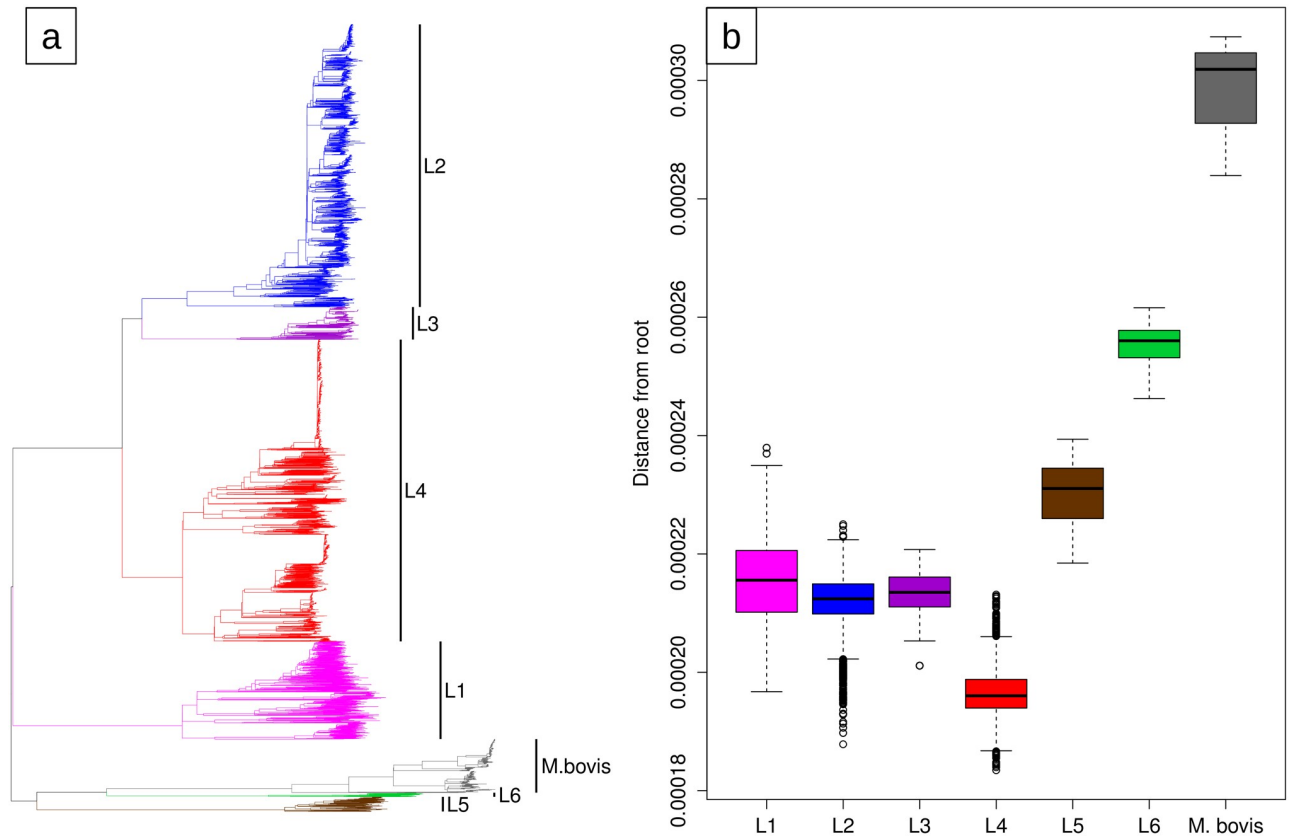
Fig 3. a) The Maximum Likelihood tree of the 6,285 strains considered in this study rooted with the genome sequence of *M. canetti*. **b)** Phylogenetic distance from the root (expected nucleotide changes per site) of MTB strains by lineage.

adaptation to the new host, several genes were under positive selection, thus leading to an increase in the accumulation of substitutions in the *M. bovis* genome. Another possibility is that the ancestor of *M. bovis* experienced a period of reduced population size, a bottleneck, and as a consequence, slightly deleterious mutations were fixed by genetic drift, resulting in a faster clock rate compared to larger populations where selection is more efficient in purging deleterious mutations [54–55].

Time dependency of the clock rate

It has been suggested that in MTB, as in other organisms, the clock rate estimation is dependent on the age of the calibration points [7, 25–26, 36, 56], and that using recent population-based samples could result in an overestimation of the clock rate, because these samples include deleterious mutations that have not yet been purged by purifying selection. However, the validity of the time dependency hypothesis has been contested in general [57], and for MTB in particular [21]. Here we used an approach similar to Rieux et al. 2014 [58] and tested whether the time dependency hypothesis was supported by our data. We repeated the analyses presented above, only this time we included the aDNA genome sequences of three MTB strains obtained from Precolumbian human remains from Peru [8]. If the clock rate estimates depend on the age of the calibration points, adding ancient genomes should result in lower clock rates. We performed this analysis with LSD, using the complete data set (6,285 strains), and with

Table 1. Results of LSD and BEAST for the MTB complex with and without aDNA samples. With BEAST we performed three different analyses, two using a sub-sample of 300 strains and different priors on the clock rate (1/x and uniform), and one using a sub-sample of 500 strains (Methods, S2 Table).

Data set	Clock rate
MTBC 6,285 + *M. canettii* (LSD)	5.59E-08 (4.12E-08, 6.17E-08)
MTBC 6,285 + *M. canettii* + aDNA (LSD)	6.93E-08 (5.48E-08, 8.42E-08)
MTBC 300 (BEAST 1/x)	8.2254E-08 (4.964E-08, 1.141E-07)
MTBC 300 +aDNA (BEAST 1/x)	7.3978E-08 (4.648E-08, 1.019E-07)
MTBC 300 (BEAST unif)	8.26E-08 (4.82E-08, 1.14E-07)
MTBC 300 +aDNA (BEAST unif)	7.1794E-08 (4.157E-08, 9.851E-08)
MTBC 500 (BEAST unif)	6.08E-08 (4.21E-08, 8.07E-08)
MTBC 500 +aDNA (BEAST unif)	5.20E-08 (3.41E-08, 7.12E-08)

https://doi.org/10.1371/journal.ppat.1008067.t001

BEAST, using the sub-sample of 300 randomly selected strains described above, and an additional independent random sub-sample of 500 strains (Methods).

With LSD, adding the aDNA samples resulted in a slightly faster clock rate, conversely all the analyses performed with BEAST resulted in marginally slower clock rates when the aDNA samples were included (Table 1). These results indicate that the effect of the age of the calibration points on the clock rate is modest, and they are corroborated by the observation that MTB mutation rates in vitro and in vivo, estimated with fluctuation assays and resequencing of strains infecting macaques, are remarkably similar to the clock rates obtained in our study (~ $3x10^{-8}$–$4x10^{-7}$) [59].

The aDNA samples considered in this study are not optimal to test the time dependency hypothesis because they belong to the *M. pinnipedii* clade of the MTB complex [8]. The modern strains of this lineage are rarely sampled, because they are infecting seals and sea lions rather than humans. The only additional aDNA samples available for MTB are L4 samples isolated from 18th century Hungarian mummies [10, 60]. However, these samples are a mix of strains with different genotypes, and cannot be easily integrated with the data and pipelines used in this study. While these results suggest that the age of the calibration points has a small effect on the clock rate estimates, they are based on only three aDNA samples. Additional aDNA samples from older periods and belonging to other lineages are necessary to reject (or confirm) the time dependency hypothesis in MTB. Recently, Sabin and colleagues [61] reported the sequencing of a high quality MTB genome from the 17th century, this data will contribute to the investigation of the time dependency hypothesis in MTB.

Dating MTB phylogenies

In most cases, the goal of molecular clock studies is not to estimate the clock rates, but rather the age of the phylogenetic tree and of its nodes. Conceptually, this means extrapolating the age of past events from the temporal information contained in the sample set. If we exclude the few aDNA samples that are available [8, 10], all MTB data sets have been sampled in the last 40 years. It is therefore evident that the age estimates of recent shallow nodes will be more accurate than medium and deep nodes. In part, this is reflected in the larger CI and HPD of the age of ancient nodes compared to more recent ones. Extrapolating the age of trees that are thousands of years old with contemporary samples is particularly challenging, because the observed data captures only a small fraction of the sample's evolutionary history, and these are the cases where aDNA samples are most valuable.

Nevertheless, the age of the MRCA of the MTB complex and of its lineages is highly relevant to understand the emergence and evolution of this pathogen and a debated topic [7–8, 24].

The LSD analyses on the tree rooted with *M. canettii* estimated the MRCA of the MTB complex to be between 2,828 and 5,758 years old (S2 Table). These results are highly similar to the ones of Bos and colleagues (2,951–5,339) which were obtained with Bayesian phylogenetics and a much smaller sample size [8]. These estimates should be taken with caution because of the intrinsic uncertainty in estimating the age of a tree that is several thousands of years old, calibrating the molecular clock with the sampling time of modern strains and only three aDNA samples. A more approachable question is the age of the MRCA of the individual MTB lineages. Here we can consider the results of four different analyses: the LSD and BEAST analyses on the individual lineages (L1-L4, and *M. bovis*), and the LSD and BEAST analyses on the complete MTB complex (including the aDNA samples), from which the age of the MRCA of the lineages can be extracted (L1-L6, and *M. bovis*). When we combined all these results, merging the CI and HPD, we obtained an estimate of the age of the MTB lineages which accounts for the uncertainty intrinsic in each analysis, but also for the differences among inference methods and models, thus providing a more conservative hypothesis. In all our analyses, the point estimates of the age of all lineages resulted to be at most 2,500 years old, and the combined CI and HPD extended to a maximum of 11,000 years ago for L2 (95% CI of the LSD analysis; Fig 4, S3 Table). The large CI of L2 obtained with LSD was maybe due to among-lineage-age variation of the clock rate in L2. While L5, L6 and *M. bovis* have younger MRCAs and narrower confidence intervals, we should note that for these lineages the sampling is much less complete compared to L1-L4, and it is possible that further sampling will add more basal strains to the tree, thus resulting in older MRCAs. For the other lineages, where the sampling is more representative of the global diversity, the confidence intervals of the age of the MRCAs extend over several thousands of years, and the point estimates of the four analyses spread over 1,000–2,000 years. This shows that we should be very careful when interpreting the results of tip-dating in MTB, especially if our goal is to estimate the age of ancient nodes such as the MRCAs of MTB lineages. Conservative researchers might want to use different methods; several model and prior combinations should be formally tested in BEAST, and the final results can be combined in one range providing an estimation of the uncertainty of the clock rate and of the age of some specific node of the tree.

Altogether our results highlight the uncertainty of calibrating MTB trees with tip-dating, they nevertheless support the results of Bos et al. 2014 [8] that found the MRCA of the MTB complex to be relatively recent, and not compatible with the Out-of-Africa hypothesis [7, 24] in which the MTB lineage differentiated in concomitance with the dispersal of *Homo sapiens* out of Africa, about 70,000 years ago. Dating analyses based on DNA samples can only reconstruct the evolutionary history of the data set as far back as the MRCA of the sample. It is possible that in the future new lineages will be sampled, and the MTB phylogeny will be updated moving the MRCA further in the past. Additionally, it is also possible that extinct lineages were circulating and causing diseases much earlier that the MRCA of the strains that are circulating now. This hypothesis is supported by the detection of molecular markers specific for MTB in archeological samples (reviewed in [62]), the oldest of them in a bison's bone about 17,500 years old [63]. Several such studies directly challenge the results of tip-dating presented here because they reported molecular markers specific to MTB lineages in archeological samples that predate the appearance of those lineages as estimated by tip-dating [64–66]. However, there is a controversy regarding the specificity of some of the used markers, and the potential contamination of some of the samples by environmental mycobacteria [67–68].

Whole genome sequences from additional aDNA samples are needed to reconcile these two diverging lines of evidence. Ideally, these samples should represent different lineages, span different periods, and be more ancient than the currently available aDNA from Peruvian human remains.
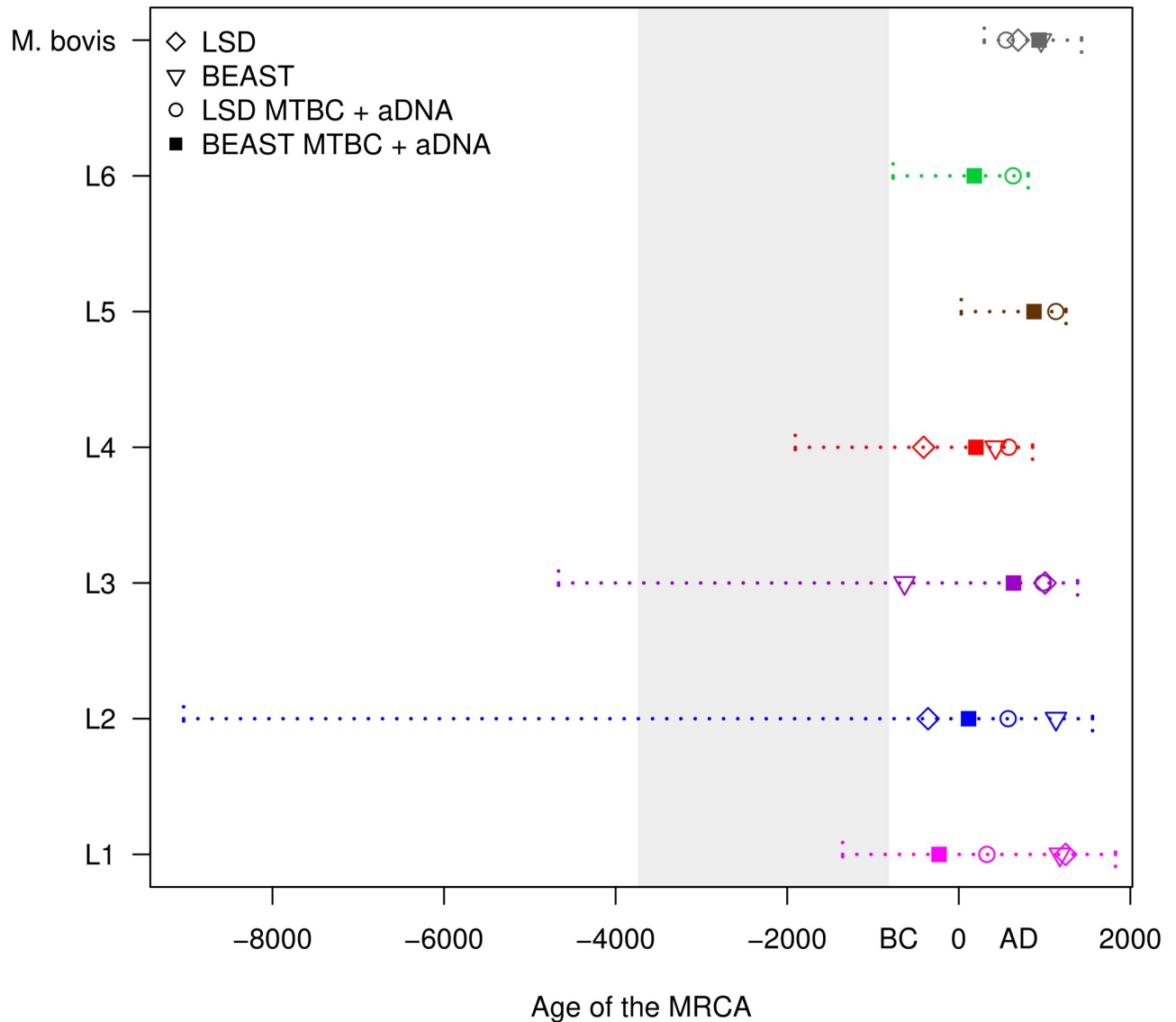
Fig 4. The inferred age of the MTB lineages. LSD: results of the LSD analysis performed on the individual lineages. BEAST: results of the BEAST analysis performed on the individual lineages (median values). LSD MTBC + aDNA: results of the LSD analysis performed on the complete data set of 6,285 strains + 3 aDNA samples, the age of the MRCA of the individual lineages was identified on the calibrated tree. BEAST MTBC + aDNA: results of the BEAST analysis performed on the random subsample of 500 strains + 3 aDNA samples, the age of the MRCA of the individual lineages was identified on the calibrated tree (median values). The confidence intervals were obtained merging the 95% CI and HPD of all analyses. The shaded area represents the age of the MRCA of the MTB complex obtained with the LSD analyses (with and without aDNA, the two 95% CI were merged). For L5 and L6 we report only the age inferred on the complete MTB complex tree, because when analyzed individually these two data sets showed a lack of temporal structure (they failed the DRT).

https://doi.org/10.1371/journal.ppat.1008067.g004

Conclusions

In this systematic study of the molecular clock of MTB, we collected the genome sequences of 6,285 strains from all over the world, divided them into multiple data sets and used different tip-dating methods to assess their temporal structure and molecular clock rates. In most cases, the clock rate could be estimated reliably only if the data sets included strains sampled for 15 or more years. We inferred an overall clock rate ranging between 10^{-8} and 5×10^{-7} nucleotide changes per site per year, independently of the data set. We explored different methodologies

and approaches to molecular clock analyses in MTB, thus providing information and guidance for future studies.

Our results support the robustness of molecular clock analyses in MTB under some conditions, but also highlight the challenges of tip-dating with large genome sizes and slow evolutionary rates. The first challenge is to sample strains for a period long enough to calibrate the clock. When this cannot be done, it is possible to include in the analysis additional strains with older dates that are phylogenetically closely related and for which the genome sequence is available in the public domain. Alternatively, for data with no temporal structure, our estimates can be used to calibrate the clock rate at 10^{-8}–$5x10^{-7}$ nucleotide changes per site per year, thus obtaining a broad (conservative) time estimate for the age of the tree and of its nodes.

The second challenge is computational. As the cost of sequencing decreases, large WGS data sets with several hundred (or thousand) of strains are becoming more common. While large data sets allow for a more precise estimation of the clock rate, they are problematic to analyze with Bayesian phylogenetics. Two possible strategies around this problem are sub-sampling large data sets or reducing the parameter space by using hybrid methods (e.g. using a fixed tree topology) [32].

Some of these challenges might be overcome through future developments. For example, the expansion of fast tip-dating algorithms (such as LSD) to incorporate relaxed clock models would benefit the analysis of large data sets. Additionally, the use of long-read sequencing data will enable the inclusion of SNPs located in repetitive regions of the genome, thus maybe shortening the sampling period needed to calibrate the molecular clock.

Methods

Bioinformatic pipeline

We identified 21,734 MTB genome sequences from the sequence read archive (S4 Table). All genome sequences were processed similarly to what was described in Menardo et al. 2018 [69].

We removed Illumina adaptors and trimmed low quality reads with Trimmomatic v 0.33 (SLIDINGWINDOW:5:20) [70]. We excluded all reads shorter than 20 bp and merged overlapping paired-end reads with SeqPrep (overlap size = 15) (https://github.com/jstjohn/SeqPrep).

We mapped the resulting reads to the reconstructed ancestral sequence of the MTB complex [7] using the mem algorithm implemented in BWA v 0.7.13 [71]. Duplicated reads were marked by the MarkDuplicates module of Picard v 2.9.1 (https://github.com/broadinstitute/picard). We performed local realignment around Indel with the RealignerTargetCreator and IndelRealigner modules of GATK v 3.4.0 [72]. We used Pysam v 0.9.0 (https://github.com/pysam-developers/pysam) to exclude reads with alignment score lower than (0.93*read_length)-(read_length*4*0.07)): this corresponds to more than 7 miss-matches per 100 bp. We called SNPs with Samtools v 1.2 mpileup [73] and VarScan v 2.4.1 [74] using the following thresholds: minimum mapping quality of 20; minimum base quality at a position of 20; minimum read depth at a position of 7X; minimum percentage of reads supporting the call 90%; no more than 90%, or less than 10% of reads supporting a call in the same orientation (strand bias filter). SNPs in previously defined repetitive regions were excluded (PPE and PE-PGRS genes, phages, insertion sequences and repeats longer than 50 bp) [53]. We excluded all strains with average coverage < 15 X. Additionally, we excluded genomes with more than 50% of the SNPs excluded due to the strand bias filter, and genomes with more than 50% of SNPs with a percentage of reads supporting the call included between 10% and 90%. We filtered out genomes with phylogenetic SNPs belonging to different lineages or sub-lineages

(only for L4) of MTB, as this is an indication that a mix of strains could have been sequenced. To do this, we used the diagnostic SNPs obtained from Steiner et al. 2014 [75] and Stucki et al. 2016 [76] for L4 sub-lineages. We excluded all strains for which we could not find the date of isolation 1) in the SRA meta-information, 2) in the associated publications, 3) from the authors of the original study after inquiry. We divided all remaining strains by lineage (L1 -L6 and *M. bovis*), and excluded strains with a number of called SNPs deviating more than three standard deviations from the mean of the respective lineage. We built SNPs alignments for all lineages including only variable positions with less than 10% of missing data. Finally, we excluded all genomes with more than 10% of missing data in the alignment of the respective lineage. After all filtering steps, we were able to retrieve 6,285 strains with high quality genome sequences and an associated date of sampling (S1 Table).

Dataset subdivision

To perform a systematic analysis of the molecular clock in MTB we considered different data sets:

1. the complete data set (6,285 strains)

2. the different lineages of MTB (L1, L2, L3, L4, L5, L6, *M. bovis*)

3. the sub-lineages of L1 (L1.1.1, L1.1.1.1, L1.1.2, L1.1.3, L1.2.1 and L1.2.2) and L2 (L2.1, L2.2.1, L2.2.2 and L2.2.1.1) as defined by Coll et al. 2014 [77]; the sub-lineages of L4 (L4.1.1, L4.1.2, L4.1.3, L4.4, L4.5, L4.6.1 and L4.10) as defined by Stucki et al. 2016 [76]. Additionally, we identified two L4 clades that were not classified by the diagnostic SNPs of Stucki et al. 2016 [76] (L4_nc and L4.1_nc, respectively, included into L4.6.2 and L4.1.2 as defined by Coll et al. 2014 [77]), and three sub-clades of L2.2.1 that were not previously designated as sub-lineages (L2.2.1_nc1, L2.2.1_nc2 and L2.2.1_nc3) (S9–S11 Figs).

4. Selected data sets representing outbreaks or local populations that have been used for molecular clock analyses in other studies

Lee et al. 2015 [15]—Mj clade outbreak among a Inuit population in Canada (L4)

Eldholm et al. 2015 [14]- Multi-drug resistant outbreak in Argentina (L4)

Eldholm et al. 2016 [20]–Afghan family outbreak in Oslo (L2)

Trewby et al. 2016 [38]–*M. bovis* in Northern Ireland

Crispell et al. 2017 [78]–*M. bovis* in New Zealand

Folkvardsen et al. 2017 [16]—C2/1112-15 outbreak in Denmark (L4)

Bainomugisa et al. 2018 [17]–Multi-drug resistant outbreak on Daru island in PNG (L2)

LSD analysis

For all data sets, we assembled SNPs alignments including variable positions with less than 10% of missing data. We inferred phylogenetic trees with raxml 8.2.11 [79] using a GTR model (-m GTRCAT -V options). Since the alignments contained only variable positions, we rescaled the branch lengths of the trees rescaled_branch_length = ((branch_length * alignment_lengths) / (alignment_length + invariant_sites)), Duchene and colleagues [32] showed that this method produced similar results compared to ascertainment bias correction. We then used the R package ape [80] to perform root to tip regression after rooting the trees in the

position that minimizes the sum of the squared residuals from the regression line. Root to tip regression is only recommended for exploratory analyses of the temporal structure of a dataset and it should not be used for hypothesis testing [28]. A more rigorous approach is the date randomization test (DRT) [81], in which the sampling dates are reshuffled randomly among the taxa and the estimated molecular clock rates estimated from the observed data is compared with the estimates obtained with the reshuffled data sets. This test can show that the observed data has more temporal information that data with random sampling times. For each dataset, we used the least square method implemented in LSD v0.3-beta [31] to estimate the molecular clock in the observed data and in 100 randomized replicates. To do this, we used the QPD algorithm allowing it to estimate the position of the root (option -r a) and calculating the confidence interval (options -f 100 and -s). We defined three different significance levels for the DRT: 1) the simple test is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. 2) The intermediate test is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. 3) The stringent test is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets.

Bayesian phylogenetic analysis

Bayesian molecular clock analyses are computationally demanding and problematic to run on large data sets. Therefore we reduced the thirteen largest data sets (MTBC, L1, L1.1.1, L1.1.1.1, L2, L2.2.1, L2.2.1.1, L2.2.1_nc1, L2.2.1_nc3, L4, L4.1.2, L4.10 and *M. bovis*) to 300 randomly selected strains. For each data set we used the Bayesian information criterion implemented in jModelTest 2.1.10 v20160303 [35] to identify the best fitting nucleotide substitution model among 11 possible schemes including unequal nucleotide frequencies (total models = 22, options -s 11 and -f). We performed Bayesian inference with BEAST2 [34]. We corrected the xml file to specify the number of invariant sites as indicated here: https://groups.google.com/forum/#!topic/beast-users/QfBHMOqImFE, and used the tip sampling year as calibration.

We ran four BEAST analyses with different settings: we used a relaxed lognormal clock model [37], the best fitting nucleotide substitution model according to the results of jModelTest, and two different coalescent priors: constant population size and exponential population growth (or shrinkage). We chose a 1/x prior for the population size $[0–10^9]$, two different priors for the mean of the lognormal distribution of the clock rate (1/x and uniform) $[10^{-10}–10^{-5}]$, a normal(0,1) prior for the standard deviation of the lognormal distribution of the clock rate [0 –infinity]. For the exponential growth rate prior, we used the standard Laplace distribution [-infinity–infinity]. For all data sets, we ran at least two runs, we used Tracer 1.7.1 [82] to identify and exclude the burn-in, to evaluate convergence among runs and to calculate the estimated sample size. We stopped the runs when at least two chains reached convergence, and the ESS of the posterior and of all parameters were larger than 200.

Analyses with the complete MTB complex and aDNA

We analyzed the complete data set of 6,285 genomes with the same methods described above. The only difference was that for the LSD analysis, we rooted the input tree using *Mycobacterium canetti* (SAMN00102920, SRR011186) as outgroup. We did this because we noticed that without outgroup, all methods placed the root on the branch separating *M. bovis* from all other lineages, and not on the branch separating MTB *sensu stricto* from the other lineages.

To test the time dependency hypothesis, we repeated the LSD and BEAST analyses on the MTB complex, adding the aDNA genome sequences of three MTB strains obtained from

Precolumbian Peruvian human remains [8]. These are the most ancient aDNA samples available for MTB. For LSD, we assigned as sampling year the confidence interval of the radiocarbon dating reported in the original publication. For BEAST, we assigned uniform priors spanning the confidence interval but we failed to reach convergence, therefore we used the mean of the maximum and minimum years in the confidence interval (SAMN02727818: 1126 [1028–1224], SAMN02727820: 1117 [1023–1211], SAMN02727821: 1211 [1141–1280]). We ran three different analyses with BEAST: we used the sub-sample of 300 strains with two different priors on the clock rate (1/x and uniform), and an independent sub-sample of 500 strain, for this last data set (500 strains) we assumed a HKY model and used a uniform prior on the clock rate (S2 Table).

To summarize the results of the BEAST analysis with the aDNA samples and retrieve the age of the MRCA of the individual lineages, we considered the analysis performed on the subset of 500 strains: we randomly sampled 5,000 trees from the posterior (after excluding the burn-in), and calculated the Maximum clade credibility tree with the software Treeannotator v2.5.0.

Supporting information

S1 Table. List of strains used in this study with sampling year and accession numbers.
(XLSX)

S2 Table. Results of BEAST and LSD for all data sets.
(XLSX)

S3 Table. The age of the MTB complex and of its lineages resulted from different analyses.
(DOCX)

S4 Table. List of all accession numbers, before filtering.
(XLSX)

S1 Fig. For each data set we report the results of the root to tip regression, where the distance from the root of the tree (expected substitution per site) is plotted against the year of sampling, the results of the date randomization test (DRT) with LSD, and the comparison of the prior and posterior distribution of the clock rate. The simple DRT is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. The intermediate DRT is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. The stringent DRT is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. Large data sets (MTBC, L1, L2, L4 and *M. bovis*) were randomly sub-sampled to 300 strains for the BEAST analysis.
(TIF)

S2 Fig. Root to tip regression analysis of L1, L5 and L6. The difference compared to S1 Fig is that the root was not placed in the position that minimizes the sum of the squared residuals from the regression line, but was obtained from the complete MTBC tree as shown in Fig 3a, and it is therefore defined by the outgroup of each of these lineages.
(TIF)

S3 Fig. For each data set we report the results of the root to tip regression, where the distance from the root of the tree (expected substitution per site) is plotted against the year of sampling, the results of the date randomization test (DRT) with LSD, and the comparison

of the prior and posterior distribution of the clock rate. The simple DRT is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. The intermediate DRT is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. The stringent DRT is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. Large data sets (L1.1.1, L1.1.1.1, L2.2.1, L2.2.1.1, L2.2.1_nc1, L2.2.1_nc3, L4.10, L4.1.2) were randomly sub-sampled to 300 strains for the BEAST analysis. (TIF)

S4 Fig. For each data set we report the results of the root to tip regression, where the distance from the root of the tree (expected substitution per site) is plotted against the year of sampling, the results of the date randomization test (DRT) with LSD, and the comparison of the prior and posterior distribution of the clock rate. The simple DRT is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. The intermediate DRT is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. The stringent DRT is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. Large data sets (L1.1.1, L1.1.1.1, L2.2.1, L2.2.1.1, L2.2.1_nc1, L2.2.1_nc3, L4.10, L4.1.2) were randomly sub-sampled to 300 strains for the BEAST analysis. (TIF)

S5 Fig. For each data set we report the results of the root to tip regression, where the distance from the root of the tree (expected substitution per site) is plotted against the year of sampling, the results of the date randomization test (DRT) with LSD, and the comparison of the prior and posterior distribution of the clock rate. The simple DRT is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. The intermediate DRT is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. The stringent DRT is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. Large data sets (L1.1.1, L1.1.1.1, L2.2.1, L2.2.1.1, L2.2.1_nc1, L2.2.1_nc3, L4.10, L4.1.2) were randomly sub-sampled to 300 strains for the BEAST analysis. (TIF)

S6 Fig. For each data set we report the results of the root to tip regression, where the distance from the root of the tree (expected substitution per site) is plotted against the year of sampling, the results of the date randomization test (DRT) with LSD, and the comparison of the prior and posterior distribution of the clock rate. The simple DRT is passed when the clock rate estimate for the observed data does not overlap with the range of estimates obtained from the randomized sets. The intermediate DRT is passed when the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. The stringent DRT is passed when the confidence interval of the clock rate estimate for the observed data does not overlap with the confidence intervals of the estimates obtained from the randomized sets. Large data sets (L1.1.1, L1.1.1.1, L2.2.1, L2.2.1.1, L2.2.1_nc1, L2.2.1_nc3, L4.10, L4.1.2) were randomly sub-sampled to 300 strains for the BEAST analysis. (TIF)

S7 Fig. Distribution of the sampling years.
(TIF)

S8 Fig. Distribution of the sampling years.
(TIF)

S9 Fig. Sub-lineages of L1 that were included in the analysis. Clades colored in gray did not pass the DRT, clades colored in black passed the DRT. *: simple DRT passed, ** intermediate DRT passed, ***: stringent DRT passed.
(TIF)

S10 Fig. Sub-lineages and outbreaks of L2 that were included in the analysis. Clades colored in gray did not pass the DRT, clades colored in black passed the DRT. *: simple DRT passed, ** intermediate DRT passed, ***: stringent DRT passed. Dotted lines represent two outbreaks from previous studies.
(TIF)

S11 Fig. Sub-lineages and outbreaks of L4 that were included in the analysis. Clades colored in gray did not pass the DRT, clades colored in black passed the DRT. *: simple DRT passed, ** intermediate DRT passed, ***: stringent DRT passed. Dotted lines represent three outbreaks from previous studies.
(TIF)

S12 Fig. Results of the DRT for all data sets ordered by genetic diversity (Watterson's estimator and number of polymorphic positions) and temporal range. Data sets with fewer strains sampled in a shorter period of time tended to fail the DRT irrespectively of the genetic diversity of the data set.
(TIF)

S13 Fig. Comparison of different priors on the clock rate (1/x prior and uniform prior). The uniform prior place most weight on high clock rates, while the 1/x prior distributes the weight through all orders of magnitude.
(TIF)

S14 Fig. Posterior distribution of the clock rate, obtained with two different priors (1/x and uniform [10^{-10}–10^{-5}]). The prior distributions for the two analyses are shown in S13 Fig.
(TIF)

S15 Fig. Posterior distribution of the clock rate, obtained with two different priors (1/x and uniform [10^{-10}–10^{-5}]). The prior distributions for the two analyses are shown in S13 Fig.
(TIF)

S16 Fig. Posterior distribution of the clock rate, obtained with two different priors (1/x and uniform [10^{-10}–10^{-5}]). The prior distributions for the two analyses are shown in S13 Fig.
(TIF)

S17 Fig. Comparison of the posterior distribution of the clock rate obtained with a constant population size and an exponential population growth prior.
(TIF)

S18 Fig. Comparison of the posterior distribution of the clock rate obtained with a constant population size and an exponential population growth prior.
(TIF)

S19 Fig. Phylogenetic tree of data set L4_nc with tips colored according to the year of sampling.
(TIF)

S20 Fig. Posterior distribution of the clock rate for L1 and L4. These are the results of the analysis with the 1/x prior on the clock rate and the exponential population growth (or shrinkage) prior.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Fabrizio Menardo, Sebastian Duchêne.

Data curation: Fabrizio Menardo, Daniela Brites.

Formal analysis: Fabrizio Menardo.

Funding acquisition: Sebastien Gagneux.

Investigation: Fabrizio Menardo, Sebastian Duchêne, Daniela Brites, Sebastien Gagneux.

Methodology: Fabrizio Menardo, Sebastian Duchêne.

Resources: Sebastien Gagneux.

Visualization: Fabrizio Menardo.

Writing – original draft: Fabrizio Menardo.

Writing – review & editing: Fabrizio Menardo, Sebastian Duchêne, Daniela Brites, Sebastien Gagneux.

References

1. Zuckerkandl E., & Pauling L. (1962). Molecular disease, evolution and genetic heterogeneity.

2. Morgan G. J. (1998). Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965. *Journal of the History of Biology*, 31(2), 155–178. PMID: 11620303

3. Kimura M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129), 624–626. https://doi.org/10.1038/217624a0 PMID: 5637732

4. Drummond A. J., Pybus O. G., Rambaut A., Forsberg R., & Rodrigo A. G. (2003). Measurably evolving populations. *Trends in ecology & evolution*, 18(9), 481–488.

5. Kapur V., Whittam T. S., & Musser J. M. (1994). Is Mycobacterium tuberculosis 15,000 years old?. *Journal of Infectious Diseases*, 170(5), 1348–1349. https://doi.org/10.1093/infdis/170.5.1348 PMID: 7963745

6. Cole S., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., . . . & Tekaia F. (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), 537. https://doi.org/10.1038/31159 PMID: 9634230

7. Comas I., Coscolla M., Luo T., Borrell S., Holt K. E., Kato-Maeda M., . . . & Gagneu S. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature genetics*, 45(10), 1176. https://doi.org/10.1038/ng.2744 PMID: 23995134

8. Bos K. I., Harkins K. M., Herbig A., Coscolla M., Weber N., Comas I., . . . & Campbell T. J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, 514(7523), 494. https://doi.org/10.1038/nature13591 PMID: 25141181

9. Merker M., Blin C., Mona S., Duforet-Frebourg N., Lecher S., Willery E., . . . & Allix-Béguec C. (2015). Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nature genetics*, 47(3), 242. https://doi.org/10.1038/ng.3195 PMID: 25599400

10. Kay G. L., Sergeant M. J., Zhou Z., Chan J. Z. M., Millard A., Quick J., . . . & Achtman M. (2015). Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature communications*, 6, 6717. https://doi.org/10.1038/ncomms7717 PMID: 25848958

11. Brynildsrud O. B., Pepperell C. S., Suffys P., Grandjean L., Monteserin J., Debech N., . . . & Fandinho F. (2018). Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. *Science Advances*, 4(10), eaat5869.

12. Liu Q., Ma A., Wei L., Pang Y., Wu B., Luo T., . . . & Gao Q. (2018). China's tuberculosis epidemic stems from historical expansion of four strains of Mycobacterium tuberculosis. *Nature ecology & evolution*, 1.

13. Rutaihwa L. K., Menardo F., Stucki D., Gygli S. M., Ley S. D., Malla B., . . . & Gagneux S. (2019). Multiple Introductions of the Mycobacterium tuberculosis Lineage 2 Beijing into Africa over centuries. *Frontiers in Ecology and Evolution*, 7, 112.

14. Eldholm V., Monteserin J., Rieux A., Lopez B., Sobkowiak B., Ritacco V., & Balloux F. (2015). Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. *Nature communications*, 6, 7119. https://doi.org/10.1038/ncomms8119 PMID: 25960343

15. Lee R. S., Radomski N., Proulx J. F., Levade I., Shapiro B. J., McIntosh F., . . . & Behr M. A. (2015). Population genomics of Mycobacterium tuberculosis in the Inuit. *Proceedings of the National Academy of Sciences*, 112(44), 13609–13614.

16. Folkvardsen D. B., Norman A., Andersen Å. B., Michael Rasmussen E., Jelsbak L., & Lillebaek T. (2017). Genomic Epidemiology of a Major Mycobacterium tuberculosis Outbreak: Retrospective Cohort Study in a Low-Incidence Setting Using Sparse Time-Series Sampling. *The Journal of infectious diseases*, 216(3), 366–374. https://doi.org/10.1093/infdis/jix298 PMID: 28666374

17. Bainomugisa A., Lavu E., Hiashiri S., Majumdar S., Honjepari A., Moke R., . . . & Coulter C. (2018). Multi-clonal evolution of multi-drug-resistant/extensively drug-resistant Mycobacterium tuberculosis in a high-prevalence setting of Papua New Guinea for over three decades. *Microbial genomics*, 4(2).

18. Kühnert D., Coscolla M., Brites D., Stucki D., Metcalfe J., Fenner L., . . . & Stadler T. (2018). Tuberculosis outbreak investigation using phylodynamic analysis. *Epidemics*, 25, 47–53. https://doi.org/10.1016/j.epidem.2018.05.004 PMID: 29880306

19. Cohen K. A., Abeel T., McGuire A. M., Desjardins C. A., Munsamy V., Shea T. P., . . . & Chapman S. B. (2015). Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of Mycobacterium tuberculosis isolates from KwaZulu-Natal. *PLoS medicine*, 12(9), e1001880. https://doi.org/10.1371/journal.pmed.1001880 PMID: 26418737

20. Eldholm V., Pettersson J. H. O., Brynildsrud O. B., Kitchen A., Rasmussen E. M., Lillebaek T., . . . & Alfsnes K. (2016). Armed conflict and population displacement as drivers of the evolution and dispersal of Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, 113(48), 13881–13886.

21. Pepperell C. S., Casto A. M., Kitchen A., Granka J. M., Cornejo O. E., Holmes E. C., . . . & Feldman M. W. (2013). The role of selection in shaping diversity of natural M. tuberculosis populations. *PLoS pathogens*, 9(8), e1003543. https://doi.org/10.1371/journal.ppat.1003543 PMID: 23966858

22. Merker M., Barbier M., Cox H., Rasigade J. P., Feuerriegel S., Kohl T. A., . . . & Andres S. (2018). Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife*, 7:e38200. https://doi.org/10.7554/eLife.38200 PMID: 30373719

23. Seo T. K., Thorne J. L., Hasegawa M., & Kishino H. (2002). A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics*, 18(1), 115–123. https://doi.org/10.1093/bioinformatics/18.1.115 PMID: 11836219

24. Wirth T., Hildebrand F., Allix-Béguec C., Wölbeling F., Kubica T., Kremer K., . . . & Meyer A. (2008). Origin, spread and demography of the Mycobacterium tuberculosis complex. *PLoS pathogens*, 4(9), e1000160. https://doi.org/10.1371/journal.ppat.1000160 PMID: 18802459

25. Ho S. Y., Phillips M. J., Cooper A., & Drummond A. J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular biology and evolution*, 22(7), 1561–1568. https://doi.org/10.1093/molbev/msi145 PMID: 15814826

26. Ho S. Y., Lanfear R., Bromham L., Phillips M. J., Soubrier J., Rodrigo A. G., & Cooper A. (2011). Time-dependent rates of molecular evolution. *Molecular ecology*, 20(15), 3087–3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x PMID: 21740474

27. Rieux A., & Balloux F. (2016). Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular ecology*, 25(9), 1911–1924. https://doi.org/10.1111/mec.13586 PMID: 26880113

28. Rambaut A., Lam T. T., Max Carvalho L., & Pybus O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus evolution*, 2(1), vew007.

29. Rambaut A., Lam T. T., Max Carvalho L., & Pybus O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus evolution*, 2(1), vew007.

30. Tong K. J., Duchêne D. A., Duchêne S., Geoghegan J. L., & Ho S. Y. (2018). A comparison of methods for estimating substitution rates from ancient DNA sequence data. *BMC evolutionary biology*, 18(1), 70. https://doi.org/10.1186/s12862-018-1192-3 PMID: 29769015

31. To T. H., Jung M., Lycett S., & Gascuel O. (2015). Fast dating using least-squares criteria and algorithms. *Systematic biology*, 65(1), 82–97. https://doi.org/10.1093/sysbio/syv068 PMID: 26424727

32. Duchêne S., Duchene D. A., Geoghegan J. L., Dyson Z. A., Hawkey J., & Holt K. E. (2018). Inferring demographic parameters in bacterial genomic data using Bayesian and hybrid phylogenetic methods. *BMC evolutionary biology*, 18(1), 95. https://doi.org/10.1186/s12862-018-1210-5 PMID: 29914372

33. Duchêne S., Geoghegan J. L., Holmes E. C., & Ho S. Y. (2016a). Estimating evolutionary rates using time-structured data: a general comparison of phylogenetic methods. *Bioinformatics*, 32(22), 3375–3379.

34. Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C. H., Xie D., . . . & Drummond A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, 10(4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537 PMID: 24722319

35. Darriba D., Taboada G. L., Doallo R., & Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8), 772.

36. Duchêne S., Holt K. E., Weill F. X., Le Hello S., Hawkey J., Edwards D. J., . . . & Holmes E. C. (2016b). Genome-scale rates of evolutionary change in bacteria. *Microbial Genomics*, 2(11).

37. Drummond A. J., Ho S. Y., Phillips M. J., & Rambaut A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS biology*, 4(5), e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

38. Trewby H., Wright D., Breadon E. L., Lycett S. J., Mallon T. R., McCormick C., . . . & Herzyk P. (2016). Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics*, 14, 26–35. https://doi.org/10.1016/j.epidem.2015.08.003 PMID: 26972511

39. Möller S., du Plessis L., & Stadler T. (2018). Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 115(16), 4200–4205.

40. Bromham L., Duchêne S., Hua X., Ritchie A. M., Duchêne D. A., & Ho S. Y. (2018). Bayesian molecular dating: opening up the black box. *Biological Reviews*, 93(2), 1165–1191. https://doi.org/10.1111/brv.12390 PMID: 29243391

41. Duchêne S., Duchêne D., Holmes E. C., & Ho S. Y. (2015). The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Molecular Biology and Evolution*, 32(7), 1895–1906. https://doi.org/10.1093/molbev/msv056 PMID: 25771196

42. Hatherell H. A., Colijn C., Stagg H. R., Jackson C., Winter J. R., & Abubakar I. (2016). Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC medicine*, 14(1), 21.

43. Holt K. E., McAdam P., Thai P. V. K., Thuong N. T. T., Ha D. T. M., Lan N. N., . . . & Thwaites G. (2018). Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nature genetics*, 1.

44. Glynn J. R., Kremer K., Borgdorff M. W., Rodriguez M. P., & Soolingen D. V. (2006). Beijing/W genotype Mycobacterium tuberculosis and drug resistance.

45. Hanekom M., Mata D., van Pittius N. G., van Helden P. D., Warren R. M., & Hernandez-Pando R. (2010). Mycobacterium tuberculosis strains with the Beijing genotype demonstrate variability in virulence associated with transmission. *Tuberculosis*, 90(5), 319–325. https://doi.org/10.1016/j.tube.2010.08.004 PMID: 20832364

46. de Steenwinkel J. E., Marian T., de Knegt G. J., Kremer K., Aarnoutse R. E., Boeree M. J., . . . & Bakker-Woudenberg I. A. (2012). Drug susceptibility of Mycobacterium tuberculosis Beijing genotype and association with MDR TB. *Emerging infectious diseases*, 18(4), 660. https://doi.org/10.3201/eid1804.110912 PMID: 22469099

47. Ribeiro S. C., Gomes L. L., Amaral E. P., Andrade M. R., Almeida F. M., Rezende A. L., . . . & Lasuns-kaia E. B. (2014). Mycobacterium tuberculosis strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *Journal of clinical microbiology*, JCM-00498.

48. Wiens K. E., Woyczynski L. P., Ledesma J. R., Ross J. M., Zenteno-Cuevas R., Goodridge A., . . . & Ray S. E. (2018). Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC medicine*, 16(1), 196. https://doi.org/10.1186/s12916-018-1180-x PMID: 30373589

49. Ford C. B., Shah R. R., Maeda M. K., Gagneux S., Murray M. B., Cohen T., . . . & Fortune S. M. (2013). Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nature genetics*, 45(7), 784. https://doi.org/10.1038/ng.2656 PMID: 23749189

50. Stimson J., Gardy J., Mathema B., Crudu V., Cohen T., & Colijn C. (2019). Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. *Molecular biology and evolution*, 36(3), 587–603. https://doi.org/10.1093/molbev/msy242 PMID: 30690464

51. Brosch R., Gordon S. V., Marmiesse M., Brodin P., Buchrieser C., Eiglmeier K., . . . & Parsons L. M. (2002). A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proceedings of the national academy of Sciences*, 99(6), 3684–3689.

52. Mostowy S., Cousins D., Brinkman J., Aranaz A., & Behr M. A. (2002). Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. *The Journal of infectious diseases*, 186(1), 74–80. https://doi.org/10.1086/341068 PMID: 12089664

53. Brites D., Loiseau C., Menardo F., Borrell S., Boniotti M. B., Warren R., . . . & Fyfe J. A. (2018). A New Phylogenetic Framework for the Animal-adapted Mycobacterium tuberculosis Complex. *Frontiers in Microbiology*, 9.

54. Ohta T. (1987). Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*, 26(1–2), 1–6.

55. Bromham L., & Penny D. (2003). The modern molecular clock. *Nature Reviews Genetics*, 4(3), 216. https://doi.org/10.1038/nrg1020 PMID: 12610526

56. Duchêne S., Holmes E. C., & Ho S. Y. (2014). Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc. R. Soc. B*, 281(1786), 20140732. https://doi.org/10.1098/rspb.2014.0732 PMID: 24850916

57. Emerson B. C., & Hickerson M. J. (2015). Lack of support for the time-dependent molecular evolution hypothesis. *Molecular ecology*, 24(4), 702–709. https://doi.org/10.1111/mec.13070 PMID: 25640964

58. Rieux A., Eriksson A., Li M., Sobkowiak B., Weinert L. A., Warmuth V., . . . & Balloux F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. *Molecular Biology and Evolution*, 31(10), 2780–2792. https://doi.org/10.1093/molbev/msu222 PMID: 25100861

59. Ford C. B., Lin P. L., Chase M. R., Shah R. R., Iartchouk O., Galagan J., . . . & Flynn J. L. (2011). Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. *Nature genetics*, 43(5), 482. https://doi.org/10.1038/ng.811 PMID: 21516081

60. Chan J. Z. M., Sergeant M. J., Lee O. Y. C., Minnikin D. E., Besra G. S., Pap I., . . . & Pallen M. J. (2013). Metagenomic analysis of tuberculosis in a mummy. *New England Journal of Medicine*, 369(3), 289–290. https://doi.org/10.1056/NEJMc1302295 PMID: 23863071

61. Sabin, S., Herbig, A., Vågene, Å. J., Ahlström, T., Bozovic, G., Arcini, C., . . . & Bos, K. I. (2019). A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex. BioRxiv, 588277.

62. Brites D., & Gagneux S. (2015). Co-evolution of M ycobacterium tuberculosis and H omo sapiens. *Immunological reviews*, 264(1), 6–24. https://doi.org/10.1111/imr.12264 PMID: 25703549

63. Rothschild B. M., Martin L. D., Lev G., Bercovier H., Bar-Gal G. K., Greenblatt C., . . . & Brittain D. (2001). Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clinical Infectious Diseases*, 33(3), 305–311. https://doi.org/10.1086/321886 PMID: 11438894

64. Taylor G. M., Murphy E., Hopkins R., Rutland P., & Chistov Y. (2007). First report of Mycobacterium bovis DNA in human remains from the Iron Age. *Microbiology*, 153(4), 1243–1249.

65. Hershkovitz I., Donoghue H. D., Minnikin D. E., Besra G. S., Lee O. Y., Gernaey A. M., . . . & Bar-Gal G. K. (2008). Detection and molecular characterization of 9000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PloS one*, 3(10), e3426. https://doi.org/10.1371/journal.pone.0003426 PMID: 18923677

66. Nicklisch N., Maixner F., Ganslmeier R., Friederich S., Dresely V., Meller H., . . . & Alt K. W. (2012). Rib lesions in skeletons from early neolithic sites in Central Germany: on the trail of tuberculosis at the onset of agriculture. *American journal of physical anthropology*, 149(3), 391–404. https://doi.org/10.1002/ajpa.22137 PMID: 23042554

67. Wilbur A. K., Bouwman A. S., Stone A. C., Roberts C. A., Pfister L. A., Buikstra J. E., & Brown T. A. (2009). Deficiencies and challenges in the study of ancient tuberculosis DNA. *Journal of Archaeological Science*, 36(9), 1990–1997.

68. Donoghue H. D., Hershkovitz I., Minnikin D. E., Besra G. S., Lee O. Y. C., Galili E., . . . & Bar-Gal G. K. (2009). Biomolecular archaeology of ancient tuberculosis: response to "Deficiencies and challenges in the study of ancient tuberculosis DNA" by Wilbur et al. (2009). *Journal of Archaeological Science*, 36 (12), 2797–2804.

69. Menardo F., Loiseau C., Brites D., Coscolla M., Gygli S. M., Rutaihwa L. K., . . . & Gagneux S. (2018). Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC bioinformatics*, 19(1), 164. https://doi.org/10.1186/s12859-018-2164-8 PMID: 29716518

70. Bolger A. M., Lohse M., & Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

71. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

72. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20:1297–303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199

73. Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21):2987–93. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

74. Koboldt D., Zhang Q., Larson D., Shen D., McLellan M., Lin L., Miller C., Mardis E., Ding L., & Wilson R. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing *Genome Research*, 22(3), 568–576. https://doi.org/10.1101/gr.129684.111 PMID: 22300766

75. Steiner A., Stucki D., Coscolla M., Borrell S., & Gagneux S. (2014). KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes. *BMC genomics*, 15(1), 881.

76. Stucki D., Brites D., Jeljeli L., Coscolla M., Liu Q., Trauner A., . . . & Gao Q. (2016). Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature genetics*, 48(12), 1535. https://doi.org/10.1038/ng.3704 PMID: 27798628

77. Coll F., McNerney R., Guerra-Assuncao J. A., Glynn J. R., Perdigao J., Viveiros M., . . . & Clark T. G. (2014). A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. *Nature communications*, 5, 4812. https://doi.org/10.1038/ncomms5812 PMID: 25176035

78. Crispell J., Zadoks R. N., Harris S. R., Paterson B., Collins D. M., de-Lisle G. W., . . . & Kao R. R. (2017). Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC genomics*, 18(1), 180. https://doi.org/10.1186/s12864-017-3569-x PMID: 28209138

79. Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

80. Paradis E., Schliep K., & Schwartz R. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 1, 3.

81. Ramsden C., Melo F. L., Figueiredo L. M., Holmes E. C., Zanotto P. M., & VGDN Consortium. (2008). High rates of molecular evolution in hantaviruses. *Molecular Biology and Evolution*, 25(7), 1488–1492. https://doi.org/10.1093/molbev/msn093 PMID: 18417484

82. Rambaut A., Drummond A. J., Xie D., Baele G., & Suchard M. A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol*, 10.