The paradoxes of Mycobacterium tuberculosis molecular evolution and consequences for the inference of tuberculosis emergence date

Rima Zein Eddine, F. Hak, A. Le Meur, C. Genestet, O. Dumitrescu, C. Guyeux, G. Senelle, C. Sola, G. Refrégier

Title : The paradoxes of *Mycobacterium tuberculosis* molecular evolution and consequences for the inference of tuberculosis emergence date

Authors

R. Zein Eddine[1], F. Hak[2], A. Le Meur[2], C. Genestet[3,4], O. Dumitrescu[3,4], C. Guyeux[5], G. Senelle[5], C. Sola[6,7], G. Refrégier[2*]

Declarations of interest: none

Authors' agreement: all authors have approved the final version of the manuscript

Author contributions

Rima Zein-Eddine, Fiona Hak, Adrien Le Meur , Ccharlotte Genestet, Gaëtan Senelle performed data curation and formal analyses of data at the basis of this work. Guislaine Refrégier created the model. Christophe Guyeux, Oana Dumitrescu, Christophe Sola, Guislaine Refrégier participated in the funding acquisition. Guislaine Refrégier wrote the initial draft and performed the final editing. Christophe Guyeux, Adrien Le Meur, Christophe Sola participated in editing the manuscript. Adrien Le Meur created the figure.

The paradoxes of *Mycobacterium tuberculosis* molecular evolution and consequences for the inference of tuberculosis emergence date

R. Zein Eddine[1], F. Hak[2], A. Le Meur[2], C. Genestet[3,4], O. Dumitrescu[3,4], C. Guyeux[5], G. Senelle[5], C. Sola[6,7], G. Refrégier[2*]

1 Laboratoire d'Optique et Biosciences, Ecole Polytechnique, Institut National de la Santé et de la Recherche Médicale : U1182, Centre National de la Recherche Scientifique : UMR7645, France

2 Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, 91190, Gif-sur-Yvette, France.

3 CIRI - Centre International de Recherche en Infectiologie, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon-1, Inserm U1111, CNRS UMR5308, Lyon, France

4 Hospices Civils de Lyon, Institut des Agents Infectieux, Laboratoire de bactériologie, Lyon, France3

5 DISC Computer Science Department, FEMTO-ST Institute, UMR 6174 CNRS, Univ. Bourgogne Franche-Comté (UBFC), 16 Route de Gray, 25000 Besançon, France;

6 Université de Paris, IAME, UMR1137, INSERM, Paris, France and AP-HP, GHU Nord, service de mycobactériologie spécialisée et de référence, Laboratoire associé du Centre National de Référence des mycobactéries et résistance des mycobactéries aux antituberculeux (CNR-MyRMA), Paris, France

7 Université Paris-Saclay, Saint-Aubin, France

*corresponding author : guislaine.refregier@universite-paris-saclay.fr lab. Ecologie Systématique et Evolution, 12 rue 128, 91190, Gif-sur-Yvette, France. Tel : +33 1 69 15 77 20.

REVIEW

Abstract

The date of tuberculosis emergence has been the subject of such a long debate that it almost resembles a Graal quest. New studies joining archaeological efforts with sequencing methods raise high hopes for solving the question. Yet, inferring the emergence of the disease based on samples from epidemics that occurred relatively long after this emergence requires a molecular clock for the *Mycobacterium tuberculosis* complex. Several clocks inferred from different types of loci and/or different samples, using both sound reasoning and reliable data, are actually very different, which we call the paradoxes of *M. tuberculosis* molecular evolution.

After having presented these paradoxes, we will remind the limits of the molecular clocks used in the different studies such as assumption of homogeneous substitution rate. We will then review recent results that shed new light on the characteristics of *M. tuberculosis* molecular evolution: traces of diverse selection pressures, the impact of host dynamics, etc. We provide some ideas on what to do next to get nearer to the Graal. Among them, the collection of additional remains from ancient tuberculosis seems still essential.

Keywords: mutation rate, substitution rate, positive selection, purifying selection, pathogen.

Highlights: Tuberculosis is still the first deadly infectious disease caused by a bacterium. It has imposed a high death toll on humankind for millennia, but what key events underlie this success is still not clear. More specifically, the date of its wide success, which we will refer to as tuberculosis emergence, is still highly debated, between 70,000 to 6,000 years Before Present. We discuss the relative reliability and the different assumptions at the basis of these datings and provide a model of all parameters affecting pathogens molecular clocks. We argue that remains with ancient *M. tuberculosis* are still needed to shed light on tuberculosis main diversification events.

1. Introduction

Identifying the role of microorganisms in infectious diseases was key to the increase in life expectancies that has characterized the XXth century [1, 2]). Reconstructing the history of a specific disease helps understanding future threats and thus avoid them or limit their impact. Phylogenetics and phylogenomics infer the common ancestry and evolutionary history of organisms based on DNA sequences, either from a selection of genes or from complete genomes [3]. It has proven to be a very powerful tool to disentangle the origin of recent epidemics [4, 5]. On a wider timescale, the picture is however less clear as exemplified by the debates on the precise origin of 1918 flu [6-8].

Tuberculosis is both an ancient disease and the presently most deadly bacteria-related infectious disease. Works of art and bones from Predynastic Egypt and the Middle Kingdom testify to the presence of the disease [9-11]. Tedious work from archaeologists continue to fuel the evidence of past traces of tuberculosis [12, 13], suggesting a high prevalence of the disease already 10,000 y Before Present (BP): Baker et al. identified osteological evidence highly suggestive of tuberculosis in Syria around 10,000 y Before Present in 10% of the remains [12]. Yet the road seems still long to make a clear history of tuberculosis disease using only osteological evidences of the disease. Interdisciplinary studies mixing archaeology and phylogenomics emerge to try to provide definite answers on the global history of tuberculosis [14], potentially reducing the interest in further archaeological explorations with or without ancient DNA. We however want here to draw attention to the hypotheses underlying these inferences. Here, we first present the different dates of tuberculosis emergence according to various phylogenetical inferences, and we clarify the hypotheses of the underlying models. Second, we will present compelling data that show how much these hypotheses are violated in the bacterium evolution. We will focus on *M. tuberculosis* complex *sensu stricto* (hereafter referred to as MTBC) that includes all human-adapted tuberculosis lineages L1-L9 and the animal strains. We exclude *M. canettii*

despite its high sequence identity with the rest of the lineages, due to its poor contribution to the present human epidemics and widely different timescales [15].

2. Various Inferences regarding the age of tuberculosis and the relying evidences

Most phylogenomics studies infer divergence dates using a molecular clock in a Bayesian framework. They assume that DNA evolves at the same pace along the whole phylogeny. This molecular clock is usually assessed either from recent heterochronous sampling of recent epidemics, or from dated fossils [16]. Pathogens codiverging with hosts (and vice-versa) benefit from a third possibility: relying on the partner's phylogeny [17]. We detail below the different types of studies that tried to infer the date of tuberculosis emergence.

The emergence date of tuberculosis was first estimated using synonymous mutations from four *M. tuberculosis* genes and the application of a mutation rate derived from *E. coli* and *S. typhimurium* [18]. The use of other bacteria's rate seems reasonable as they evolve drug resistance in similar proportions in lab experiments (David 1971). This study suggested that *M. tuberculosis* clonal expansion dates back to 15,300-20,400 years Before Present (BP). However, a closer look at the data shows that this inference derived from only four synonymous mutations from a sample of 31 isolates. In addition, these rare mutations occurred in a heterogeneous subselection of 4 genes with very diverse functions (housekeeping genes targeted by drugs, antigens) [18]. In addition, it is difficult to reconstru[15]ct how well the selected isolates represented the global diversity. They might have corresponded only to L4 as it was the most widely sampled lineage at this time, or to modern MTBC (including L2 and potentially L3 lineages), or to the whole diversity including L1 and the L5-L6-animal strains group. If the proposed dating stands for L4 emergence, this would mean that global MTBC expansion would be twice older. Kapur et al. dating might therefore only roughly reflect the actual date of tuberculosis emergence.

The second type of estimation derived from the mutation rate of minisatellites data [19]. This approach has the advantage of circumventing potential biases due to low single nucleotide mutation rate. The diversity of MIRU profiles leads to date the MTBC Most Recent Common Ancestor (MRCA) to 40,000 years BP. This dating corresponds to a coalescence time of main L4 sublineages such as LAM (L4.3) of 7,000 years BP [19].

Another type of estimations is based on a molecular clock derived from *M. tuberculosis* ancient DNA. The corresponding studies took both synonymous and non-synonymous mutations without distinction. This choice is motivated by the quest of precision in the mutation rate estimation. It is supported by the clonal evolution of MTBC and an average ratio between synonymous and non-

synonymous mutations (dN/dS) close to 1 (~0.7) [20]. Such a ratio suggests that no selection occured. Yet, another possibility is the combination of positive/diversifying selection (dN/dS higher than 1) to negative/purifying selection (dN/dS below 1) (see further discussion below). The first ancient genome was sequenced from a Peruvian mummy infected by *M. pinnipedii* [21]. The inferred mutation rate was estimated at ~5. 10^{-8} subst/site/year. Applying this rate to the global phylogeny induces to date MTBC MRCA at ~6,000 y before present [21]. Interestingly, new estimations based on L4 MTBC isolated from Hungarian mummies come out with similar estimations (Brynildsrud et al, 2017; Sabin et al, 2020).

At last, another study inferring a molecular clock used the codivergence of *M. tuberculosis* with its human hosts. The peculiar history and very restricted expansion of L4.2.2 sublineage allowed Refrégier al (2016) to support the same molecular clock as Bos et al., dating MTBC MRCA to ~6,000 BP. Of note, the authors do not argue in favour of this hypothesis as discussed below (Refrégier et al, 2016). Interestingly, this article pinpoints the fact that the molecular clock inferred from heterochronous sampling of recent epidemics is around twice faster as found by several independent studies at around 10^{-7} subst/site/year [22-24].

Apart from the studies using molecular clocks, Comas et al. dated MTBC clonal expansion using the human phylogeny. Based on some similarities between the phylogenetic trees and geographic patterns of *M. tuberculosis* and modern *H. sapiens* involved in the Out-of-Africa migration event, emergence was reevaluated to ~70,000 y before Present, questioning the Neolithic nature of tuberculosis emergence (Comas et al, 2013). The corresponding global mutation rate deriving from this inference is ~3. 10-9 subst/site/year which is around 1 order of magnitude less than that of *Y. pestis* and 3 orders less than *S. pneumoniae* , and also 1 order of magnitude less than that of studies based on ancient DNA [25].

Altogether, the estimated dates of tuberculosis emergence span a very wide time range, going from 6,000 to 70,000 years BP. Younger inferences were obtained using both synonymous and non-synonymous mutations of ancient MTBC isolates, whereas inferences based solely on synonymous mutations or minisatellites data or codivergence with the host come up with elder age of emergence. We will next examine deeper why inferences based on both synonymous and non-synonymous mutations might be erroneous.

3. Assumptions at the basis of molecular dating

Errors of several orders of magnitude can happen when inferring dates with sequences that do not evolve with constant substitution rates *i.e.* with constant molecular clock [26]. Let us remind here that substitution rate differs from the mutation rate by two characteristics: substitution looks at accumulated modifications in the population after selection action, and substitution rate is usually measured per time (per year) instead of per generation. Altogether, both the bacterial multiplication rate, whether intra or interpatient, and selection are evolutionary forces controlling the impact of mutation rate on the molecular clock (Grey ovals, Fig. 1). We will now detail how substitution rate (and thus molecular clock) can vary among genes and how selection and bacterial multiplication may influence it.

Many nucleotides in a genome do not evolve with the same substitution rate. This is obviously true between genes when one has undergone horizontal gene transfer and thus changed cellular environment that controls mutation rate, nucleotide concentrations, for all types of mutations [27]. The divergence between single nucleotide substitution rate is largely inflated between synonymous and non-synonymous mutations as a consequence of selection: positive/diversifying selection is characterized by the fixation of non-synonymous mutations conferring a selective advantage to the organism, and negative/purifying selection corresponds to the elimination of non-synonymous mutations that reduce organism fitness. This is actually the reason why the non-synonymous/synonymous substitution ratio, the dN/dS, is indicative of different selection pressures between genes [20]. Several studies have in fact documented the existence of positive/diversifying selection pressures on MTBC genes. This includes genes targeted by drugs but also other genes involved in host-pathogen interaction [28-31]. Recent exploration of trajectories of dN/dS showed strikingly different dN/dS trajectories supporting a major past selection pressure on epitopes, and a more recent pressure on toxin-antitoxin systems [32]. The selection on MTBC epitopes may relate to the reciprocal selection of MTBC onto human genome [33]. Experimental evolution selecting for biofilm formation also showed a high propensity for adaptation of MTBC, especially active upon regulatory proteins, as well as an important role for epistatic interactions [34] (gene/sequences types in red in Fig. 1). Moreover, mutation rate also depends on DNA repair systems and on oxidative damage as recently shown by Moreno-Molina *et al.* [35] and Genestet *et al.*[31]. Last, duplicated genes such as PE-PPE undergo occasional intragenomic recombination, which should also be taken into account if their mutations are taken into account [36]. Obviously thus, the pace at which non-synonymous mutations accumulate in different types of genes varies due to selection, potentially heterogeneous mutation rate. This advocates against the common practice of merging synonymous and non-synonymous mutations when building a molecular clock [37]. The current studies using ancient DNA all suffer from this limitation.

We showed that intra-host selection acting upon non-synonymous mutations impacts MTBC molecular clock. We want to underline in addition that these selections are likely to have varied through time and space. Indeed, intra-host selection depends on patient's drug treatment, but also on the characteristics of the immune system (efficiency, adaptation). These characteristic first depend on the host genetics [33]. They obviously also highly depend on other diseases. For instance, HIV-status is a strong predictor of tuberculosis disease and tuberculosis regained strength with HIV dissemination around the 80ies [38], but other diseases such as diabetes are also drivers of tuberculosis [39]. And of course recently, COVID has shown to have indirectly impacted tuberculosis transmission by reducing the diagnostic measures efficiency [40]. Other diseases may have greatly impacted the ability of the immune system to fight tuberculosis over time and space, either because they exhaust the body, reduce immune efficiency or because they may provide cross-immunity such as leprosy [41](Fig. 1). Other key parameters for MTBC opportunities to disperse are: demography, which is impacted by social parameters such as wars, economic wealth, etc and impacts migration; ecological changes linked to present global warming or past cooling (the Little Ice Age) [42]

A last important point to bear in mind when considering molecular clock *versus* mutation rate is that mutation rate is usually measured as a number of nucleotide changes per generation. This is because most mutations are linked to replication infidelity and DNA repair [43-45]. But generations can have different length time. Molecular clock may thus largely depend on the pace at which MTBC multiplies inside the host (intrahost multiplication). Ford et al questioned this point by comparing the accumulation of MTBC mutations in infected macaques carrying either latent or active tuberculosis [46]. They could not detect a difference between the two categories but the power they had to do so was limited due to the low number of replicates and the relatively short time of investigation, which allowed to accumulate 0 to 3 substitutions in 91 to 507 days depending on the animal survival. In addition, the study did not discriminate between synonymous and non-synonymous mutations.

We thus showed that non-synonymous mutations may have diverse substitution rates *i.e.* molecular clocks, with one key evolutionary force at play being selection. This should prompt us to focus on synonymous mutations. The synonymous mutation rate was explored shortly after the sequencing of the second human MTBC strain, CDC1551 [47]. Interestingly, this rate appeared to vary highly between genes. Recent deep exploration of diversity also suggests that various mutation rates prevail along the genome of bacteria [48]. Synonymous mutations could thus be an incomplete solution to the problem of identifying a reliable molecular clock.

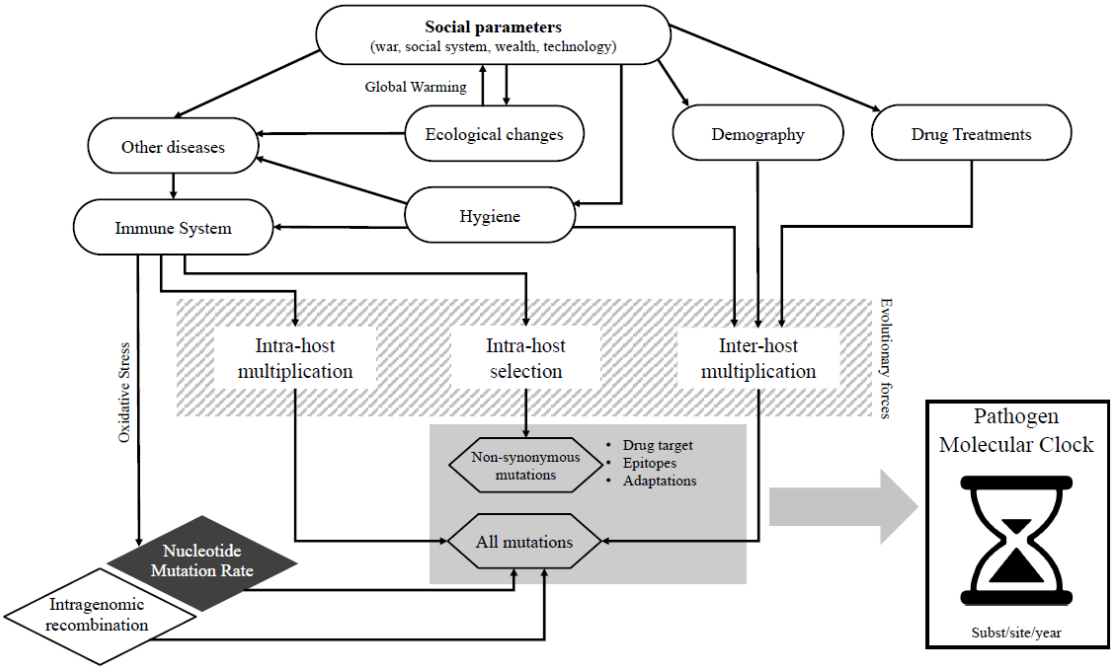

Figure 1 – Parameters affecting molecular pathogens' clock, a model built for MTBC

We distinguish two types of targets: all mutations or only non-synonymous mutations (grey hexagons). The main drivers of the difference between mutation and the molecular clock are intra and inter-host multiplication and intra-host selection (evolutionary forces rectangle). Parameters affecting intra and inter-host processes are indicated above. They include community parameters (social characteristics, demography), global disease status (hygiene, other diseases prevalence) which impacts directly or indirectly the immune status of the population and the opportunities for transmission *i.e.* inter-host multiplication.

4. Conclusions

Since the model of epidemiologic transition, pathogens are thought to have shaped the history of humankind (Caldwell 2001). *Mycobacterium tuberculosis*, with its incredibly low mutation rate, pushes researchers working on phylogenetic inferences to merge data that clearly have a very diverse dynamics due to selection, but also due to demography, the impact of other diseases and mutation rate. Altogether, we have shown that no safe dating can be presently provided for tuberculosis emergence. This holds for major diversification events of the agent of tuberculosis such as Lineage 4 expansion. As a reminder, we may speak for major diversification event for this lineage as 9 sublineages have been consistently considered at an equivalent position in the taxonomy (L4.1 to L4.9). This major

diversification event could be due to Roman Empire expansion or Christian Crusades. In addition to the emergence date of tuberculosis, this 1,000 y uncertainty on this major diversification event in MTBC is also far from being solved.

Now that the reliability of sequencing is improving, that powerful methods exist and that the amount of data is exploding, turning towards synonymous mutations and implementing in rich phylodynamics model may help revisiting the possible history of tuberculosis. Yet, all assumptions underlying phylodynamics, including host demography, MTBC mutation rate, selection, etc. should be kept in mind to avoid exaggerated inferences. Independent confirmations are needed. They may come from other diseases, human history, etc. Altogether, archaeological research, with more bones to gnaw on, will definitively be of great help to strengthen or lower support of phylogenetic conclusions.

Funding

RZ has received funding from the European Union'S Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 899987. ALM benefits from a PhD fellowship from ANRS-MIE 2022-2, grant agreement n° 22476.

Acknowledgements

T. Giraud and E. Cambau are acknowledged for fruitful discussions on theoretical and practical grounds related to the subject.

Author contributions

Rima Zein-Eddine, Fiona Hak, Adrien Le Meur , Ccharlotte Genestet, Gaëtan Senelle performed data curation and formal analyses of data at the basis of this work. Guislaine Refrégier created the model. Christophe Guyeux, Oana Dumitrescu, Christophe Sola, Guislaine Refrégier participated in the funding acquisition. Guislaine Refrégier wrote the initial draft and performed the final editing. Christophe Guyeux, Adrien Le Meur, Christophe Sola, Rima Zein-Eddine participated in editing the manuscript. Adrien Le Meur created the figure.

5. References

1. Roser, M., E. Ortiz-Ospina, and H. Ritchie. *Life Expectancy*. 2019 [cited 2013 Jan. 2023]; Available from: https://ourworldindata.org/life-expectancy.
2. Caldwell, J.C., *Population health in transition.* Bull World Health Organ, 2001. **79**(2): p. 159-60.
3. Rascovan, N., et al., *Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline.* Cell, 2019. **176**(1-2): p. 295-305 e10.
4. Weill, F.X., et al., *Genomic history of the seventh pandemic of cholera in Africa.* Science, 2017. **358**(6364): p. 785-789.
5. Piarroux, R., et al., *Understanding the cholera epidemic, Haiti.* Emerg Infect Dis, 2011. **17**(7): p. 1161-8.
6. Gibbs, M.J. and A.J. Gibbs, *Molecular virology: was the 1918 pandemic caused by a bird flu?* Nature, 2006. **440**(7088): p. E8; discussion E9-10.
7. Antonovics, J., M.E. Hood, and C.H. Baker, *Molecular virology: was the 1918 flu avian in origin?* Nature, 2006. **440**(7088): p. E9; discussion E9-10.
8. Taubenberger, J.K., *Influenza hemagglutinin attachment to target cells: 'birds do it, we do it...'.* Future Virol, 2006. **1**(4): p. 415-418.
9. Crubezy, E., et al., *Identification of Mycobacferium DNA in an Egyptian Pott's disease of 5 400 years old. Identificdtion dilDN de Mycobacterium dims un mul de POE &-yptien de 5400 ans.* C. R. Acad. Sci. Paris, 1997. **321**.
10. Zink, A.R., et al., *Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping.* J Clin Microbiol, 2003. **41**(1): p. 359-67.
11. Wells, C., *Bones, bodies, and disease; evidence of disease and abnormality in early man*. Ancient peoples and places,. 1964, New York,: Praeger. 288 p.
12. Baker, O., et al., *Prehistory of human tuberculosis: Earliest evidence from the onset of animal husbandry in the Near East.* . Paléorient, 2017. **43**(2).
13. Spekker, O., et al., *A rare case of calvarial tuberculosis from the Avar Age (8th century CE) cemetery of Kaba-Bitozug (Hajdu-Bihar county, Hungary) - Pathogenesis and differential diagnostic aspects.* Tuberculosis (Edinb), 2022. **135**: p. 102226.
14. Sabin, S., et al., *A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex.* Genome Biol, 2020. **21**(1): p. 201.
15. Gutierrez, M.C., et al., *Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis.* PLoS Pathog, 2005. **1**(1): p. e5.
16. Ho, S.Y., et al., *Time-dependent rates of molecular evolution.* Mol Ecol, 2011. **20**(15): p. 3087-101.
17. Wirth, T., A. Meyer, and M. Achtman, *Deciphering host migrations and origins by means of their microbes.* Mol Ecol, 2005. **14**(11): p. 3289-306.
18. Kapur, V., T.S. Whittam, and J.M. Musser, *Is Mycobacterium tuberculosis 15,000 years old?* J Infect Dis, 1994. **170**(5): p. 1348-9.
19. Wirth, T., et al., *Origin, spread and demography of the Mycobacterium tuberculosis complex.* PLoS Pathog, 2008. **4**(9): p. e1000160.
20. Yang, Z., et al., *Codon-substitution models for heterogeneous selection pressure at amino acid sites.* Genetics, 2000. **155**(1): p. 431-49.
21. Bos, K.I., et al., *Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis.* Nature, 2014.
22. Bryant, J.M., et al., *Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data.* BMC Infect Dis, 2013. **13**: p. 110.

23. Roetzer, A., et al., *Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study.* PLoS Med, 2013. **10**(2): p. e1001387.
24. Walker, T.M., et al., *Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study.* Lancet Infect Dis, 2013. **13**(2): p. 137-46.
25. Comas, I., et al., *Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans.* Nat Genet, 2013. **45**(10): p. 1176-82.
26. Kuo, C.H. and H. Ochman, *Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria.* Biol Direct, 2009. **4**: p. 35.
27. Vos, M., et al., *Rates of Lateral Gene Transfer in Prokaryotes: High but Why?* Trends Microbiol, 2015. **23**(10): p. 598-605.
28. Chiner-Oms, A. and I. Comas, *Large genomics datasets shed light on the evolution of the Mycobacterium tuberculosis complex.* Infect Genet Evol, 2019. **72**: p. 10-15.
29. Osorio, N.S., et al., *Evidence for diversifying selection in a set of Mycobacterium tuberculosis genes in response to antibiotic- and nonantibiotic-related pressure.* Mol Biol Evol, 2013. **30**(6): p. 1326-36.
30. Tantivitayakul, P., et al., *Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in Mycobacterium tuberculosis.* Sci Rep, 2020. **10**(1): p. 8024.
31. Genestet, C., et al., *Mycobacterium tuberculosis genetic features associated with pulmonary tuberculosis severity.* Int J Infect Dis, 2022. **125**: p. 74-83.
32. Chiner-Oms, A., et al., *Gene evolutionary trajectories in Mycobacterium tuberculosis reveal temporal signs of selection.* Proc Natl Acad Sci U S A, 2022. **119**(17): p. e2113600119.
33. Kerner, G., et al., *Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years.* Am J Hum Genet, 2021. **108**(3): p. 517-524.
34. Smith, T.M., et al., *Rapid adaptation of a complex trait during experimental evolution of Mycobacterium tuberculosis.* Elife, 2022. **11**.
35. Moreno-Molina, M., et al., *Genomic analyses of Mycobacterium tuberculosis from human lung resections reveal a high frequency of polyclonal infections.* Nat Commun, 2021. **12**(1): p. 2716.
36. Liu, X., et al., *Evidence for recombination in Mycobacterium tuberculosis.* J Bacteriol, 2006. **188**(23): p. 8169-77.
37. Menardo, F., et al., *The molecular clock of Mycobacterium tuberculosis.* PLoS Pathog, 2019. **15**(9): p. e1008067.
38. WHO, *Global Tuberculosis Control : Surveillance, Planning, Financing.* 2004, WHO, Geneva, Switzerland.
39. Mathema, B., et al., *Drivers of Tuberculosis Transmission.* J Infect Dis, 2017. **216**(suppl_6): p. S644-S653.
40. WHO, *Global tuberculosis report 2022*. 2022, World Helath Organization: Geneva. p. 1-68.
41. Hambridge, T., et al., *Mycobacterium leprae transmission characteristics during the declining stages of leprosy incidence: A systematic review.* PLoS Negl Trop Dis, 2021. **15**(5): p. e0009436.
42. Menardo, F., *Understanding drivers of phylogenetic clustering and terminal branch lengths distribution in epidemics of Mycobacterium tuberculosis.* Elife, 2022. **11**.
43. Castaneda-Garcia, A., et al., *A non-canonical mismatch repair pathway in prokaryotes.* Nat Commun, 2017. **8**: p. 14246.
44. Kunkel, T.A. and K. Bebenek, *DNA replication fidelity.* Annu Rev Biochem, 2000. **69**: p. 497-529.
45. Rock, J.M., et al., *DNA replication fidelity in Mycobacterium tuberculosis is mediated by an ancestral prokaryotic proofreader.* Nat Genet, 2015. **47**(6): p. 677-81.
46. Ford, C.B., et al., *Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection.* Nat Genet, 2011. **43**(5): p. 482-6.

47. Hughes, A.L., R. Friedman, and M. Murray, *Genomewide Pattern of Synonymous Nucleotide Substitution in Two Complete Genomes of Mycobacterium tuberculosis.* Emerg Infect Dis, 2002. **8**(11): p. 1342-6.
48. Jee, J., et al., *Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing.* Nature, 2016. **534**(7609): p. 693-6.

Figure