



Peer Community In Evolutionary Biology

How the tubercle bacillus got its genome: modernising, modelling, and making sense of the stories we tell

B. Jesse Shapiro based on peer reviews by 2 anonymous reviewers

Christoph Stritt, Sebastien Gagneux (2023) How do monomorphic bacteria evolve? The *Mycobacterium tuberculosis* complex and the awkward population genetics of extreme clonality. *EcoEvoRxiv*, ver. 3, peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.32942/X2GW2P>

Submitted: 16 December 2022, Recommended: 30 June 2023

Cite this recommendation as:

Shapiro, B. (2023) How the tubercle bacillus got its genome: modernising, modelling, and making sense of the stories we tell. *Peer Community in Evolutionary Biology*, 100644. [10.24072/pci.evolbiol.100644](https://doi.org/10.24072/pci.evolbiol.100644)

Published: 30 June 2023

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

In this instructive review, Stritt and Gagneux offer a balanced perspective on the evolutionary forces shaping *Mycobacterium tuberculosis* and make the case that our instinct for storytelling be balanced with quantitative models. *M. tuberculosis* is perhaps the best-known clonal bacterial pathogen – evolving largely in the absence of horizontal gene transfer. Its genome is full of puzzling patterns, including much higher GC content than most intracellular pathogens (which suggests efficient selection to resist AT-skewed mutational bias) but a very high ratio of nonsynonymous to synonymous substitution rates ($dN/dS \sim 0.5$, typically interpreted as weak selection against deleterious amino acid changes).

The authors offer alternative explanations for these patterns, framing the question: is *M. tuberculosis* evolution shaped mainly by drift or by efficient selection? They propose that this question can only be answered by accounting for the pathogen's extreme clonality. A clonal lifestyle can have its advantages, for example when adaptations must arise in a particular order (Kondrashov and Kondrashov 2001). An important disadvantage highlighted by the authors are linkage effects: without recombination to shuffle them up, beneficial mutations are linked to deleterious mutations in the same genome (hitchhiking) and purging deleterious mutations also purges neutral diversity across the genome (background selection). The authors propose the latter – efficient purifying selection and strong linkage – as an explanation for the low genetic diversity observed in *M. tuberculosis*. This is of course not exclusive of other related explanations, such as clonal interference (Gerrish and Lenski 1998). They also champion the use of forward evolutionary simulations (Haller and Messer 2019) to model the interplay between selection, recombination, and demography as a powerful alternative to traditional backward coalescent models.

At times, Stritt and Gagneux are pessimistic about our existing methods – arguing that dN/dS and homoplasies “tell us little about the frequency and strength of selection.” Even though I favour a more optimistic view, I fully agree that our traditional population genetic metrics are sensitive to a slew of different deviations from a standard neutral evolution model and must be interpreted with caution. As I and others have argued, the extent of recombination (measured as the amount of linkage in a genome) is a key factor in determining how best to test for natural selection (Shapiro et al. 2009) and to conduct genotype-phenotype association studies (Chen and Shapiro 2021) in microbes. While this article is focused on the well-studied *M. tuberculosis* complex, there are many parallels with other clonal bacteria, including pathogens and symbionts. Whatever your favourite bug, we must all be careful to make sure the stories we tell about them are not “just so tales” but are supported, to the extent possible, by data and quantitative models.

References:

- Chen, Peter E., and B. Jesse Shapiro. 2021. “Classic Genome-Wide Association Methods Are Unlikely to Identify Causal Variants in Strongly Clonal Microbial Populations.” bioRxiv.
<https://doi.org/10.1101/2021.06.30.450606>
- Gerrish, P. J., and R. E. Lenski. 1998. “The Fate of Competing Beneficial Mutations in an Asexual Population.” *Genetica* 102-103 (1-6): 127-44.
<https://doi.org/10.1023/A:1017067816551>
- Haller, Benjamin C., and Philipp W. Messer. 2019. “SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model.” *Molecular Biology and Evolution* 36 (3): 632-37.
<https://doi.org/10.1093/molbev/msy228>
- Kondrashov, F. A., and A. S. Kondrashov. 2001. “Multidimensional Epistasis and the Disadvantage of Sex.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (21): 12089-92.
<https://doi.org/10.1073/pnas.211214298>
- Shapiro, B. Jesse, Lawrence A. David, Jonathan Friedman, and Eric J. Alm. 2009. “Looking for Darwin’s Footprints in the Microbial World.” *Trends in Microbiology* 17 (5): 196-204.
<https://doi.org/10.1016/j.tim.2009.02.002>
- Stritt, C., Gagneux, S. (2023). How do monomorphic bacteria evolve? The Mycobacterium tuberculosis complex and the awkward population genetics of extreme clonality. EcoEvoRxiv, ver.3 peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.32942/X2GW2P>

Reviews

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.32942/X2GW2P>

Version of the preprint: 1

Authors’ reply, 16 June 2023

[Download author’s reply](#)

Decision by **B. Jesse Shapiro**, posted 18 April 2023, validated 19 April 2023

We invite a revision of your preprint

Thank you for submitting your preprint, and I apologize for the delay in obtaining reviews.

Both I and the two reviewers agreed that you address a very interesting topic and that the preprint makes a useful contribution to evolutionary microbiology. The reviewers agreed that the manuscript could benefit from some clarification and shortening of certain sections to better highlight the key points, and to ensure the manuscript is accessible and useful to as broad an audience as possible.

I hope you will carefully consider these constructive comments and I look forward to seeing your revised manuscript.

Reviewed by anonymous reviewer 2, 17 February 2023

This is a very interesting and relevant review about MTBC evolution. The authors clearly state the assumptions of population genetics models and which ones are met by the known biology and evolution of MTBC.

Nevertheless, I think the flow and structure of the article can be improved.

For example, linkage and linked selection is mentioned in the introduction, but in the section on positive selection it is not clear which consequences this has for the inference of selection.

I also find it odd to start with recombination instead of mutation since the latter is the basic process generating diversity and all other processes act on this diversity. Of course, the rationale of the authors can be different, but then that must be clear in the flow of the text.

It is also unclear why genetic drift and purifying selection are addressed in the same section (and positive selection in a different one) although drift interferes with selection independent of the direction of selection. Also, this results in dN/dS being introduced twice (lines 368 and 616).

The manuscript is also a bit lengthy and I would suggest to shorten sections that are not directly relevant for MTBC or for clonal evolution, e.g., the section on DCT starting on line 112.

The translation of the per-site mutation rate into the per-year mutation rate is very simplistic (line 220). Does this assume that all mutations are neutral and how are dynamics in the population included? The authors correctly distinguish between the mutation rate and the molecular clock rate earlier in the manuscript (line 191), so they should also correctly distinguish them throughout the manuscript.

It is unclear which conclusions the authors want to state in the paragraph starting at line 439. They search for an explanation for the high dN/dS and invoke selection at synonymous sites. But then it is shown that there is evidence for positive selection at synonymous sites, which would result in low dN/dS and not high ones.

The discussion section contains a highly relevant appeal for including proper simulations in data analysis. Nevertheless an actual discussion is missing. I suggest to add a discussion on the consequences of clonal evolution on MTBC genome evolution, for example on their genomic architecture, on the efficiency of selection, or on linkage and the distribution of fitness effects (what about epistasis?). Such a conclusion is really expected by the reader since the authors state in the introduction "In this review, we present the main hypotheses about what drives the evolution of the MTBC, and how they have been arrived at." (line 91) So, what are these main hypotheses?

Further comments:

line 239: It is unclear what the authors want to say with the sentence starting at line 239. How is the AT bias reflected by stress-induced mutagenesis and how does this relate to the GC rich genome?

line 605: I would suggest to remove "human" from the sentence to focus on MTBC migration instead. An explanation might be added that this is driven by human migration.

line 691: It would be easier to follow, if the purpose of the simulation was mentioned, before the simulation details are described.

Fig 4c,d: What are the solid and dashed boxes?

Line 710: Where can Fig. B1 be found?

Reviewed by anonymous reviewer 1, 17 April 2023

[Download the review](#)