Rapid adaptation of a complex trait during experimental evolution of *Mycobacterium tuberculosis*

Tracy M. Smith[1*], Madison A. Youngblom[2*], John F. Kernien[2], Mohamed A. Mohamed[2], Lindsey L. Bohr[2], Tatum D. Mortimer[3], Mary B. O'Neill[4], Caitlin S. Pepperell[2,5]

[*] These authors contributed equally to this work

[1] Lodo Therapeutics Corporation, New York City, NY, United States
[2] Department of Medical Microbiology and Immunology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States
[3] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[4] Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, CNRS, Paris 75015, France
[5] Department of Medicine (Infectious Diseases), School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*M. tb*), is a leading cause of death due to infectious disease. TB is not traditionally associated with biofilms, but *M. tb* biofilms are linked with drug and immune tolerance and there is increasing recognition of their potential role in the recalcitrance of TB infections. Here we used *M. tb* experimental evolution to investigate this complex phenotype and identify candidate loci controlling biofilm formation. We identified novel candidate loci, adding to our understanding of the genetic architecture underlying *M. tb* biofilm development. Under selective pressure to grow as a biofilm, regulatory mutations rapidly swept to fixation and were associated with changes in multiple traits including extracellular matrix production, cell size, and growth rate. Genetic and phenotypic paths to enhanced biofilm growth varied according to the genetic background of the parent strain, suggesting that epistatic interactions are important in *M. tb* adaptation to changing environments.

Introduction

In 2019 an estimated 10 million people fell ill with tuberculosis (TB), and one quarter of the world's population is estimated to be infected with its causative agent *Mycobacterium tuberculosis* (*M. tb*) (Global Tuberculosis Report 2019, WHO). New strategies for diagnosis, treatment and control of TB are urgently needed. From an evolutionary perspective, *M. tb* stands out among bacterial pathogens for its strict association with human hosts, limited genetic diversity and clonal evolution (Eldholm and Balloux, 2016). We might expect these features to constrain adaptation of *M. tb*, yet TB remains a challenging infection to treat due to the bacterium's ability to persist in the face of antibiotic and immune pressure and to acquire novel drug resistances. In order to better treat and control TB, we need to understand the sources of *M. tb*'s robustness and to identify its vulnerabilities. Experimental evolution is a powerful tool for illuminating these strengths and vulnerabilities and has led to important insights into fundamental processes guiding microbial adaptation.

Biofilms are increasingly recognized as a relevant growth form for bacteria in their natural environments (Costerton et al., 1999). TB is not traditionally thought of as a biofilm infection. However, *M. tb* cells spontaneously aggregate and elucidate extracellular matrix when grown *in vitro* (Dubos and Davis, 1946; Bacon et al., 2014), suggesting they are naturally inclined to grow as biofilms. Autopsy studies have long identified *M. tb* aggregates in human tissues during TB infection (Canetti, 1955; Nyka, 1977, 1967, 1963; Nyka and O'Neill, 1970). More recent research has demonstrated specific biomarkers of *M. tb* biofilms in human autopsy specimens and animal models of TB (Chakraborty et al., 2021). The presence of *M. tb* biofilms during TB infection is of major practical significance as growth within a biofilm allows *M. tb* cells to survive otherwise lethal concentrations of antibiotics and to evade immune responses (Ojha et al., 2008; Ackart et al., 2014; Trivedi et al., 2016; Chakraborty et al., 2021). Identifying the mechanisms of biofilm development by *M. tb* can thus aid the development of new, more effective therapies for TB (Wang et al., 2013; Ackart et al., 2014; Richards et al., 2019; Chakraborty et al., 2021).

The genetic determinants of *M. tb* biofilm formation have been investigated with candidate gene approaches and phenotypic characterization of knockout, knockdown and overexpression mutants (Ojha et al., 2008; Pang et al., 2012; Sambandan et al., 2013; Wolff et al., 2015; Rastogi et al., 2017; Yang et al., 2017; Richards et al., 2019; Hegde, 2020, p. 20; Bharti et al., 2021; Chakraborty et al., 2021). Here, we use a complementary approach based on serial passaging of *M. tb* clinical isolates under selective pressure to grow as a biofilm. This approach has the advantage of maintaining the integrity of complex networks of genes and their regulators while enabling discovery of subtle genetic

changes with an impact on biofilm phenotype. It is also unbiased with respect to the choice of candidate loci. During experimental evolution of six closely related *M. tb* isolates passaged over months to years we find: 1) Rapid adaptation in response to selection imposed in our system, with development of more robust biofilm growth in all strains, 2) Changes in a range of *M. tb* phenotypes in association with few, presumably pleiotropic mutations, 3) Predominant impacts on gene dosage among mutations that emerged during the experiment, 4) Implication of *M. tb* loci not previously known to be involved in biofilm development, and 5) Apparent effects of strain genetic background in shaping the adaptive path to the phenotype under selection.

Results

Sample

Pellicles are a specific type of biofilm in which bacteria form aggregates at air-liquid interfaces (Kobayashi, 2007); *M. tb* has been shown to form pellicle biofilms *in vitro* (Kerns et al., 2014). For our study, six closely related isolates of *M. tb*, from three sub-clades of the Euro-American lineage L4 (4.9, 4.4.1.2, and 4.4.1.1 shown in Figure 1), were grown as pellicles following a published protocol (Kulka et al., 2012). We passaged these six populations in pellicle form every 5-7 weeks as described in the Methods (Figure 8). Each pellicle population was passaged at least eight times over a period of two years.

Phenotypic changes

At each passage we photographed the pellicle and described its growth according to the following criteria: proportion of liquid surface covered, presence of climbing (attachment to and growth up the sides of the flask), thickness of growth and continuity of growth (versus discontinuous patches). Although the *M. tb* strains were closely related (i.e. separated by 100s of SNPs), differences in biofilm phenotype were evident prior to passaging (Figure 1). During the initial few passages, phenotypes changed for all strains and then stabilized between passages 2 and 8 (Figure 1). We performed extended passaging for four strains, which were carried out to passage 16 or 20; we did not observe any further phenotypic changes during this extended passaging (Supplementary Figure 1).

Over the course of the experiment all populations evolved more robust biofilms characterized by an increase in surface coverage and thicker, reticulating growth (Figure 1). MT72 was an exception, as it developed a typical confluent biofilm after two passages, but then evolved discontinuous, thick growth

covering only a small fraction of the surface. This contrasts with the other evolved pellicles, which cover the entire surface of the liquid, climb up the side of the flask, and are more confluent (Figure 1).

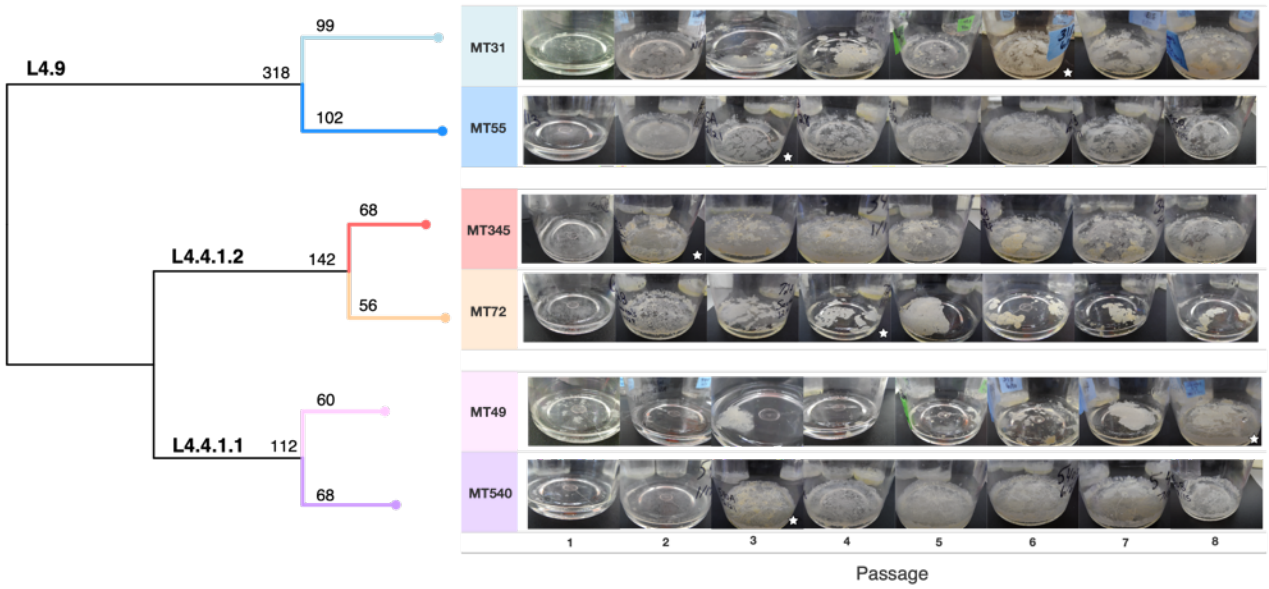

Figure 1: Whole genome sequence phylogeny of ancestral populations used for passaging and photos of pellicles throughout the experiment. Populations fall into three distinct lineages with SNP distances given for each sub-lineage and strain. Pellicle photos show rapid change in phenotype, followed by stabilization through the rest of the experiment. Passage at which point we determined the pellicle phenotype to be stabilized is indicated by a star for each population. Photos of pellicles were taken after 5-7 weeks of growth.

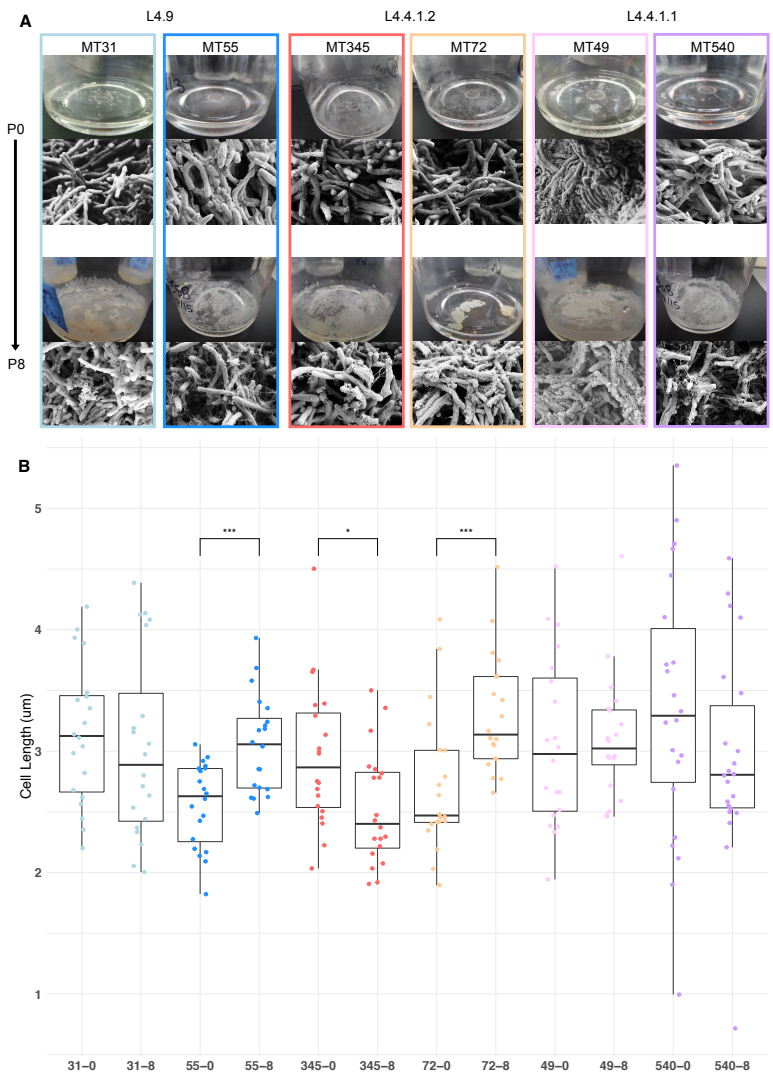

Figure 2: SEM images of ancestral and evolved biofilms show changes in extracellular matrix (A) and cell length (B). SEMs were taken after 5 weeks of biofilm growth, shown alongside photo of pellicle from that passage. Cell lengths were measured across two biological replicates for each strain, except MT540-0 for which only one replicate was available. Each dot represents a single cell length measurement. Significant differences in cell length between ancestral and evolved pairs shown across the top of panel B as calculated by Mann-Whitney U test (with Benjamini-Hochberg correction). P-value legend: *: <0.05, **: <0.01, ***: <0.001.

To further characterize *M. tb* biofilm phenotypes we performed scanning electron microscopy (SEM) of biofilm samples from ancestral strains and paired evolved strains that had undergone eight rounds of passaging. The appearance of extracellular matrix (ECM) was variable among strains at baseline, but all strains exhibited increased production of ECM after serial passaging (Figure 2). We observed changes in ECM appearance including increased webbing (MT55, MT540) and increased production of globules (MT31, MT345, MT72, MT49) (Figure 2). Bacterial cell shape has been documented to have direct fitness consequences in many species (Yang et al., 2016). We measured bacilli length of ancestral and evolved strains and found that three strains (MT55, MT72, MT345) had significant changes in cell length (Figure 2). The direction of the change varied among strains: MT55 and MT72 evolved longer bacilli whereas MT345 evolved shorter bacilli (Figure 2).

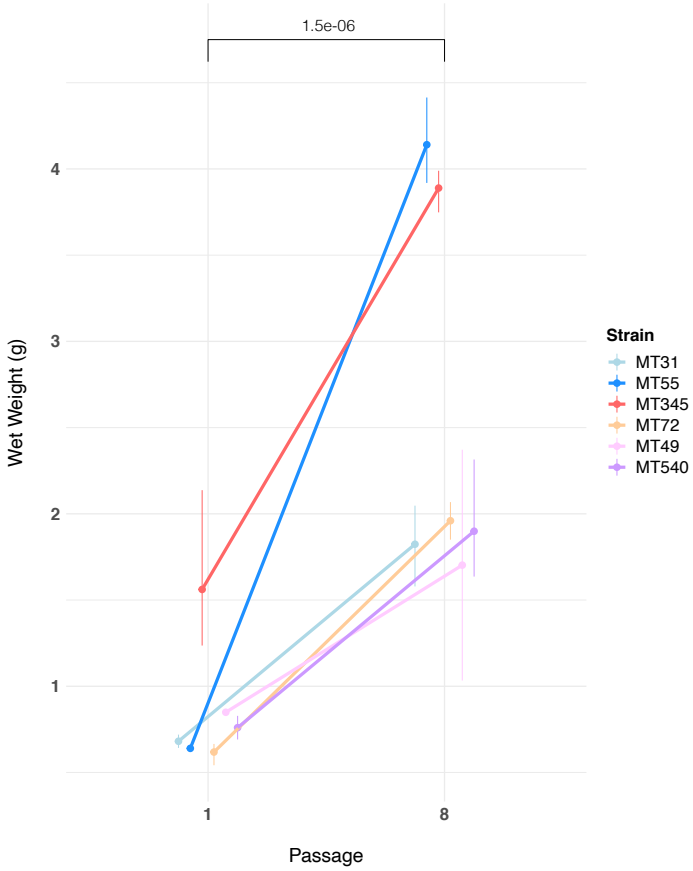

Figure 3: Pellicle wet weights measured in triplicate after 5 weeks of growth, for ancestral populations and populations after eight passages. Pellicle wet weights increased for all genetic backgrounds following passaging (Mann-Whitney U test, p-value = 1.5e-06). The magnitude of this change varied, with two genetic backgrounds (MT55 & MT345) showing relatively dramatic increases in pellicle weight. We also observed a trend toward increased variability in wet weights after passaging as a pellicle. Error bars represent range of wet weights across replicates.

Wet weights

To obtain another quantitative measure of phenotypic adaptation to biofilm growth, we developed a protocol for measuring the wet weight of a pellicle biofilm. Wet weights of ancestral biofilms were similar except for MT345, which was heavier in keeping with its confluent morphotype (Figures 1 and 3). We observed a significant (Mann-Whitney U test, p-value = 1.5e-06) increase in wet weight for all populations after eight passages (Figure 3). MT55 and MT345 had particularly dramatic increases in pellicle biomass. We also observed a trend towards increased variability of wet weights between replicate cultures of the evolved biofilms (Figure 3).

Planktonic growth trade-offs

In order to investigate whether adaptation to pellicle growth involved trade-offs with fitness in planktonic culture, we compared planktonic growth rates of ancestral and evolved populations. Some evolved strains grew faster in planktonic culture following pellicle passaging (MT31, MT345, MT49) whereas other strains exhibited similar growth rates – if slightly different kinetics – when compared to their ancestral strains (MT55, MT72; Figure 4). One evolved strain (MT540) had a slower growth rate than its ancestor (Figure 4C; E).

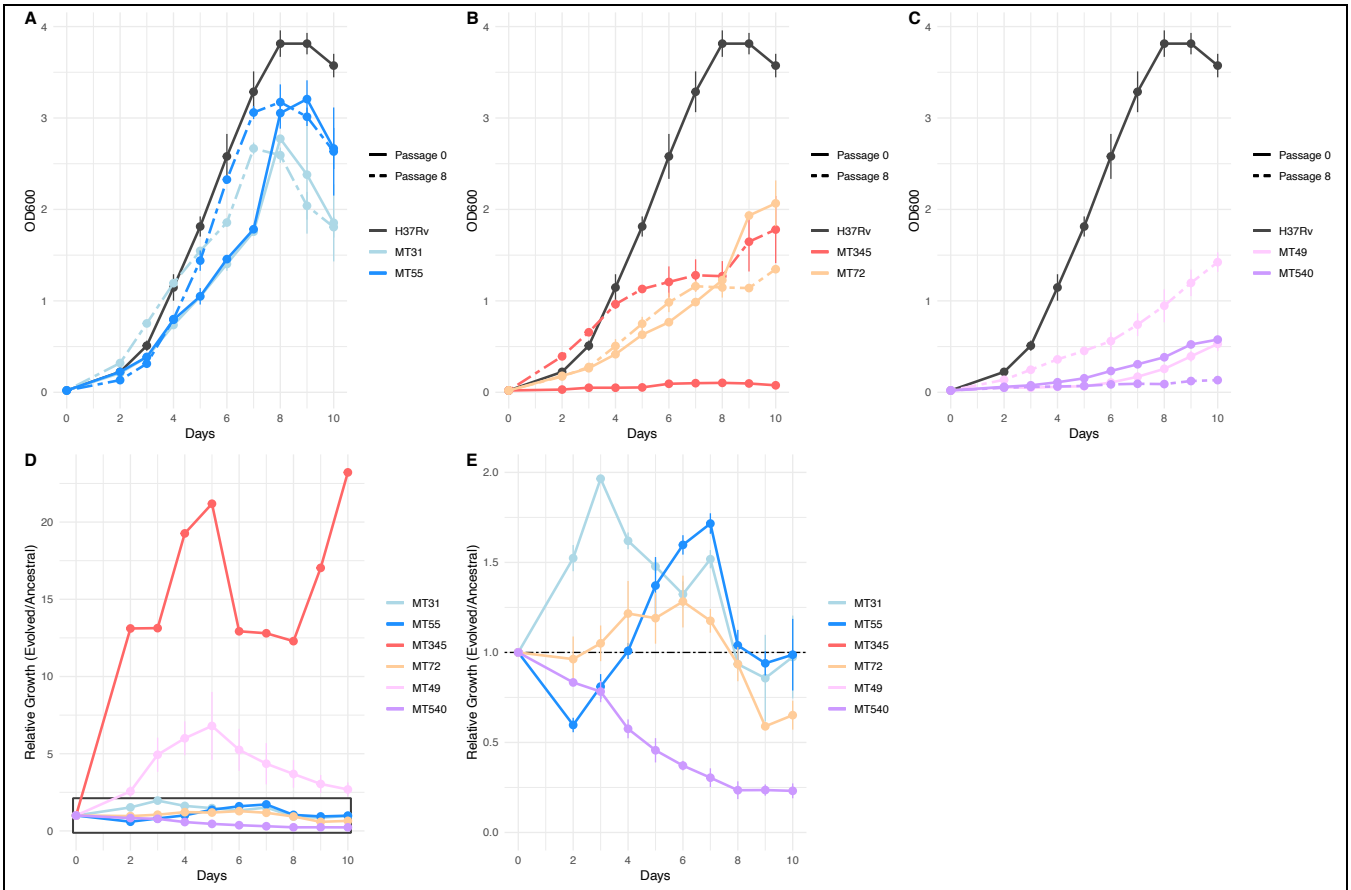

Figure 4: Planktonic growth (A, B, C) and relative fitness (D, E) curves comparing the growth of strains after 8 pellicle passages, to the ancestral strains. Growth curves were performed in triplicate, and the mean OD_{600} value is plotted with error bars representing +/- 1 standard deviation. Relative fitness was calculated at each timepoint as OD_{600} Passage 8 / OD_{600} Passage 0. Panel E shows boxed region from panel D in more detail.

Candidate loci involved in biofilm formation

Genomic DNA (gDNA) was extracted from whole biofilm communities for (pooled) sequencing every 4 passages (except MT31 and MT49 which were first sequenced at passage 8). To identify loci potentially responsible for the observed changes in pellicle phenotype, we identified mutations that rose to fixation (variant at 100% in the population) or disappeared (variant at 0%, reference allele at 100%) over the course of our experiment. After filtering and manual curation of the results, we identified 14 single nucleotide polymorphisms (SNPs), 2 duplications and 2 deletions that met these criteria (Table 1). The starting frequency of these alleles varied from 0-55% and fixation occurred by the 8th passage for all but two of these variants (position 4,115,505 in MT31 and position 1,416,232 in MT72) (Table 1). Candidate genes are annotated with a range of functions with regulation (*regX3, phoP, embR, Rv2488c*) and lipid metabolism (*Rv2186c, mgtA, fadE22*) being the most common (Table 1). A search

of ~40,000 isolates from a global dataset (see Methods), revealed these mutations to be exceedingly rare or completely absent from the database (Table 1). The most commonly identified SNP was at position 3,690,947 upstream of *lpdA* (identified in our study in strain MT49), at a frequency of 0.0015 in this dataset (Table 1). One caveat is that genes in repetitive families (PE, PPE, PE-PGRS) were excluded from our analysis due to difficulties accurately resolving these regions with short-read sequencing data. It is possible that variants contributing to enhanced biofilm formation lie undetected in these regions.

Table 1: Small number of variants which became fixed (>95% or <5% frequency) throughout the course of the experiment. Including both SNPs and INDELs identified by Popoolation2, Breseq and/or Pilon. Highlighted in bold are two instances of convergent adaptation we observed in the L4.9 and L4.4.1.1 sub-lineages.

Fixed Mutations by Strain								
Strain	Position*	Proportion at passage 0	Fixed at Passage	Type	Mutation	Gene	Product	Proportion in Global Dataset
L4.9								
MT31	2279599	40%	8†	Synonymous	D157D (GAT-GAC)	acg	conserved hypothetical, possible nitroreductase	0
MT31	2447977	39%	8	Synonymous	I6I (ATC-ATT)	Rv2186c	conserved hypothetical	0.000025
MT31	**3553976**	**–**	**8**	**~120 kb DUP**	**–**	**Rv3188 [. . .] Rv3296**	**–**	**–**
MT31	4115505	49%	20†	Synonymous	P116P (CCG-CCT)	nth	endonuclease III	0.00010
MT55	**3589531**	**–**	**4**	**~120 kb DUP**	**–**	**Rv3212 [. . .] Rv3324c**	**–**	**–**
L4.4.1.2								
MT345	580968	1.5%	4	Non-synonymous	M54V (ATG-GTG)	regX3	two component sensory transduction protein RegX	0
MT345	852317	<1%	4	Non-synonymous	R237L (CGC-CTC)	phoP	two component system response transcriptional positive regulator PhoP	0
MT345	3876836	0%	8	Intergenic	-14/+54 (C-A)	truA / rplQ	tRNA pseudouridine synthase A / 50S ribosomal protein L17	0
MT72	499768	5.7%	4	1 bp DEL - frameshift	pos. 56/654	renU (mutT3)	nudix hydrolase	0
MT72	652668	9.1%	4	Non-synonymous	R30S (CGC-AGC)	Rv0561c	oxidoreductase	0.000025
MT72	1416232	30%	16†	Synonymous	C372C (TGT-TGC)	embR	transcriptional regulator	0
MT72	4168394	0%	4	1 bp DEL-noncoding	pos. 50/86	serV	tRNA-Ser (GGA)	0
L4.4.1.1								
MT49	1939638	55%	8†	Non-synonymous	P14T (CCG-ACG)	cmk	cytidylate kinase	0
MT49	2799039	0%	8	Synonymous	G614G (GGC-GGT)	Rv2488c	LuxR family transcriptional regulator	0.000025
MT49	**3690947**	**0%**	**8**	**Intergenic**	**-9/-194 (A-G)**	**lpdA / Rv3304**	**NAD(P)H quinone reductase / hypothetical**	**0.0015**
MT540	648780	6.0%	4	Non-synonymous	F82S (TTC-TCC)	mgtA	GDP-mannose-dependent alpha-mannosyltransferase	0
MT540	3423753	0%	4	Non-synonymous	D559Y (GAT-TAT)	fadE22	acyl-CoA dehydrogenase FadE22	0
MT540	**3690952**	**0%**	**8‡**	**Intergenic**	**-14/-189 (G-C)**	**lpdA / Rv3304**	**NAD(P)H quinone reductase / hypothetical**	**0.00010**

* Coordinates in H37Rv genome
† Mutation fell out of population, fixed = 0%
‡ At 85% frequency at passage 8

We observed two instances of convergent adaptation, i.e. the same locus being subject to repeated mutation. Interestingly each pair of convergent mutations appeared in the same sub-lineage. A large (~120kb) duplication fixed in populations MT31 and MT55, which both belong to sub-lineage L4.9. The

second convergent mutation, an intergenic SNP upstream of NAD(P)H quinone reductase *lpdA,* within a transcription factor binding site (TFBS) and a non-coding RNA, occurred in MT49 and MT540 that both belong to sub-lineage L4.4.1.1 (Table 1).

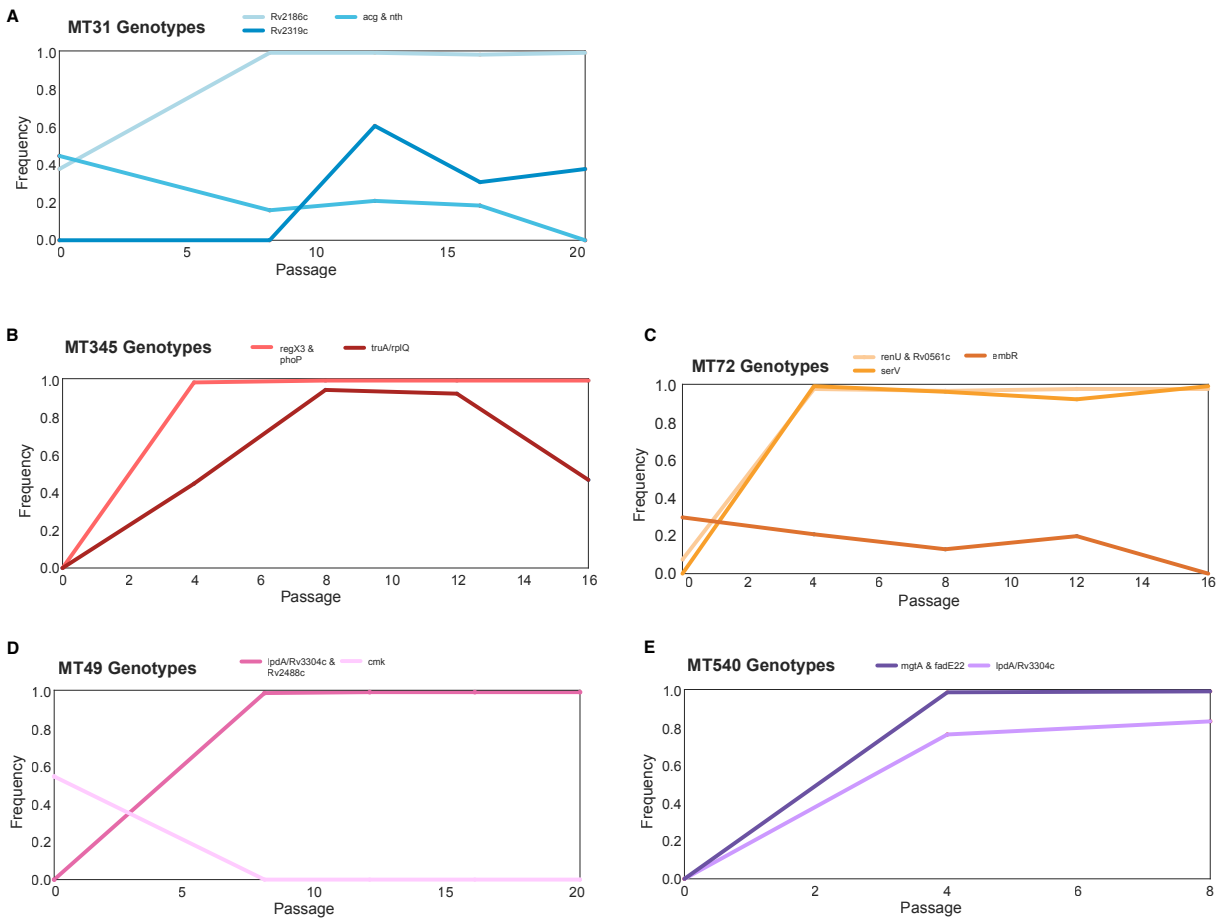

Figure 5: Trajectories of mutations that changed >=30% over the course of the experiment show just how few mutations were acquired. The mutations which became fixed almost always did so in the first or second sequenced passage, and we identified very few mutations which significantly declined in frequency. Mutation frequency data was calculated with Popoolation2 and plots were made using Lolipop (see Methods).

Allele frequency dynamics for all variants that changed >30% over the course of the experiment are shown in Figure 5. In most populations (MT31, MT345, MT72, MT540) we observed secondary mutations arising on the background of the first fixed mutations (Figure 5). Some secondary mutations reached high frequencies (>95%) but did not remain at high frequency for the duration of the experiment (Figure 5B). We identified a single mutation (Figure 5A, P286L *Rv2319*) with a moderate change in frequency (>30%) that did not become fixed. Allele frequency trajectories were otherwise

characterized by rare variants that remained rare. Few changes in allele frequency were observed for MT55: aside from the large duplication that reached fixation we did not identify any other mutations in this strain with >= 30% change in frequency. We calculated genome-wide measures of diversity (Tajima's D, pi and theta) for sequence data at each timepoint (Supplementary Table 1). Our values concur with previously estimated values for *M. tb* (O'Neill et al., 2015) and are consistent with purifying selection.

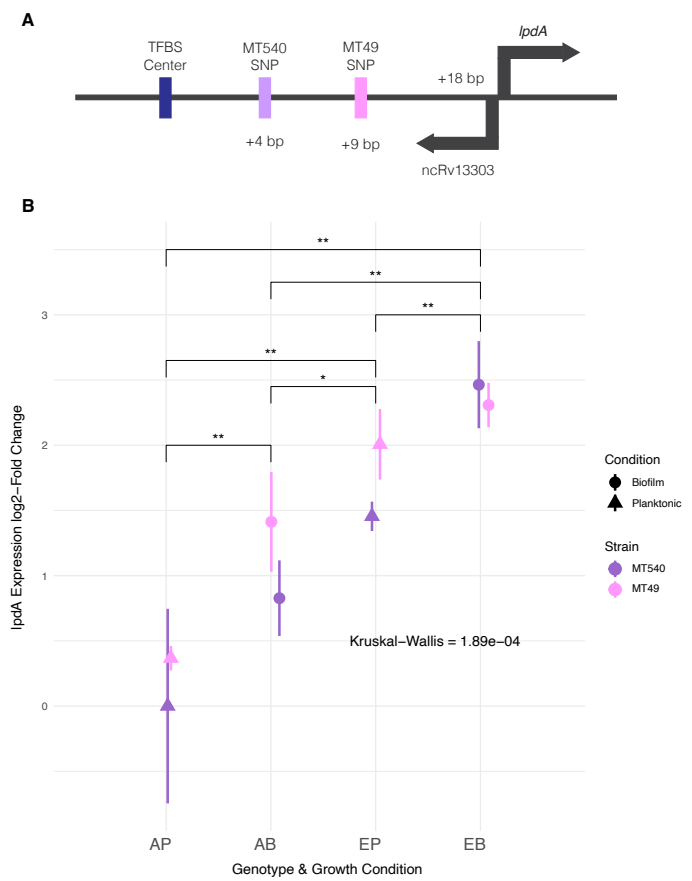

Figure 6: (A) Diagram of convergent adaptation in L4.4.1.1 strains MT540 and MT49 in the region upstream of *lpdA* within a TFBS and ncRNA. (B) Expression of *lpdA* in ancestral and evolved strains as measured by qPCR, showing significant increase in expression (Kruskal-Wallis test, p-value = 1.89e-04, pairwise comparisons performed with Mann-Whitney U test with Benjamini-Hochberg correction, p-value <0.05) in evolved (E) strains when compared to ancestral (A) strains, as well as during biofilm (B) growth when compared to planktonic (P) growth. P-value legend: *: <0.05, **: <0.01.

Convergent adaptation affecting expression of lpdA

We identified convergent adaptation in MT49 and MT540 (both of lineage L4.4.1.1) in the region upstream of NAD(P)H quinone reductase *lpdA* (Table 1). These intergenic SNPs are five base pairs apart and lie within the transcription factor binding site (TFBS) for *Rv1719,* which is known to downregulate *lpdA* (Rustad et al., 2014). We hypothesized that a SNP within the TFBS would affect TF binding and therefore expression of *lpdA.* We sought to quantify *lpdA* expression by qPCR in ancestral populations and compare that to expression in evolved populations containing the SNP of interest. Evolved populations had significantly higher (p-value <0.05, Mann-Whitney U-test) expression of *lpdA* (Figure 6). Additionally, we observed significantly increased expression of *lpdA* during biofilm growth

when compared to planktonic growth, suggesting a possible role for this gene in biofilm formation (Figure 6). In addition to being within a TFBS affecting expression of *lpdA*, these SNPs also lie within a non-coding RNA (ncRNA) that is expressed from the opposite strand of *lpdA* (Figure 6). This ncRNA was identified by RNA-sequencing (Arnvig et al., 2011; Gerrick et al., 2018) but has yet to be characterized.

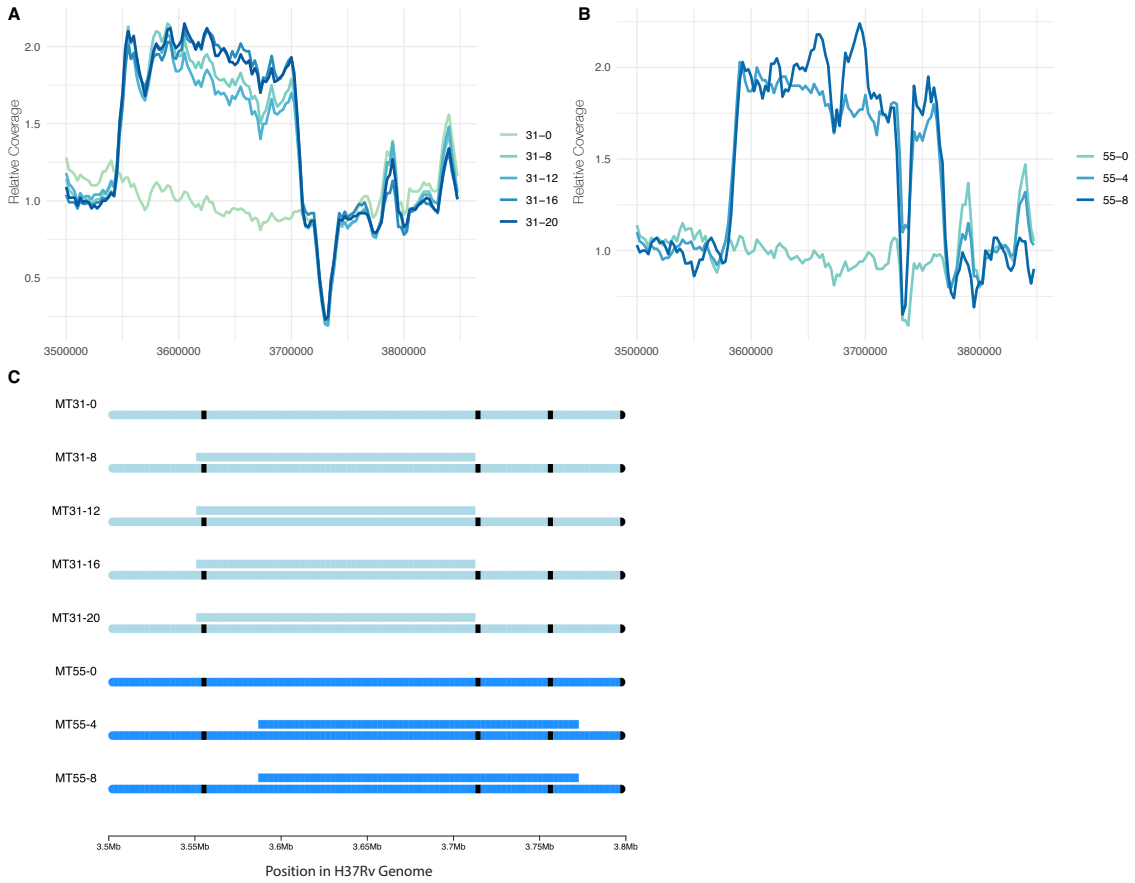

Figure 7: Convergent evolution of a large duplication in L4.9 strains MT31 and MT55, within the same genomic region flanked by mycobacterial IS6110 insertion sequences. Sliding window coverage plots show the substantial increase in coverage in this region, and a chromosome plot shows the duplications' coordinates and overall stability in both populations over many passages.

Convergent evolution of large duplication

We identified the evolution of a large (~120kb) duplication in both of the strains from lineage L4.9 (MT31 and MT55, Table 1). The duplications were initially called using Pilon, and in order to confirm its coordinates, as well as its absence in the ancestral population, we made sliding window coverage plots in this region and confirmed ~ 2X coverage in this region, indicating the presence of a duplication (Figure 7). The coordinates of the duplication appear to be slightly different in the two populations with

the duplication in MT55 starting about 3kb after the duplication in MT31 (Figure 7). We determined the duplication to be fixed after the first sequenced passage in each population (passage 4 and 8 for MT55 and MT31, respectively). After fixation, the duplication appears to be stable in both populations, as far out as 20 passages in MT31, as we do not see any reduction in coverage across or within the duplicated region at these later passages (Figure 7). There are a total of 172 genes within the duplicated regions in MT31 and MT55 (*metU* through *folD*). To identify a possible functional enrichment within these genes we performed a GO term enrichment analysis and compared our results to the results from randomly selected groups of contiguous genes from throughout the genome (see Methods). While the initial analysis identified significant enrichment of genes involved in nucleotide metabolism within the duplication (Supplementary Table 2), this enrichment does not appear significant when compared with our randomly selected gene sets (Supplementary Figure 2).

Regulatory pathways involved in biofilm formation

Table 2: Overlap between our genes of interest (GOI) and the transcription factors (TF) that regulate their expression. P-value calculated based on overlap of GOI with genes significantly differentially expressed by TF.

Regulatory Inputs and Outputs for GOIs

	TF	TF Name	Total Genes*	GOI Overlap*	P-value	GOI
Regulators Shared Among GOIs						
ChIP-Seq	Rv2989	–	94	2	4.09×10^{-2}	*lpdA, phoP*
Transcription Factor Overexpression	Rv3681c	*whiB4*	3	1	1.04×10^{-2}	*acg, embR, phoP**
	Rv1267c	*embR*	5	1	1.73×10^{-2}	*acg, embR*, phoP**
	Rv0576	–	900	7	2.11×10^{-2}	*acg*, cmk*, embR*, fadE22*, lpdA*, renU(mutT3)*, Rv2186c, Rv2488c**
	Rv2989	–	8	1	2.75×10^{-2}	*acg, lpdA*, phoP*
	Rv1719	–	12	1	4.10×10^{-2}	*acg, lpdA*, Rv2488c*
GOIs Regulated by GOIs						
	Rv0491	*regX3*	31	0	–	*acg*
	Rv2488c	–	0	0	–	*acg, embR*
	Rv1267c	*embR*	5	1	–	*acg, phoP**
	Rv0757	*phoP*	795	2	–	*acg, mgtA*, fadE22**

* Significant (p-value < 0.05) differential expression

Having identified two instances of convergent evolution at the level of individual loci, we investigated further for evidence of convergent evolution at the regulatory pathway level. Fourteen genes of interest (GOI) encompass the coding genes containing or adjacent to variants we identified as rising to fixation during the experiment (Table 1). Using the MTB Network Portal v2 database of transcription factor overexpression (TFOE) and ChIP-Seq data, we identified regulatory pathways that contained GOI. In Table 2 we list the transcription factors (TFs) that are predicted to affect expression of multiple GOI. Of our GOI, *phoP* and *acg* are the most highly regulated, as has been noted previously for *phoP*, which is known to play a central role in the *M. tb* regulatory network (Li et al., 2019). The function of *acg* is unknown. It is annotated as a conserved hypothetical, but according to *in vivo* studies may play a

important role in infection (Hu and Coates, 2011; Singh et al., 2019). Additionally, we used the transcription factor overexpression (TFOE) data to determine if GOI annotated as regulators (*regX3, phoP, Rv2488c* and *embR*) regulated other GOI (Table 2). We found again that *acg* appears to be highly regulated, and that *phoP* also regulates *mgtA* and *fadE22*, which are both hypothesized to have a role in cell-wall processes. Taken together these results indicate that *M. tb* has a highly complex and tightly regulated network and suggests that biofilm formation involves a complex network of interacting loci.

Discussion

Variation within the Mycobacterium tuberculosis Complex (MTBC)

It has long been recognized that the MTBC is divided into well differentiated, globally circulating lineages (Hirsh et al., 2004). As an obligate human pathogen, *Mycobacterium tuberculosis* is subject to the vagaries of human behavior, and this behavior has had a prominent role in structuring *M. tb* populations (Pepperell et al., 2011; Brynildsrud et al., 2018; Liu et al., 2018; Mulholland et al., 2019; O'Neill et al., 2019). In addition to genetic variation, phenotypic differences have been described among strains of *M. tb*. Differences have been observed in a range of phenotypes including metabolism (Øyås et al., 2020), gene expression (Domenech et al., 2017), DNA methylation (Modlin et al., 2020), *in vitro* stress responses (Tizzano et al., 2021), and biofilm morphotype (Pang et al., 2012).

We have shown previously that *M. tb* populations are genetically sub-divided over fine scales as a result of neutral genetic drift (Pepperell et al., 2010). Here, we find that *M. tuberculosis* strains from within a single lineage (lineage 4, 'Euro-American'), separated by few genetic differences (Figure 1), exhibit variation across a range of phenotypes. Prior to the imposition of selection pressures *in vitro*, differences among strains were evident in pellicle morphotype (Figure 1), cell length (Figure 2), pellicle wet weight (Figure 3) and planktonic growth rate (Figure 4). In some cases, phenotypic clustering occurred at the sub-lineage scale – e.g., planktonic growth curves of L4.9 isolates – and in others sub-lineage strain pairs exhibited marked differences – e.g., planktonic growth rate and pellicle wet weight of the L.4.1.1.2 strains MT345 and MT72.

Genetic and phenotypic responses to selection for biofilm growth

Regardless of phenotypic starting point, all strains of *M. tuberculosis* in our study showed evidence of adaptation in response to selection for pellicle growth. The gross morphology of *M. tb* pellicles changed over the course of the experiment (Figure 1), pellicle wet weights increased for all strains (Figure 3) and increases in the amount of extracellular matrix were evident in scanning electron micrographs (SEM)

for all strains (Figure 2). This demonstrates that our method was effective in exerting selection for pellicle growth. More broadly, it shows that experimental evolutionary approaches can be applied successfully to complex phenotypes and fastidious, slow growing, non-model organisms.

During our experiment pellicle phenotypes stabilized between passage 2 and 8 (Figure 1); extended passaging beyond 8 passages did not produce further phenotypic changes (Supplementary Figure 1). Fixation of novel and previously segregating mutations occurred over the same time frame (Table 1; Figure 5), and we did not observe any further sweeps over the course of the experiment. Within 8 passages we thus saw phenotypic and genotypic stabilization across all populations. This phenomenon is frequently observed during experimental evolution, where rapid adaptation is followed by diminishing returns after continued passaging (Phillipe et al, 2007; Barrick et al, 2009; Khan et al, 2011).

We observed few mutations over the course of the experiment (Table 1), and some of these are likely to be hitchhikers linked to advantageous alleles that swept to fixation (Table 1, Figure 5). These few mutations appeared to have pleiotropic effects, as we observed changes in phenotypes other than the one under selection. 'Off target' phenotypic changes varied among strains; for example, planktonic growth rates decreased after passaging for MT540 whereas they increased for MT49, MT345, MT31 and MT55. Cell length increased for MT72 and MT55, whereas it decreased for MT345. Experimentally evolved populations generally manifest first-pass, large-effect mutations during initial adaptation (Buskirk et al., 2017); our results suggest that these mutations in *M. tb* are broad in their scope and variable in their effects. The latter phenomenon may be due to a large genetic target for biofilm growth, the impacts of hitchhiking mutations and/or genetic background of the strains.

The mutations we identified here were either exceedingly rare or absent in our search of a database of almost 40,000 *M. tb* genomes. This suggests that the loci hit in our experiment are under strong purifying selection in *M. tb*'s natural environment such that the mutations we observed are associated with substantial fitness costs to the bacterium during human infection. These fitness costs could relate to the trait under selection, e.g. cellular adhesion and the elucidation of extracellular matrix, to off-target effects such as planktonic growth and cell size, and/or phenotypes unmeasured here. Experiments such as this one can illuminate impermissible paths in *M. tb* adaptation and reveal specific vulnerabilities with the potential to be exploited in TB therapeutics.

Two strains in our study evolved a large (~120kb) duplication in the same genomic region, which fixed by passage 8 (Table 1, Figure 7). The duplication persisted without evidence of decay in extended

passaging in our study (up to passage 20 in MT31). Duplications in this region of the genome have been identified previously in isolates from Lineages 2 and 4 (Domenech et al, 2010; Domenech et al, 2014; Weiner et al, 2012). Duplications at this site are likely facilitated by flanking *IS6110* elements (Figure 7C) which can induce unequal homologous recombination. Prior research has shown that the duplication confers a fitness disadvantage *in vivo* (as measured in a mouse model) and appears to be selected for *in vitro* during axenic culture (Domenech et al, 2014).

Here we find this apparently non-specific lab adaptation to be associated with enhanced pellicle formation. Large-scale genome rearrangements including duplications have previously been identified as a rough adaptive 'first draft' followed by further refinement (Lynch & Conery, 2000; Zhang, 2003). The duplication was the only mutation that fixed in MT55, which exhibited a dramatic increase in pellicle weight (Figure 3) and increased production of ECM (Figure 2) following passaging. The apparent impacts of the duplication were not limited to pellicle growth, as MT55 also exhibited an increase in cell length (Figure 2) and a faster rate of growth during planktonic culture, as did its sister strain MT31 that also encoded the mutation (Figure 4). Taken together, our findings suggest that the production of ECM and other aspects of biofilm growth are tightly regulated during natural infection, and that first pass mutations outside of this environment result in de-repression of biofilm production. Our findings further suggest that biofilm formation is co-regulated with a range of phenotypes, consistent with prior research on biofilms, persistence and drug tolerance (Richards et al., 2019).

Epistasis

The 120kb duplication occurred on two strains from the same sub-lineage (MT31 and MT49 from L4.9) and did not occur in any other genetic background. A similar phenomenon emerged in association with sub-lineage L4.4.1.1, where both strains underwent mutation at a transcription factor binding site upstream of *lpdA*, with attendant impacts on gene expression. The repeated mutation at this locus provides strong support for its involvement in biofilm formation. In addition, the association of this mutation and the deletion with specific sub-lineages suggests that *M. tb* strains' genetic backgrounds played a role in shaping the adaptive path to enhanced biofilm growth.

Epistasis, which refers to the phenotypic impact of interactions among mutations at distinct loci, has been observed previously during experimental evolution of microbes [e.g. (Fisher et al., 2019)]. Sign epistasis is a form of epistasis in which a mutation is advantageous on one genetic background and deleterious on another (Weinreich et al., 2005). The presence of sign epistasis can hinder access to a fitness optimum, and thus it represents a constraint on adaptation (Weinreich et al., 2005; Poelwijk et

al., 2007). Our identification of mutations only in association with specific sub-lineages is consistent with sign epistasis, as is the observation of parallel adaptation in this small sample of strains: it suggests, for example, that mutations upstream of *lpdA* may be deleterious on some genetic backgrounds, and/or that alternatives are deleterious on the L.4.1.1.1 background. A careful review of the existing literature reveals additional evidence of sign epistasis in *M. tb*. For example, we and others have found evidence suggesting that *M. tb* lineage affects the mutational path to drug resistance (Mortimer et al., 2018; Farhat et al., 2019; Castro et al., 2019). Even more striking, Manson *et al* identified a stereotypic order of resistance mutation acquisition in a large survey of drug resistant *M. tb* isolates, a finding that was subsequently replicated (Manson et al., 2017; Ektefaie et al., 2021). This suggests that the acquisition of an initial resistance mutation alters the fitness associated with subsequent mutations, as has been observed in other settings (Silva et al., 2011). More direct evidence for sign epistasis comes from mutagenesis experiments, in which the lethality of gene knockouts has been shown to vary according to *M. tb* strain genetic background (Maksymiuk et al., 2015; Carey et al., 2018).

Upregulation of lpdA during biofilm growth

As noted above, we identified convergent adaptation in the region upstream of *lpdA* (*Rv3303c*) on L.4.1.1.1 backgrounds. *lpdA* is annotated as a NAD(P)H quinone reductase, which is known to be expressed *in vivo* (Deb et al, 2002) and to contribute to virulence in mice (Akhtar et al, 2006). It has not previously been linked with biofilm formation. The mechanism of its contribution to virulence *in vivo* is thought to be by protection against oxidative stress (Akhtar et al, 2006). The strains in our study evolved variants 5 bp apart, both of which lie within the TFBS for *Rv1719* (Figure 6A). These SNPs also lie within ncRNA *ncRv13303c*, which was originally identified by RNA sequencing (Dejesus et al, 2017; Arnvig et al, 2011) but is yet uncharacterized. Previous work characterizing variable-number tandem repeats (VNTR) in this intergenic region have proposed a possible hairpin structure that lies within *ncRv13303c* and could affect its function (Zheng et al, 2008). Differences in VNTR copy number and promoter sequence in this region have been observed between different strains of mycobacteria, and even linked to differences in expression of *lpdA* (Perez-Lago et al, 2013; Akhtar et al, 2009; Zheng et al, 2008).

We hypothesized that SNPs in the TFBS upstream of *lpdA* would affect TF binding and expression of *lpdA*. This was indeed the case, as evolved strains had significantly higher expression of *lpdA* (Figure 6B). Additionally, we found expression of this gene to be higher during biofilm growth when compared

to planktonic growth. Interestingly, *lpdA* is within the bounds of the large duplications we identified in strains MT31 and MT55, which may affect the dosage of this gene in those evolved strains.

We identified the two SNPs upstream of *lpdA* at rare frequencies among clinical isolates of *M. tb*: in a search of almost 40,000 *M. tb* genomes, we identified the MT49 SNP in 61 genomes and the MT540 SNP was found in 4 genomes (Table 1). Interestingly, the MT49 mutation was shown to sweep to fixation in a deep sequencing study of serial sputum samples chronicling *M. tb* within host dynamics (Trauner et al, 2017). Taken together, these observations suggest that variants at the TFBS upstream of *lpdA* can be transiently selected during infection but are otherwise deleterious, possibly due to a role for biofilm formation at a specific phase of natural infection.

The regulatory system as a target of selection in M. tb

The L4.9 duplication and SNPs upstream of *lpdA* presumably act via regulation/gene dosage. Regulatory activity was also common among the other loci that accumulated mutations in our experiment, including phosphorelay signal transduction and DNA binding activity (Table 1). We identified fixed SNPs in four transcriptional regulators: *regX3, phoP, embR* and *Rv2488c* (Table 1). These genes in turn regulate other candidate genes identified in our study (Table 2). *RegX3* and *phoP* belong to two-component regulatory systems (2CRS), whereas *embR* and *Rv2488c* are of the ompB and luxR families, respectively. Many of *M. tb*'s 2CRS have been implicated in virulence and survival in host conditions (Li et al, 2019) and this is true of *regX3*, which with its partner *senX3* is required for virulence in mice (Parish et al, 2003; Li et al, 2019). Of note, *regX3* regulates genes that are known to be involved in *M. tb* biofilm development (Richards et al., 2019). Similarly, while *Rv2488* has not previously been directly implicated in *M. tb* biofilm growth, luxR regulators are involved in quorum sensing in gram negative bacteria and have been hypothesized to play a similar role in *Mycobacteria* (Chen and Xie, 2011; Sharma et al., 2014). Our results provide support for further investigation of quorum sensing in *M. tb* and its relationship to biofilm development.

Table 2 highlights regulatory connections among candidate genes identified in this study. Several candidate genes are clustered on the regulatory network. In some cases, this is because they are responsive to a master regulator such as *Rv0576*, which downregulates 900 genes including 7 out of 14 of our genes of interest. Little is known about *Rv0576,* which is a member of the *arsR* family and is involved in response to metal ion stress (Gao et al, 2012). Several candidate genes are themselves responsive to a large number of regulators, e.g. *acg*, which was found in all of the regulatory pathways we identified (Table 2). This gene is annotated as a conserved protein that is expressed within the

dosR regulon and is required for growth and virulence in mice and macrophages (Hu & Coates, 2011). The *dosR* regulon is important for *M. tb* persistence, a phenotype that overlaps with biofilm growth (Richards et al., 2019). Three of our highly regulated genes are regulated by *whiB4* – a TF that responds to oxidative stress and is involved in modulating virulence and dissemination during infection (Chawla et al, 2012).

Broadly speaking, these results show that selection in our system appears to have targeted genes that regulate many other genes (hubs, e.g. *PhoP*), and genes under regulation by many genes (e.g. *acg*), whose expression is likely responsive to a broad range of conditions. This reflects general features of *M. tb's* regulatory network, which exhibits a high degree of connectivity between regulatory pathways (Galagan et al, 2013; Chauhan et al, 2016) as well as a hierarchical structure in which master regulators can rapidly calibrate global patterns of gene expression (Chauhan et al, 2016; Parvati Sai Arun et al, 2018). Similar findings have also been obtained in other systems, where experimentally evolved populations have shown adaptation in transcription factors and other global regulators, with global impacts on gene expression (Phillipe et al, 2007; Ali & Seshasayee, 2020; Rodriguez-Verdugo et al, 2016; Saxer et al, 2014; Conrad et al, 2010). These mutations occur quickly to enable rapid adaptation to new environments, and it is hypothesized that secondary selection may act to refine gene expression after the initial burst of adaptation (Ali & Seshasayee, 2020; Rodriguez-Verdugo et al, 2016). Thorpe et al previously found evidence of positive selection at promoter sites in natural populations of *M. tb*, suggesting that adaptation in natural as well as experimental populations is facilitated by regulatory mutations (Thorpe 2017). The broad potential impacts of regulators such as *regX3, phoP, embR* and *Rv2488c* and large-scale duplications likely helps to explain the wide range of phenotypic changes we observed. Further passaging of our isolates will reveal whether these adaptations can be further refined or whether pellicle growth remains inextricably linked to a range of phenotypes.

Of note, none of the genes identified in our study have previously been implicated in biofilm development, with the exception of *renU/mutT3*. Interestingly, previous research (Wolff 2015) found that *renU* knockouts were defective in pellicle growth, whereas a frameshift mutation in *renU* was associated with enhanced biofilm growth in MT72 in our study. The novelty of our findings highlights the potential for experimental evolution to complement more traditional approaches of interrogating important pathogen phenotypes, particularly in elucidating complex interactions and the phenotypic impacts of subtle genetic changes.

From some perspectives, the adaptation of *M. tb* seems impossibly constrained: subject to genetic drift imposed by its host's unpredictable behavior, subject to the limitations of a fully linked genome, navigating a fitness landscape rendered complex by powerful epistatic interactions. Yet *M. tb* adapts rapidly in both natural and experimental settings. *M. tb*'s complex regulatory architecture exhibits many of the features that characterize robust systems – such as redundancy, modularity, and multiple feedback mechanisms (Kitano, 2004) – and we hypothesize that it is among the reasons *M. tb* continues to persist and thrive across the globe.

Materials and Methods

Bacterial strains and growth conditions.

Clinical strains were initially isolated from sputum samples. Briefly, sputum samples were decontaminated and struck on Löwenstein–Jensen slants. *Mycobacterium tuberculosis* growth was inoculated into Middlebrook 7H9 broth (HiMedia) containing 0.2% w/v glycerol, 10% v/v OADC supplement (oleic acid, albumin, D-glucose, catalase; Becton Dickinson) and 0.05% w/v Tween-80 (7H9OTG) and incubated at 37^0C with 5% CO_2 for 3-5 weeks. Cultures were sub-cultured in 7H9OTG prior to the start of this experiment. 6 clinical strains were selected for experimental passage (MT31, MT49, MT55, MT72, MT345, and MT540). The terminal culture (Passage 0) of each strain was grown to an OD_{600} ~1 in planktonic culture and inoculated into media for biofilm growth (Passage 1; Figure 8). Aliquots of passage 0 were frozen, and genomic DNA was isolated for sequencing. For planktonic growth, 250 µL was inoculated into 5 mL 7H9OTG, in duplicate, and incubated at 37^0C with shaking, until grown to an OD_{600} ~1. For biofilm growth, 250 µL of planktonic culture was inoculated into 25 mL Sauton's medium (for 1L: 0.5g KH_2PO_4, 0.5g $MgSO_4$, 4g L-Asparagine, 2g Citric acid, 0.05g Ammonium Iron (III) citrate, 60mL glycerol, adjust pH to 7.0 with NaOH) containing 0.1% w/v $ZnSO_4$, in duplicate, and incubated at 37^0C, with 5% CO_2, without shaking for 5-7 weeks. Biofilms were grown as previously described (Kulka et al., 2012): 250 mL bottles (Corning, 430281) were incubated with a tight-fitting cap for 3 weeks, and then with a loose-fitting cap for the remainder of growth.

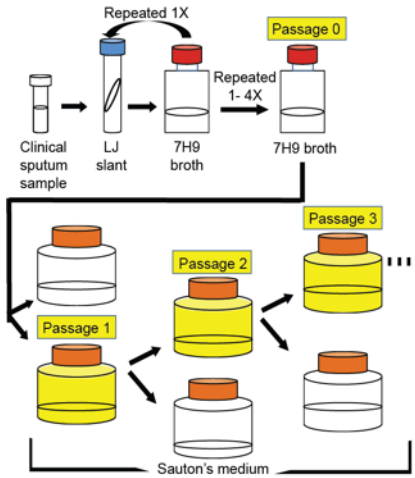

Figure 8: Protocol for serial passage of *M.tb* pellicles

Experimental evolution

As illustrated in Figure 8, duplicate populations of 6 clinical strains were grown as a biofilm. After 5-7 weeks, the more robust of the two pellicles was selected for passage. 0.3 g (wet weight) of pellicle was inoculated into 25 ml Sauton's in duplicate and grown as previously described. The pellicle not selected for passage was discarded. Between 8 and 20 total pellicle passages per strain were performed. Every 4 passages culture were frozen at -80^0C and gDNA was extracted for sequencing (excluding MT31 and MT49, for whom the first passage that was frozen and had gDNA extracted was Passage 8).

Pellicle phenotypes and wet weights

Pellicle growth was photo documented and phenotyped after 5-7 weeks of growth during the course of the experiment. Wet weights were taken after regrowing both ancestral and Passage 8 strains from cryovial stocks, as described above, inoculating first into 7H9OTG and then into Sauton's in triplicate. Wet weight measurements were taken after 5 weeks of pellicle growth by removing the spent media beneath the pellicle and weighing. Then, the tare weight of the empty bottle (which was measured before culture inoculation) was subtracted.

SEM

For scanning electron microscopy experiments, samples were taken from *M. tuberculosis* pellicle cultures after 5 weeks of growth and placed on poly-L-lysine-treated plastic coverslips (13 mm, Thermonax plastic for cell culture). Samples were then fixed overnight in a 4% formaldehyde, 2% glutaraldehyde solution in Dulbecco's phosphate-buffered saline (-calcium, -magnesium) (DPBS) (Hyclone Laboratories Inc., Logan, UT). Following fixation, samples were washed with DPBS, then treated with 1% osmium tetroxide for 1 hour. Following osmium tetroxide treatment, samples were washed with DPBS. Next, cells underwent ethanol dehydration, which was followed by critical point drying. Following that, samples were then placed on aluminum stubs and sputter-coated with 20 nm platinum. The samples were imaged at 3 kV by a Zeiss GeminiSEM 450 scanning electron microscope.

To measure cell length via scanning electron microscopy, SEM images taken at 2000X were opened on ImageJ. The pixel length of the scale bar was measured with the "straight" tool. Next, the image was cropped to remove any area without *M. tb* cells. Six frames per image were selected using the "unbiased frames" macro in ImageJ. The cell length in pixels of a fully exposed *M. tb* cells was measured in each frame using the "straight" tool. A random number generator was used to select frames for the remaining four cell lengths. After all cell lengths were measured, pixel length was converted to µm using the scale bar prior to analysis. Cell lengths were measured across two biological replicates for all strains except SK540-0, where only one replicate was available.

Planktonic growth curves

To compare rates of growth between ancestral and evolved populations, we inoculated 5 mL 7H9OTG with 250 µL thawed culture from a previously frozen cryovial for each of the ancestral strains and passage 8 of each of the evolved strains. Cultures were incubated at 37^0C with shaking, until grown to an OD_{600} ~ 1. 15mL fresh 7H9OTG was seeded at a starting OD_{600} of 0.02 in triplicate and incubated at 37^0C with shaking. OD_{600} measurements were taken every day for 10 days (starting at day 2).

DNA extraction

gDNA for whole genome sequencing was isolated by following a modified Qiagen DNeasy Blood and Tissue Kit protocol. Briefly, cultures were pelleted and resuspended in TE buffer in 2 mL screw-cap tubes containing ~250µg of 0.1mm glass beads. The samples were heated to 80^0C for 50 minutes to sterilize the suspensions, then frozen at -80^0C for at least 1 hour to aid cellular lysis. Tubes were vortexed for 3 minutes for mechanical lysis. Suspensions were then incubated with 3% w/v mutanolysin (Sigma-Aldrich, M9901) for 1 hour at 37^0C. Beads and cellular debris were pelleted, and supernatants were transferred to tubes for DNA purification following the protocol detailed in the Qiagen DNeasy Blood and Tissue Handbook, starting with step 4 of "Protocol: Pretreatment for Gram-Positive Bacteria."

Library preparation and sequencing

Library preparation was performed using a modified Nextera protocol as described by (Baym et al., 2015) with a reconditioning PCR with fresh primers and polymerase for an additional 5 PCR cycles to minimize chimeras and a two-step bead based size selection with target fragment size of 650 bp. Libraries were submitted to the University of Wisconsin-Madison Biotechnology Center. Quality and quantity of the finished libraries were assessed using an Agilent DNA High Sensitivity chip (Agilent Technologies, Santa Clara, CA) and Qubit dsDNA HS Assay Kit, respectively. Libraries were standardized to 2 µM. Samples were sequenced generating paired end reads using Illumina HiSeq or NovaSeq (depending on when the sample was sequenced).

Reference guided assembly and variant calling.

Genome assemblies were performed using an in-house reference-guided assembly pipeline (https://github.com/pepperell-lab/RGAPepPipe). Briefly, raw data was checked for quality with FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed using TrimGalore v0.6.4 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads were mapped to the H37Rv reference genome using BWA-MEM v0.7.12 (Li, 2013), and Samtools v1.11 (Li et al., 2009) view and sort were used to process SAM and BAM files. Picard v1.183 (http://broadinstitute.github.io/picard/) was used to remove duplicates and add read information. Reads were locally re-aligned using GATK v3.5 (DePristo et al., 2011) and variants were identified using Pilon v1.16 (Walker et al., 2014). Finally, assembly quality was determined using Qualimap BamQC (Okonechnikov et al., 2016).

Lineage Typing and Phylogenetic tree

We used SNP-IT (Lipworth et al., 2019) to identify the lineages of our ancestral strains. To identify the sub-lineage of our ancestral strains we obtained an alignment of known L4 isolates from a collaborator. We created a SNP alignment of the assemblies from the RGAPepPipe using Snp-sites v 2.4.1 (Page et al., 2016) and constructed a maximum likelihood phylogeny using RAxML v8.2.3 (Stamatakis, 2014). Next, we visualized the phylogeny using Dendroscope (Huson and Scornavacca, 2012) and lineage typed our isolates by comparing their location in the phylogeny to known L4 subtypes. The phylogenetic tree in Figure 1 was plotted and colored in R using ggtree (Yu et al., 2017).

Identification of fixed SNPs

To identify genetic changes occurring after repeated passaging, we used Popoolation2 v1.201 (Kofler et al., 2011b) and breseq v0.35.0 (Deatherage and Barrick, 2014). BAM files from RGAPepPipe (see above) were filtered for quality and used with Samtools mpileup (Li et al., 2009) to tabulate mutation frequencies across sequenced timepoints. The mpileup file was converted to a sync file using Popoolation2 and finally converted the sync file to a format suitable for downstream analysis using Python (code available at https://github.com/myoungblom/mtb_ExpEvo). Variants were filtered for quality by minimum coverage of 20X, minimum alternate allele count of 5 and a minimum allele frequency of 5%. We also removed variants in repetitive families of genes (PE, PPE, and PE-PGRS gene families) from analysis. Sequence data were also assembled and variants called and annotated using breseq v0.35.0 (Deatherage and Barrick, 2014) to confirm variants called using Popoolation2. Breseq was run in population mode for each sequenced timepoint, with all default parameters except one: “polymorphism-minimum-total-coverage-each-strand” was set to 4. Then the mutational frequencies of each timepoint were compared using gdtools (Deatherage and Barrick, 2014) compare run with default parameters. To identify mutations that had significant frequency changes throughout the experiment we looked for variants that were at a low starting frequency and rose to fixation (>95% frequency) or conversely, started at intermediate frequency and disappeared from the population (<5% frequency). Based on these criteria, we identified 14 fixed SNPs and 2 fixed single base pair deletions (Table 1).

Structural variant detection

To identify potential insertions, deletions, and duplications of interest in our genomic sequence data, we called structural variants using Pilon v1.16 (Walker et al., 2014) as a part of RGAPepPipe. We removed all insertions, deletions, and duplications that were located entirely in regions annotated as repetitive genes (PE, PPE, and PE-PGRS gene families). We also removed variants that were identified in

ancestral strains and/or that did not remain fixed in subsequent passages. Additionally, we manually curated potential variants by visually inspecting the alignments in IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) and removed variants that appeared to be mis-mappings or in regions of poor coverage. After curation of the data, we identified two large duplications that fixed in our datasets (Table 1). Sliding window coverage plots to visualize the duplications in MT31 and MT55 were made using Samtools bedcov (Li et al., 2009) using in-house scripts (code available at https://github.com/myoungblom/mtb_ExpEvo). Relative coverage was calculated by dividing each window coverage by the average coverage across the assembly – as given by BamQC (Okonechnikov et al., 2016) – and plotted in R.

GO term enrichment analysis

To identify possible functional enrichments in the genes within the large duplication that evolved in MT31 and MT55 (Table 1), we performed a GO term enrichment analysis. GO terms for all *M. tb* genes were downloaded from Quick GO (https://www.ebi.ac.uk/QuickGO/). We used the R package topGO (https://bioconductor.org/packages/release/bioc/html/topGO.html) to perform an enrichment analysis of the genes present in one or both duplications (a total of 172 genes). We calculated significance using a Fishers exact test and filtering for p-values < 0.01 (Supplementary Table 2). We performed the same enrichment analysis on 100 randomly selected, contiguous sets of 172 genes to test the greater significance of our results (Supplementary Figure 2).

Genotype plots

To better understand the structure of the population over passaging we created genotypes by clustering variants with similar trajectories throughout the experiment. Output from Popoolation2 as described above, was filtered for based on >= 30% frequency change between the final passage and ancestral populations, as well as filtering out all mutations within 1000bp of each other (we identified that these mutations are often the result of repetitive regions and/or poor mapping). Lolipop (https://github.com/cdeitrick/Lolipop) was used with default parameters to cluster mutations into genotypes and create plots.

Population genetics statistics

Population genetics estimates (Tajima's D, pi and theta) were calculated as previously described (O'Neill et al., 2015). Briefly, we used Popoolation (Kofler et al., 2011a) to estimate these parameters in sliding windows across the genome. To reduce biases caused by variable coverage we randomly sub-sampled read data without replacement to uniform coverage of 50X; we repeated this process 9 times

to further reduce biases. Genome-wide averages were calculated by averaging all windows across the genome, across all replicate sub-sampled data. Scripts available at https://github.com/myoungblom/mtb_ExpEvo.

Network analysis

Genes of interest (GOI) were defined as coding genes containing or just downstream of the fixed mutations we identified (Table 1). Network analysis of GOI was performed by accessing the transcription factor overexpression data available in the MTB Network Portal (http://networks.systemsbiology.net/mtb/content/TFOE-Searchable-Data-File). Transcription factors identified as significantly contributing to regulation of more than one of our GOIs were included in Table 2. We also used this same dataset to identify GOIs which are themselves regulators and affect expression of other GOIs (Table 2).

Fixed SNPs in global datasets

We searched for the presence of our fixed mutations in publicly available sequencing data using a searchable COBS index of bacterial genomes curated from the European Nucleotide Archive (Blackwell et al., 2021). We used the Python interface to search the COBS index (Bingmann et al., 2019). Our query sequences included 50bp in either direction of the SNP of interest, with a k-mer matching threshold of 1. To determine the frequency of isolates in the COBS index with each SNP of interest, we needed to determine the number of *M. tb* genomes in the database: we searched the database using the full H37Rv 16S rRNA sequence (accession: NR_102810.2) which returned 39,941 results – this total number of *M. tb* isolates was used to calculate frequency of our SNPs of interest (Table 1).

RNA isolation

To investigate the consequence of the convergent SNPs we identified in MT540 and MT49, within a TFBS upstream of *lpdA,* we examined *lpdA* expression in ancestral and evolved populations of these strains, during both planktonic and biofilm growth. Frozen cryovials of ancestral populations (MT540-0, MT49-0) and evolved populations (MT540-4, MT49-12) were thawed, and 250 µl of each was inoculated into 25 ml 7H9OTG and incubated at 37^0C with shaking and grown to an OD_{600} of ~1. 250 µl of culture was used to inoculate 25 ml Sauton's for biofilm growth. The remaining culture was pelleted, resuspended in 3 ml RNAprotect Bacteria Reagent (Qiagen), divided into 3 x 1 ml aliquots, and frozen at -80^0C until RNA extraction. Biofilm cultures were grown as previously described, and pellicles were harvested at 5 weeks. Pellicles were broken up and suspended in 3 ml RNAprotect, divided into 3 x 1 ml aliquots, and frozen at -80^0C until RNA extraction. RNA was extracted from each 1 ml aliquot (three

replicates/ sample), using the illustra RNAspin Mini RNA Isolation Kit (GE Healthcare) following the "Protocol for total RNA purification from up to 10^9 bacterial cells," which includes a 15-minute room-temperature incubation with RNase-free DNase I. RNA integrity was assessed by bleach gel electrophoresis (O'Neill et al., 2018) and quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc).

RT-qPCR

~100 ng of each RNA sample was used for cDNA synthesis with random hexamer primers using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) and included a 5-minute incubation at 65^0C and a reaction temperature of 44^0C for GC-rich templates. Real-time quantitative PCR was performed in EU Fast Cycler PCR 96-well plates (BIOplastics) using the StepOnePlus Real-Time PCR System (v. 2.3, Applied Biosystems). We used sets of primers previously published and verified (Akhtar et al., 2009) to measure the expression of *lpdA* normalized to the expression of the endogenous control, *sigA.* All real-time PCR assays were run in a total reaction volume of 20 µl comprised of 2X Fast SYBR Green Master Mix (Applied Biosystems), 200 nM of both forward and reverse primers (Integrated DNA Technologies, Coralville, IA, USA), and 3 µl of cDNA. RT-qPCR cycling parameters were set as follows: an initial AmpliTaq Fast DNA Polymerase, UP Activation step of 20 seconds at 95^0C, followed by 40 cycles of 3 seconds at 95^0C and 30 seconds at 60^0C. Each reaction was repeated three times with three independent cDNA samples. Negative controls consisting of no-template reaction mixtures were run with all reactions. Melting-curve analysis was carried out to confirm the specificity of the amplified product. After baseline correction and determination of the threshold settings, the Factor-qPCR ratio method (Ruijter et al., 2015) was used to remove between-run variation due to the experiment including multiple plates. The factor corrected ΔCT values of the three biological replicates were then averaged and the $\Delta\Delta$CTs were calculated by subtracting a reference value (MT540-0, planktonic growth mean) from each sample. The expression fold change was calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001), as PCR efficiencies had previously been found to be similar. Low and high fold changes were calculated by adding or subtracting the standard deviation from the fold change. The results are expressed as log2 of the fold changes. Kruskal-Wallis test was used to determine the presence of significant log2 fold differences in the dataset (p-value = 1.89e-04). Then, a Wilcox test was performed to compare log2 fold changes between sample types (AP, AB, EP, EB) to identify significant differences between pairs. A p-value less than 0.05 was considered significant.

Data Availability

Raw sequence data has been submitted to NCBI under the project accession PRJNA720149.

Author contributions

TMS and CSP conceived of the study. TMS, MAY and TDM conducted experiments. TMS and JFK performed SEM and image analysis. TMS, MAY, MAM, LLB, MBO and CSP performed data analysis. MAY drafted the manuscript with feedback from CSP. All authors provided feedback and contributed to manuscript revision.

Funding

This work was funded by the National Institutes of Health, NIAID R01AI113287 to CSP. MAY was funded by National Science Foundation Graduate Research Fellowship Program under grant No. DGE-1747503. LLB was funded by National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award No. T32-AI55391. TDM and MBO were funded by National Science Foundation Graduate Research Fellowship Program under grant No. DGE-1256259 with additional funding for TDM by National Institute of Health National Research Service Award No. T32-GM07215.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We would like to acknowledge Soleil Young and Sydney Fry for their help phenotyping pellicles. We also thank the University of Wisconsin Biotechnology center for providing sequencing services.

References

Ackart, D.F., Hascall-Dove, L., Caceres, S.M., Kirk, N.M., Podell, B.K., Melander, C., Orme, I.M., Leid, J.G., Nick, J.A., Basaraba, R.J., 2014. Expression of antimicrobial drug tolerance by attached communities of *Mycobacterium tuberculosis*. Pathog. Dis. 70, 359–369. https://doi.org/10.1111/2049-632X.12144

Akhtar, P., Singh, S., Bifani, P., Kaur, S., Srivastava, B.S., Srivastava, R., 2009. Variable-number tandem repeat 3690 polymorphism in Indian clinical isolates of Mycobacterium tuberculosis and its influence on transcription. J. Med. Microbiol. 58, 798–805. https://doi.org/10.1099/jmm.0.002550-0

Aranda, P.S., LaJoie, D.M., Jorcyk, C.L., 2012. Bleach Gel: A Simple Agarose Gel for Analyzing RNA Quality. Electrophoresis 33, 366–369. https://doi.org/10.1002/elps.201100335

Arnvig, K.B., Comas, I., Thomson, N.R., Houghton, J., Boshoff, H.I., Croucher, N.J., Rose, G., Perkins, T.T., Parkhill, J., Dougan, G., Young, D.B., 2011. Sequence-Based Analysis Uncovers an Abundance of Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. PLoS Pathog. 7. https://doi.org/10.1371/journal.ppat.1002342

Bacon, J., Alderwick, L.J., Allnutt, J.A., Gabasova, E., Watson, R., Hatch, K.A., Clark, S.O., Jeeves, R.E., Marriott, A., Rayner, E., Tolley, H., Pearson, G., Hall, G., Besra, G.S., Wernisch, L., Williams, A., Marsh, P.D., 2014. Non-Replicating Mycobacterium tuberculosis Elicits a Reduced

Infectivity Profile with Corresponding Modifications to the Cell Wall and Extracellular Matrix. PLOS ONE 9, e87329. https://doi.org/10.1371/journal.pone.0087329
Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., Kishony, R., 2015. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes (No. biorxiv;013771v2).
Bharti, S., Maurya, R.K., Venugopal, U., Singh, R., Akhtar, Md.S., Krishnan, M.Y., 2021. Rv1717 Is a Cell Wall - Associated β-Galactosidase of Mycobacterium tuberculosis That Is Involved in Biofilm Dispersion. Front. Microbiol. 11, 611122. https://doi.org/10.3389/fmicb.2020.611122
Bingmann, T., Bradley, P., Gauger, F., Iqbal, Z., 2019. COBS: A Compact Bit-Sliced Signature Index, in: Brisaboa, N.R., Puglisi, S.J. (Eds.), String Processing and Information Retrieval, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 285–303. https://doi.org/10.1007/978-3-030-32686-9_21
Blackwell, G.A., Hunt, M., Malone, K.M., Lima, L., Horesh, G., Alako, B.T.F., Thomson, N.R., Iqbal, Z., 2021. Exploring bacterial diversity via a curated and searchable snapshot of archived DNA sequences. bioRxiv 2021.03.02.433662. https://doi.org/10.1101/2021.03.02.433662
Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., Fandinho, F., Silva, M.A. da, Perdigao, J., Portugal, I., Viveiros, M., Clark, T., Caws, M., Dunstan, S., Thai, P.V.K., Lopez, B., Ritacco, V., Kitchen, A., Brown, T.S., Soolingen, D. van, O'Neill, M.B., Holt, K.E., Feil, E.J., Mathema, B., Balloux, F., Eldholm, V., 2018. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci. Adv. 4, eaat5869. https://doi.org/10.1126/sciadv.aat5869
Buskirk, S.W., Peace, R.E., Lang, G.I., 2017. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proc. Natl. Acad. Sci. 114, 8330–8335. https://doi.org/10.1073/pnas.1702314114
Canetti, G., 1955. The Tubercle Bacillus in the Pulmonary Lesion of Man. Springer, New York.
Carey, A.F., Rock, J.M., Krieger, I.V., Chase, M.R., Fernandez-Suarez, M., Gagneux, S., Sacchettini, J.C., Ioerger, T.R., Fortune, S.M., 2018. TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities. PLOS Pathog. 14, e1006939. https://doi.org/10.1371/journal.ppat.1006939
Castro, R.A.D., Ross, A., Kamwela, L., Reinhard, M., Loiseau, C., Feldmann, J., Borrell, S., Trauner, A., Gagneux, S., 2019. The Genetic Background modulates the Evolution of Fluoroquinolone-Resistance in Mycobacterium tuberculosis. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msz214
Chakraborty, P., Bajeli, S., Kaushal, D., Radotra, B.D., Kumar, A., 2021. Biofilm formation in the lung contributes to virulence and drug tolerance of Mycobacterium tuberculosis. Nat. Commun. 12, 1606. https://doi.org/10.1038/s41467-021-21748-6
Chen, J., Xie, J., 2011. Role and regulation of bacterial LuxR-like regulators. J. Cell. Biochem. 112, 2694–2702. https://doi.org/10.1002/jcb.23219
Costerton, J.W., Stewart, P.S., Greenberg, E.P., 1999. Bacterial Biofilms: A Common Cause of Persistent Infections. Science 284, 1318–1322. https://doi.org/10.1126/science.284.5418.1318
Deatherage, D.E., Barrick, J.E., 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. Clifton NJ 1151, 165–188. https://doi.org/10.1007/978-1-4939-0554-6_12
DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. https://doi.org/10.1038/ng.806
Domenech, P., Zou, J., Averback, A., Syed, N., Curtis, D., Donato, S., Reed, M.B., 2017. Unique Regulation of the DosR Regulon in the Beijing Lineage of Mycobacterium tuberculosis. J. Bacteriol. 199. https://doi.org/10.1128/JB.00696-16

Dubos, R.J., Davis, B.D., 1946. FACTORS AFFECTING THE GROWTH OF TUBERCLE BACILLI IN LIQUID MEDIA. J. Exp. Med. 83, 409–423.
Ektefaie, Y., Dixit, A., Freschi, L., Farhat, M.R., 2021. Globally diverse Mycobacterium tuberculosis resistance acquisition: a retrospective geographical and temporal analysis of whole genome sequences. Lancet Microbe S2666524720301956. https://doi.org/10.1016/S2666-5247(20)30195-6
Eldholm, V., Balloux, F., 2016. Antimicrobial Resistance in Mycobacterium tuberculosis: The Odd One Out. Trends Microbiol. 24, 637–648. https://doi.org/10.1016/j.tim.2016.03.007
Farhat, M.R., Freschi, L., Calderon, R., Ioerger, T., Snyder, M., Meehan, C.J., Jong, B. de, Rigouts, L., Sloutsky, A., Kaur, D., Sunyaev, S., Soolingen, D. van, Shendure, J., Sacchettini, J., Murray, M., 2019. GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. Nat. Commun. 10, 2128. https://doi.org/10.1038/s41467-019-10110-6
Fisher, K.J., Kryazhimskiy, S., Lang, G.I., 2019. Detecting genetic interactions using parallel evolution in experimental populations. Philos. Trans. R. Soc. B Biol. Sci. 374, 20180237. https://doi.org/10.1098/rstb.2018.0237
Gerrick, E.R., Barbier, T., Chase, M.R., Xu, R., François, J., Lin, V.H., Szucs, M.J., Rock, J.M., Ahmad, R., Tjaden, B., Livny, J., Fortune, S.M., 2018. Small RNA profiling in Mycobacterium tuberculosis identifies MrsI as necessary for an anticipatory iron sparing response. Proc. Natl. Acad. Sci. 115, 6464–6469. https://doi.org/10.1073/pnas.1718003115
Global Tuberculosis Report 2019 [WWW Document], n.d. URL https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-report-2019 (accessed 3.9.21).
Hegde, S.R., 2020. Computational Identification of the Proteins Associated With Quorum Sensing and Biofilm Formation in Mycobacterium tuberculosis. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.03011
Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., Small, P.M., 2004. Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci U A 101, 4871–6. https://doi.org/10.1073/pnas.0305627101
Hu, Y., Coates, A.R.M., 2011. Mycobacterium tuberculosis acg Gene Is Required for Growth and Virulence In Vivo. PLOS ONE 6, e20958. https://doi.org/10.1371/journal.pone.0020958
Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Syst. Biol. sys062. https://doi.org/10.1093/sysbio/sys062
Kerns, P.W., Ackhart, D.F., Basaraba, R.J., Leid, J.G., Shirtliff, M.E., 2014. Mycobacterium tuberculosis pellicles express unique proteins recognized by the host humoral response. Pathog. Dis. 70, 347–358. https://doi.org/10.1111/2049-632X.12142
Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826–837. https://doi.org/10.1038/nrg1471
Kobayashi, K., 2007. Bacillus subtilis Pellicle Formation Proceeds through Genetically Defined Morphological Changes. J. Bacteriol. 189, 4920–4931. https://doi.org/10.1128/JB.00157-07
Kofler, R., Orozco-terWengel, P., Maio, N.D., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlötterer, C., 2011a. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. PLOS ONE 6, e15925. https://doi.org/10.1371/journal.pone.0015925
Kofler, R., Pandey, R.V., Schlötterer, C., 2011b. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27, 3435–3436. https://doi.org/10.1093/bioinformatics/btr589
Kulka, K., Hatfull, G., Ojha, A.K., 2012. Growth of Mycobacterium tuberculosis Biofilms. J. Vis. Exp. JoVE. https://doi.org/10.3791/3820
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map

format and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Li, X., Lv, X., Lin, Y., Zhen, J., Ruan, C., Duan, W., Li, Y., Xie, J., 2019. Role of two-component regulatory systems in intracellular survival of Mycobacterium tuberculosis. J. Cell. Biochem. 120, 12197–12207. https://doi.org/10.1002/jcb.28792
Lipworth, S., Jajou, R., de Neeling, A., Bradley, P., van der Hoek, W., Maphalala, G., Bonnet, M., Sanchez-Padilla, E., Diel, R., Niemann, S., Iqbal, Z., Smith, G., Peto, T., Crook, D., Walker, T., van Soolingen, D., 2019. SNP-IT Tool for Identifying Subspecies and Associated Lineages of *Mycobacterium tuberculosis* Complex. Emerg. Infect. Dis. 25, 482–488. https://doi.org/10.3201/eid2503.180894
Liu, Q., Ma, A., Wei, L., Pang, Y., Wu, B., Luo, T., Zhou, Y., Zheng, H.-X., Jiang, Q., Gan, M., Zuo, T., Liu, M., Yang, C., Jin, L., Comas, I., Gagneux, S., Zhao, Y., Pepperell, C.S., Gao, Q., 2018. China's tuberculosis epidemic stems from historical expansion of four strains of Mycobacterium tuberculosis. Nat. Ecol. Evol. 2, 1982. https://doi.org/10.1038/s41559-018-0680-6
Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25, 402–408. https://doi.org/10.1006/meth.2001.1262
Maksymiuk, C., Ioerger, T., Balakrishnan, A., Bryk, R., Rhee, K., Sacchettini, J., Nathan, C., 2015. Comparison of transposon and deletion mutants in Mycobacterium tuberculosis: The case of rv1248c, encoding 2-hydroxy-3-oxoadipate synthase. Tuberculosis 95, 689–694. https://doi.org/10.1016/j.tube.2015.08.009
Manson, A.L., Cohen, K.A., Abeel, T., Desjardins, C.A., Armstrong, D.T., Barry Iii, C.E., Brand, J., TBResist Global Genome Consortium, Chapman, S.B., Cho, S.-N., Gabrielian, A., Gomez, J., Jodals, A.M., Joloba, M., Jureen, P., Lee, J.S., Malinga, L., Maiga, M., Nordenberg, D., Noroc, E., Romancenco, E., Salazar, A., Ssengooba, W., Velayati, A.A., Winglee, K., Zalutskaya, A., Via, L.E., Cassell, G.H., Dorman, S.E., Ellner, J., Farnia, P., Galagan, J.E., Rosenthal, A., Crudu, V., Homorodean, D., Hsueh, P.-R., Narayanan, S., Pym, A.S., Skrahina, A., Swaminathan, S., Van der Walt, M., Alland, D., Bishai, W.R., Cohen, T., Hoffner, S., Birren, B.W., Earl, A.M., 2017. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat. Genet. advance online publication. https://doi.org/10.1038/ng.3767
Modlin, S.J., Conkle-Gutierrez, D., Kim, C., Mitchell, S.N., Morrissey, C., Weinrick, B.C., Jacobs, W.R., Ramirez-Busby, S.M., Hoffner, S.E., Valafar, F., 2020. Drivers and sites of diversity in the DNA adenine methylomes of 93 Mycobacterium tuberculosis complex clinical isolates. eLife 9, e58542. https://doi.org/10.7554/eLife.58542
Mortimer, T.D., Weber, A.M., Pepperell, C.S., 2018. Signatures of Selection at Drug Resistance Loci in Mycobacterium tuberculosis. mSystems 3, e00108-17. https://doi.org/10.1128/mSystems.00108-17
Mulholland, C.V., Shockey, A.C., Aung, H.L., Cursons, R.T., O'Toole, R.F., Gautam, S.S., Brites, D., Gagneux, S., Roberts, S.A., Karalus, N., Cook, G.M., Pepperell, C.S., Arcus, V.L., 2019. Dispersal of Mycobacterium tuberculosis Driven by Historical European Trade in the South Pacific. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02778
Nyka, W., 1977. The chromophobic tubercle bacilli and the problem of endogenous reactivation of tuberculosis. Mater. Medica Pol. Pol. J. Med. Pharm. 9, 175–185.
Nyka, W., 1967. Method for staining both acid-fast and chromophobic tubercle bacilli with carbolfuschsin. J. Bacteriol. 93, 1458–1460.
Nyka, W., 1963. STUDIES ON MYCOBACTERIUM TUBERCULOSIS IN LESIONS OF THE HUMAN LUNG. A NEW METHOD OF STAINING TUBERCLE BACILLI IN TISSUE SECTIONS. Am. Rev. Respir. Dis. 88, 670–679. https://doi.org/10.1164/arrd.1963.88.5.670

Nyka, W., O'Neill, E.F., 1970. A NEW APPROACH TO THE STUDY OF NON-ACID-FAST MYCOBACTERIA. Ann. N. Y. Acad. Sci. 174, 862–871. https://doi.org/10.1111/j.1749-6632.1970.tb45605.x

Ojha, A.K., Baughn, A.D., Sambandan, D., Hsu, T., Trivelli, X., Guerardel, Y., Alahari, A., Kremer, L., Jacobs, W.R., Hatfull, G.F., 2008. Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. 69, 164–174. https://doi.org/10.1111/j.1365-2958.2008.06274.x

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294. https://doi.org/10.1093/bioinformatics/btv566

O'Neill, M.B., Mortimer, T.D., Pepperell, C.S., 2015. Diversity of Mycobacterium tuberculosis across Evolutionary Scales. PLOS Pathog. 11, e1005257. https://doi.org/10.1371/journal.ppat.1005257

O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Mol. Ecol. mec.15120. https://doi.org/10.1111/mec.15120

Øyås, O., Borrell, S., Trauner, A., Zimmermann, M., Feldmann, J., Liphardt, T., Gagneux, S., Stelling, J., Sauer, U., Zampieri, M., 2020. Model-based integration of genomics and metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. 117, 8494–8502. https://doi.org/10.1073/pnas.1915551117

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genomics 2. https://doi.org/10.1099/mgen.0.000056

Pang, J.M., Layre, E., Sweet, L., Sherrid, A., Moody, D.B., Ojha, A., Sherman, D.R., 2012. The Polyketide Pks1 Contributes to Biofilm Formation in Mycobacterium tuberculosis. J. Bacteriol. 194, 715–721. https://doi.org/10.1128/JB.06304-11

Pepperell, C., Hoeppner, V.H., Lipatov, M., Wobeser, W., Schoolnik, G.K., Feldman, M.W., 2010. Bacterial genetic signatures of human social phenomena among M. tuberculosis from an Aboriginal Canadian population. Mol Biol Evol 27, 427–40. https://doi.org/10.1093/molbev/msp261

Pepperell, C.S., Granka, J.M., Alexander, D.C., Behr, M.A., Chui, L., Gordon, J., Guthrie, J.L., Jamieson, F.B., Langlois-Klassen, D., Long, R., Nguyen, D., Wobeser, W., Feldman, M.W., 2011. Dispersal of Mycobacterium tuberculosis via the Canadian fur trade. Proc Natl Acad Sci U A 108, 6526–31. https://doi.org/10.1073/pnas.1016708108

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445, 383–386. https://doi.org/10.1038/nature05451

Rastogi, S., Singh, A.K., Pant, G., Mitra, K., Sashidhara, K.V., Krishnan, M.Y., 2017. Down-regulation of PE11, a cell wall associated esterase, enhances the biofilm growth of Mycobacterium tuberculosis and reduces cell wall virulence lipid levels. Microbiology 163, 52–61. https://doi.org/10.1099/mic.0.000417

Richards, J.P., Cai, W., Zill, N.A., Zhang, W., Ojha, A.K., 2019. Adaptation of Mycobacterium tuberculosis to Biofilm Growth Is Genetically Linked to Drug Tolerance. Antimicrob. Agents Chemother. 63. https://doi.org/10.1128/AAC.01213-19

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754

Ruijter, J.M., Ruiz Villalba, A., Hellemans, J., Untergasser, A., van den Hoff, M.J.B., 2015. Removal of between-run variation in a multi-plate qPCR experiment. Biomol. Detect. Quantif., Special Issue: Advanced Molecular Diagnostics for Biomarker Discovery – Part I 5, 10–14. https://doi.org/10.1016/j.bdq.2015.07.001

Rustad, T.R., Minch, K.J., Ma, S., Winkler, J.K., Hobbs, S., Hickey, M., Brabant, W., Turkarslan, S., Price, N.D., Baliga, N.S., Sherman, D.R., 2014. Mapping and manipulating the Mycobacterium

tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network. Genome Biol. 15, 502. https://doi.org/10.1186/PREACCEPT-1701638048134699
Sambandan, D., Dao, D.N., Weinrick, B.C., Vilchèze, C., Gurcha, S.S., Ojha, A., Kremer, L., Besra, G.S., Hatfull, G.F., Jacobs, W.R., 2013. Keto-Mycolic Acid-Dependent Pellicle Formation Confers Tolerance to Drug-Sensitive Mycobacterium tuberculosis. mBio 4. https://doi.org/10.1128/mBio.00222-13
Sharma, I.M., Petchiappan, A., Chatterji, D., 2014. Quorum sensing and biofilm formation in mycobacteria: Role of c-di-GMP and methods to study this second messenger: QS, biofilm formation and c-di-GMP in mycobacteria. IUBMB Life 66, 823–834. https://doi.org/10.1002/iub.1339
Silva, R.F., Mendonça, S.C.M., Carvalho, L.M., Reis, A.M., Gordo, I., Trindade, S., Dionisio, F., 2011. Pervasive Sign Epistasis between Conjugative Plasmids and Drug-Resistance Chromosomal Mutations. PLOS Genet. 7, e1002181. https://doi.org/10.1371/journal.pgen.1002181
Singh, S., Sharma, M., Chaudhry, A., Sharma, S., 2019. Rv2626c and Rv2032 activate TH1 response and downregulate regulatory T cells in peripheral blood mononuclear cells of tuberculosis patients. Comp. Immunol. Microbiol. Infect. Dis. 62, 46–53. https://doi.org/10.1016/j.cimid.2018.11.016
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192. https://doi.org/10.1093/bib/bbs017
Tizzano, B., Dallenga, T.K., Utpatel, C., Behrends, J., Homolka, S., Kohl, T.A., Niemann, S., 2021. Survival of hypoxia-induced dormancy is not a common feature of all strains of the Mycobacterium tuberculosis complex. Sci. Rep. 11, 2628. https://doi.org/10.1038/s41598-021-81223-6
Trivedi, A., Mavi, P.S., Bhatt, D., Kumar, A., 2016. Thiol reductive stress induces cellulose-anchored biofilm formation in Mycobacterium tuberculosis. Nat. Commun. 7, 11392. https://doi.org/10.1038/ncomms11392
Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9, e112963. https://doi.org/10.1371/journal.pone.0112963
Wang, F., Sambandan, D., Halder, R., Wang, J., Batt, S.M., Weinrick, B., Ahmad, I., Yang, P., Zhang, Y., Kim, J., Hassani, M., Huszar, S., Trefzer, C., Ma, Z., Kaneko, T., Mdluli, K.E., Franzblau, S., Chatterjee, A.K., Johnsson, K., Mikusova, K., Besra, G.S., Fütterer, K., Robbins, S.H., Barnes, S.W., Walker, J.R., Jacobs, W.R., Schultz, P.G., 2013. Identification of a small molecule with activity against drug-resistant and persistent tuberculosis. Proc. Natl. Acad. Sci. 110, E2510–E2517. https://doi.org/10.1073/pnas.1309171110
Weinreich, D.M., Watson, R.A., Chao, L., 2005. Perspective: Sign Epistasis and Genetic Constraint on Evolutionary Trajectories. Evolution 59, 1165–1174.
Wolff, K.A., Peña, A.H. de la, Nguyen, H.T., Pham, T.H., Amzel, L.M., Gabelli, S.B., Nguyen, L., 2015. A Redox Regulatory System Critical for Mycobacterial Survival in Macrophages and Biofilm Development. PLOS Pathog. 11, e1004839. https://doi.org/10.1371/journal.ppat.1004839
Yang, D.C., Blair, K.M., Salama, N.R., 2016. Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. Microbiol. Mol. Biol. Rev. 80, 187–203. https://doi.org/10.1128/MMBR.00031-15
Yang, Y., Thomas, J., Li, Y., Vilchèze, C., Derbyshire, K.M., Jacobs, W.R., Ojha, A.K., 2017. Defining a temporal order of genetic requirements for development of mycobacterial biofilms. Mol. Microbiol. https://doi.org/10.1111/mmi.13734

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8, 28–36. https://doi.org/10.1111/2041-210X.12628