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

Tracy M. Smith[1*], Madison A. Youngblom[2,3*], John F. Kernien[3], Mohamed A. Mohamed[3], Sydney S. Fry[3], Lindsey L. Bohr[2,3], Tatum D. Mortimer[4], Mary B. O'Neill[5], Caitlin S. Pepperell[3,6]

[*] These authors contributed equally to this work

[1] New York Genome Center, New York City, NY, United States
[2] Microbiology Doctoral Training Program, University of Wisconsin-Madison
[3] Department of Medical Microbiology and Immunology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, United States
[4] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[5] Laboratoire de Biochimie (LBC), Chimie Biologie et Innovation, ESPCI Paris, PSL Université, CNRS UMR8231, Paris, France
[6] 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 contribution to 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 the 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 secrete 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

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

76 **Results**

77

78 *Sample*

79 Pellicles are a specific type of biofilm in which bacteria form aggregates at air-liquid interfaces
80 (Kobayashi, 2007); *M. tb* has been shown to form pellicle biofilms *in vitro* (Kerns et al., 2014). For our
81 study, we chose six closely related isolates of *M. tb*, from three sub-clades of the Euro-American
82 lineage L4 (4.9, 4.4.1.2, and 4.4.1.1 shown in Figure 1). The study design enabled us to identify the
83 impacts of genetic background over multiple scales, including comparisons between sub-lineages (4.9
84 and 4.4), sub-sub-lineages (4.4.1.2 and 4.4.1.1) and individual strains. In addition, the strain selection
85 encompassed a variety of biofilm phenotypes such that we could identify impacts of ancestral
86 phenotype on bacterial adaptation in our passaging system. The strains were grown as pellicles
87 following a published protocol (Kulka et al., 2012). We passaged these six populations in pellicle form
88 every 5-7 weeks as described in the Methods. Each pellicle population was passaged at least eight
89 times over a period of two years (Supplementary Figure 1; Supplementary Table 1). To investigate the
90 specificity of adaptations observed during pellicle passaging, we also passaged the six strains in
91 planktonic culture: we passaged three independently evolving populations per strain, and each
92 population was passaged four times.

93

94 *Phenotypic changes of pellicles*

95 At each passage we photographed the pellicle and described its growth according to the following
96 criteria: proportion of liquid surface covered, presence of climbing (attachment to and growth up the
97 sides of the flask), thickness of growth and continuity of growth (versus discontinuous patches).
98 Although the *M. tb* strains were closely related (i.e., separated by 100s of SNPs), differences in biofilm
99 phenotype were evident prior to passaging (Figure 1). During the initial few passages, phenotypes
100 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).

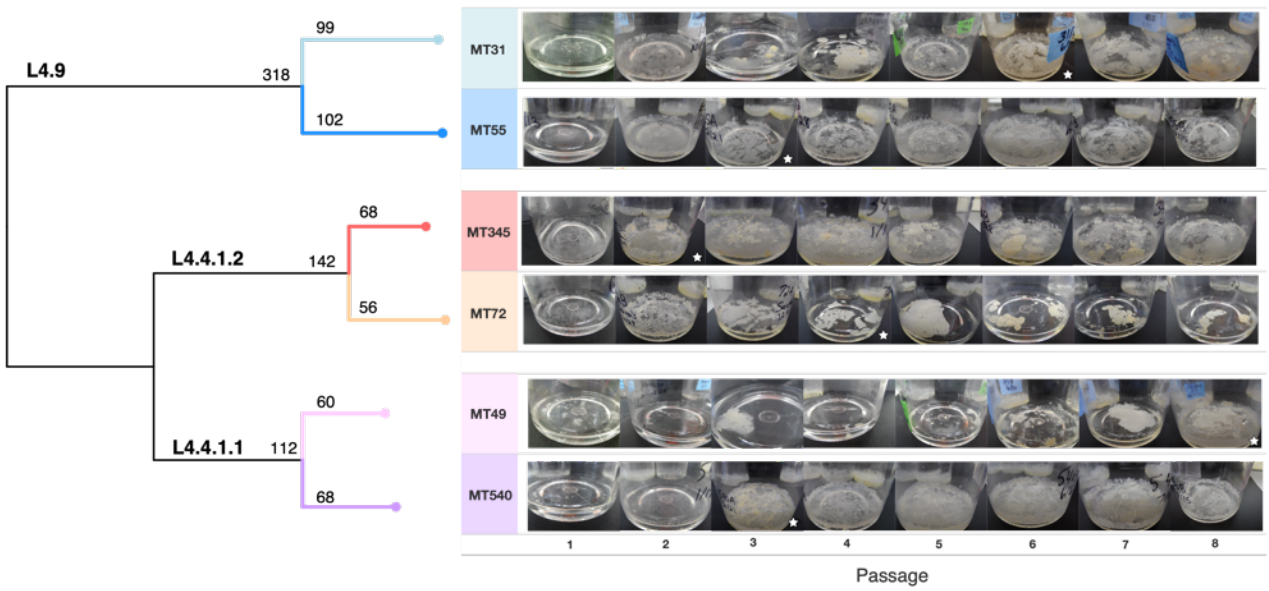


Figure 1: Whole genome sequence phylogeny of ancestral populations used for passaging and photos of pellicles throughout the experiment. Populations fall into three clades with SNP distances given for each sub-lineage and strain. Pellicle photos show rapid change in phenotype, followed by stabilization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

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

118 changes in cell length (Figure 2). The direction of the change varied among strains: MT55 and MT72
119 evolved longer bacilli whereas MT345 evolved shorter bacilli (Figure 2).
120

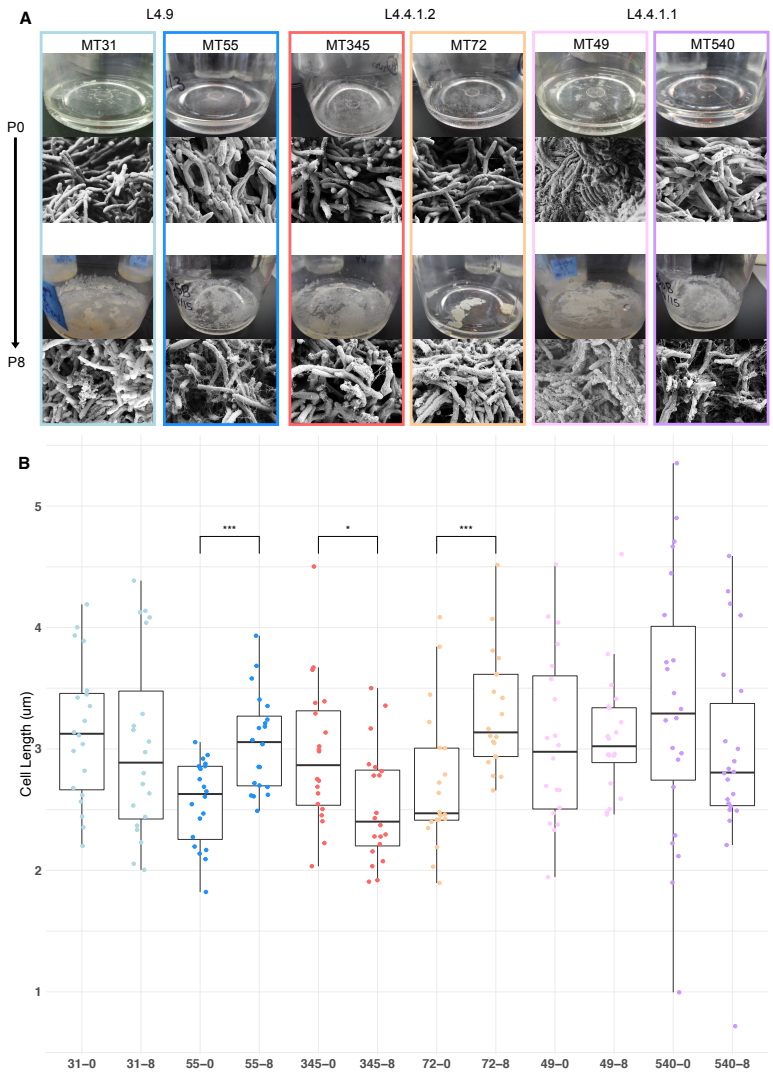


Figure 2: SEM images of ancestral and evolved biofilms show changes in extracellular matrix (A) and cell length (B). SEMs were taken after 5 weeks of biofilm growth, shown alongside photo of pellicle from that passage. Cell lengths were measured across two biological replicates for each strain, except 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 (Mann-Whitney U test with Benjamini-Hochberg correction). *P*-value legend: *: <0.05, **: <0.01, ***: <0.001.

121
122 *Wet weights*
123 To obtain another quantitative measure of phenotypic adaptation to biofilm growth, we developed a
124 protocol for measuring the wet weight of a pellicle biofilm. Wet weights of ancestral biofilms were similar
125 except for MT345, which was heavier in keeping with its confluent morphotype (Figures 1 and 3). We
126 observed a significant (Mann-Whitney U test, *P*-value = 1.5e-06) increase in wet weight for all
127 populations after eight passages (Figure 3). Of note, we observed a substantial increase in wet weight
128 in MT345, which formed a robust pellicle at baseline, suggesting that the potential to increase pellicle
129 biomass remains among high biofilm forming isolates of *M. tb*. Indeed, MT345 along with MT55 had

130 particularly dramatic increases in pellicle biomass: an average of 2.3g and 3.5g, respectively. We also
131 observed a trend towards increased variability of wet weights between replicate cultures of the evolved
132 biofilms (Figure 3).
133

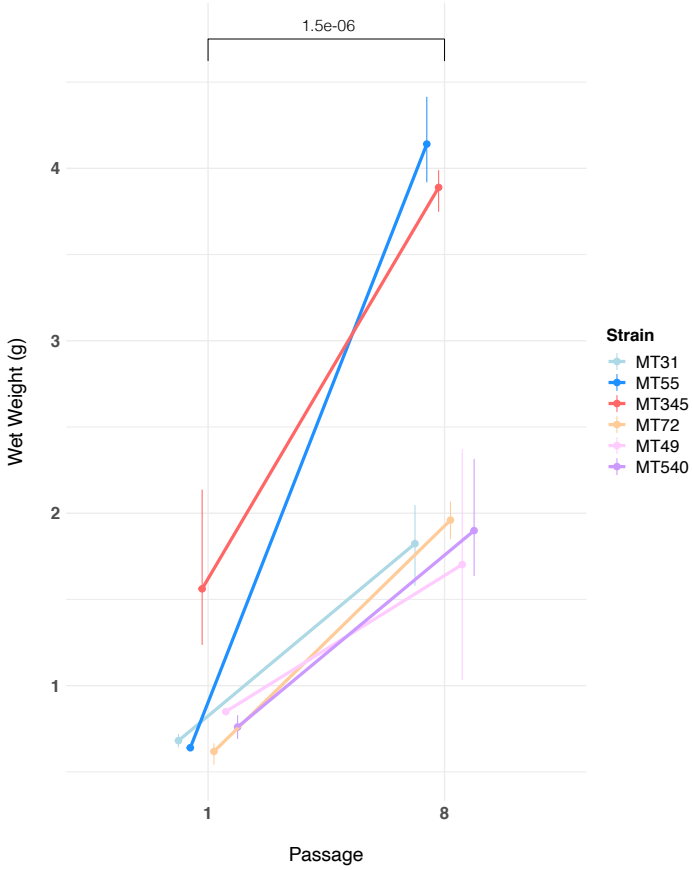


Figure 3: Pellicle wet weights measured in triplicate after 5 weeks of growth, for ancestral populations and populations after eight passages. Pellicle wet weights increased for all genetic backgrounds following passaging (Mann-Whitney U test, *P*-value = 1.5e-06). The magnitude of this change varied, 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.

134
135
136 *Planktonic growth trade-offs during pellicle adaptation*
137 To investigate whether adaptation to pellicle growth involved trade-offs with fitness in planktonic culture,
138 we compared planktonic growth rates of ancestral and evolved populations. Some evolved strains grew
139 faster in planktonic culture following pellicle passaging (MT31, MT345, MT49) whereas other strains
140 exhibited similar growth rates – if slightly different kinetics – when compared to their ancestral strains
141 (MT55, MT72; Figure 4). One evolved strain (MT540) had a slower growth rate than its ancestor (Figure
142 4C; E).

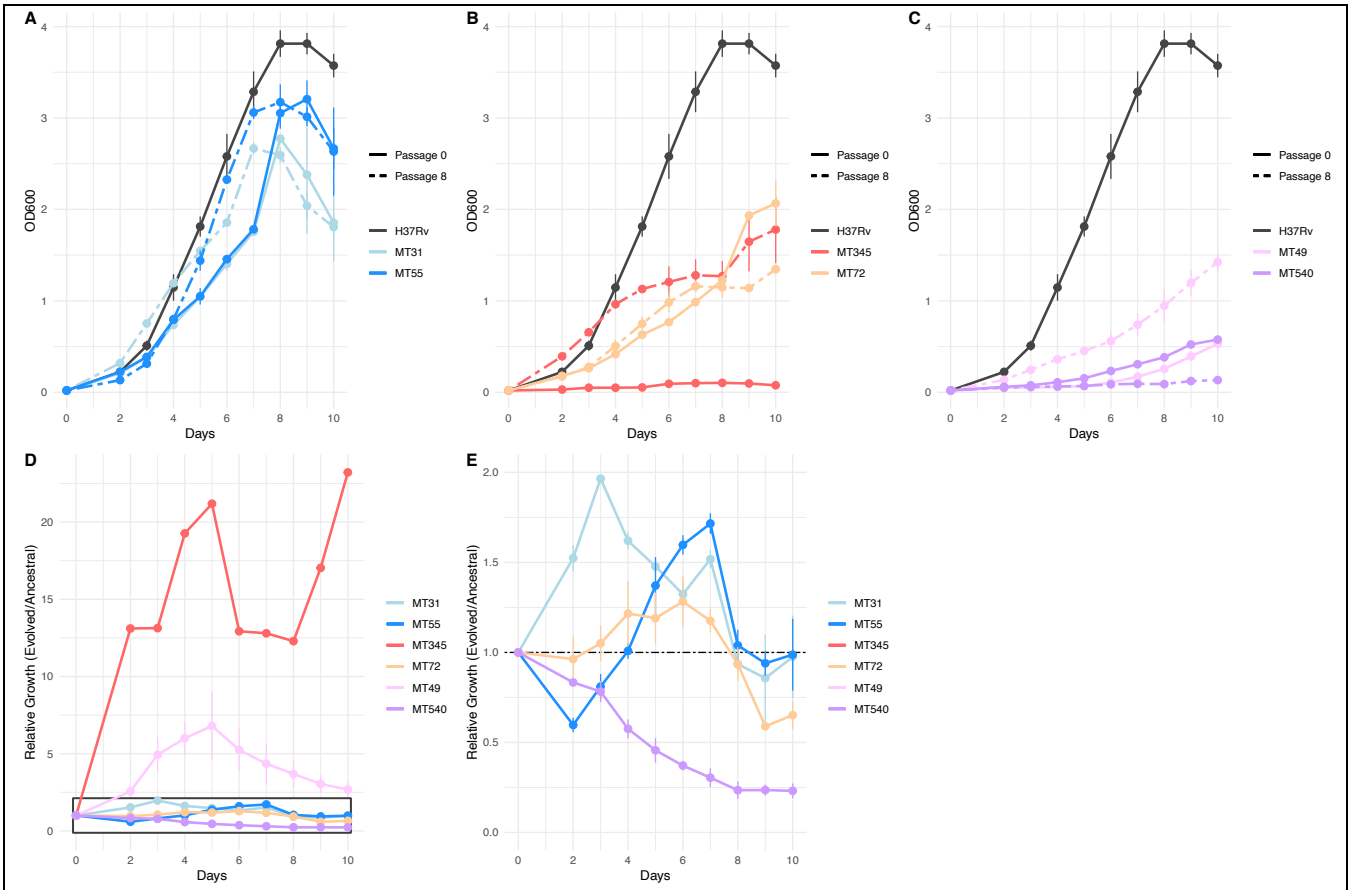


Figure 4: Planktonic growth (A, B, C) and relative fitness (D, E) curves comparing the growth of strains after 8 pellicle passages, to the ancestral strains. Growth curves were performed in triplicate, and the mean OD_{600} value is plotted with error bars representing +/- 1 standard deviation. Relative 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; Supplementary Table 1). 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

156 (Table 1). A search of ~40,000 isolates from a global dataset (see Methods), revealed these mutations
157 to be exceedingly rare or completely absent from the dataset (Table 1). The most commonly identified
158 SNP was at position 3,690,947 upstream of *lpdA* (identified in our study in strain MT49), at a frequency
159 of 0.0015 in this dataset (Table 1). One caveat is that genes in repetitive families (PE, PPE, PE-PGRS)
160 were excluded from our analysis due to difficulties accurately resolving these regions with short-read
161 sequencing data. It is possible that variants contributing to enhanced biofilm formation lie undetected in
162 these regions.
163

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. These mutations are exclusive to strains passaged as pellicles, see Supplementary Table 2 for mutations arising in strains passaged planktonically.

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

164
165
166

167 We did not identify any overlap in mutations that arose following pellicle and planktonic passaging,
168 either in specific SNPs or genes containing SNPs (Table 1, Supplementary Table 2). This suggests that
169 the candidate variants for pellicle adaptation are unlikely to represent generalized lab adaptations.
170

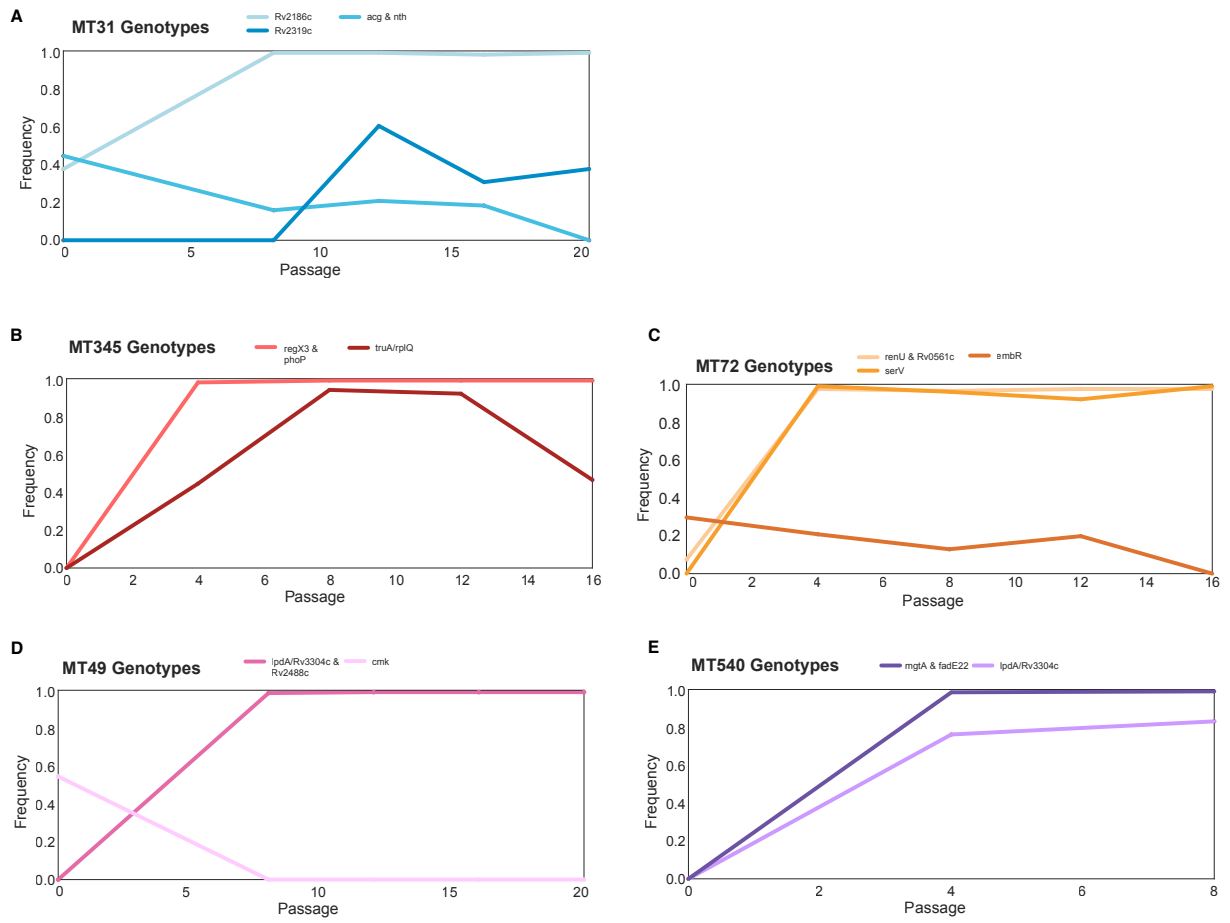


Figure 5: Trajectories of mutations that changed >=30% over the course of the experiment show few mutations arose during passaging. The mutations that became fixed almost always did so in the first or second sequenced passage, and we identified very few mutations with substantial declines in frequency.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

171

172 Allele frequency dynamics for all variants that changed >30% over the course of pellicle passaging are
173 shown in Figure 5. In most populations (MT31, MT345, MT72, MT540) we observed secondary
174 mutations arising on the background of the first fixed mutations (Figure 5). Some secondary mutations
175 reached high frequencies (>95%) but did not remain at high frequency for the duration of the
176 experiment (Figure 5B). We identified a single mutation in the evolved pellicles (Figure 5A, P286L
177 *Rv2319c*) 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 and nucleotide diversity pi) for sequence data at each timepoint (Supplementary Figure 3). Our values concur with previously estimated values for *M. tb* (O'Neill et al., 2015) and are consistent with genome-wide purifying selection (Pepperell et al., 2010, 2013). Bacterial populations within pellicles were more diverse (characterized by higher values of pi) and exhibited a stronger skew to rare variants (lower Tajima's D) than those found in planktonically grown communities (Supplementary Figure 3). This observation points to the complexity of biofilm communities and distinct demographic and/or selective forces associated with pellicle versus planktonic growth.

Convergent adaptation

We observed two instances of convergent adaptation, i.e. the same locus being subject to repeated mutation during pellicle adaptation. 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). Similar patterns were evident in the planktonically passaged populations of bacteria. We identified convergence in these experiments, with mutations arising repeatedly at the same loci (Supplementary Table 2, Supplementary Figure 2). We also observed a likely impact of genetic background on adaptation, with loci subject to repeated mutation in an apparently strain- or sub-lineage specific manner (Supplementary Table 2, Supplementary Figure 2).

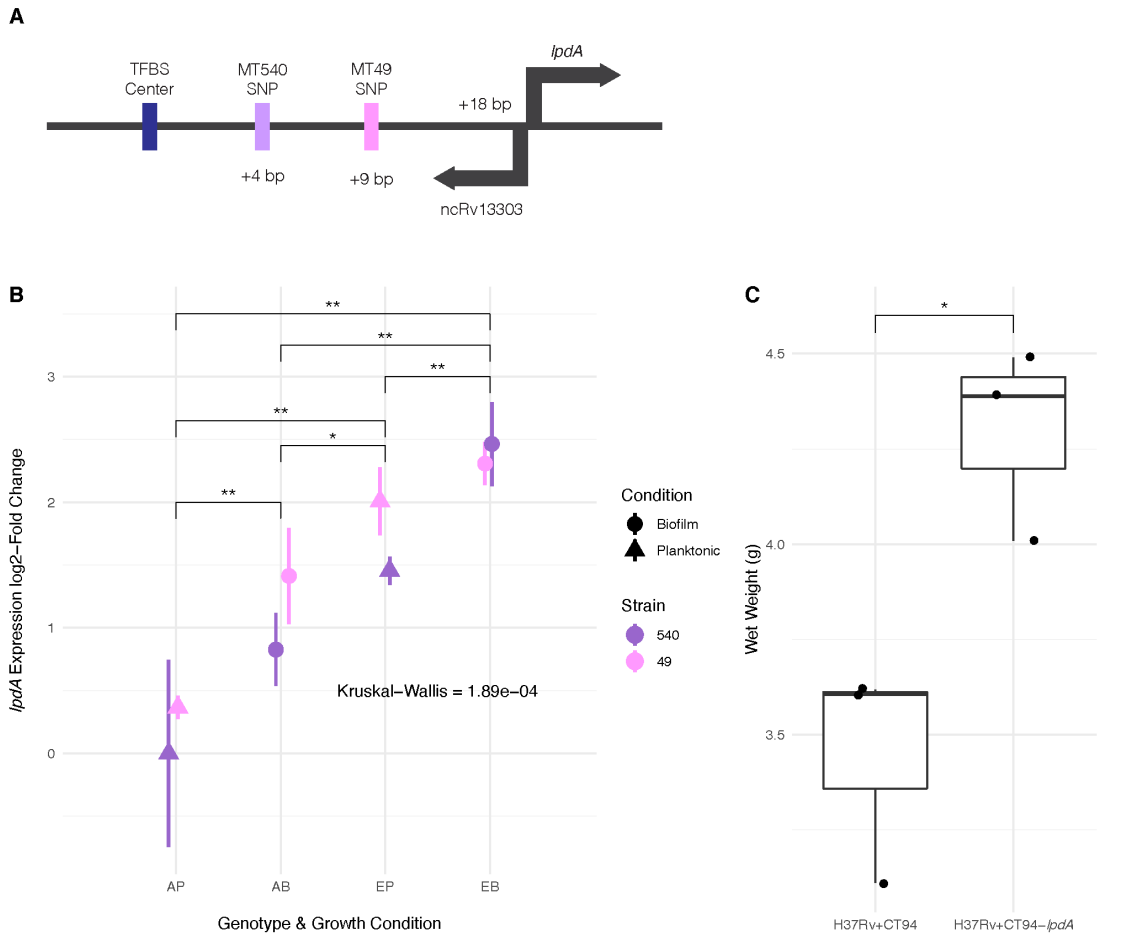


Figure 6: (A) Diagram of convergent adaptation in L4.4.1.1 strains MT540 and MT49 in the region upstream of *lpdA* within a TFBS and ncRNA. **(B)** Expression of *lpdA* differed significantly between sample groups (Kruskal-Wallis test, *P*-value = 1.89e-04). Pairwise comparisons between groups (Mann-Whitney U test with Benjamini-Hochberg correction) revealed significantly increased expression in evolved (E) strains when compared to ancestral (A) strains, as well as during biofilm (B) growth when compared to planktonic (P) growth. **(C)** Overexpression of *lpdA* in H37Rv results in significantly increased biofilm wet weights (*n*=3 for each construct). *P*-value legend: *: <0.05, **: <0.01.

205

206 *Convergent adaptation affecting expression of lpdA*

207 We identified convergent adaptation in evolved pellicles of strains MT49 and MT540 (both of lineage
208 L4.4.1.1) in the region upstream of NAD(P)H quinone reductase *lpdA* (Table 1). These intergenic SNPs
209 are five base pairs apart and lie within the transcription factor binding site (TFBS) for *Rv1719,* which is
210 known to downregulate *lpdA* (Rustad et al., 2014). We hypothesized that a SNP within the TFBS would
211 affect TF binding and therefore expression of *lpdA.* We sought to quantify *lpdA* expression by qPCR in
212 ancestral populations and compare that to expression in evolved populations containing the SNP of

213 interest. Evolved populations had significantly higher (*P*-value <0.05, Mann-Whitney U-test) expression
214 of *lpdA* (Figure 6). Additionally, we observed significantly increased expression of *lpdA* during biofilm
215 growth when compared to planktonic growth, suggesting a role for this gene in biofilm formation (Figure
216 6). In addition to being within a TFBS affecting expression of *lpdA*, these SNPs also lie within a non-
217 coding RNA (ncRNA) that is expressed from the opposite strand of *lpdA* (Figure 6). This ncRNA was
218 identified by RNA-sequencing (Arnvig et al., 2011; Gerrick et al., 2018) but has yet to be characterized.
219 In order to further investigate the role of *lpdA* in pellicle growth, we measured pellicle phenotypes in
220 H37Rv after introducing a second copy of the gene with an integrative, constitutive overexpression
221 plasmid. Pellicles from bacteria with the additional copy of *lpdA* had significantly higher mass than
222 those from bacteria transformed with an empty vector (Figure 6C), with no change in gross phenotype
223 (Supplementary Figure 4).
224

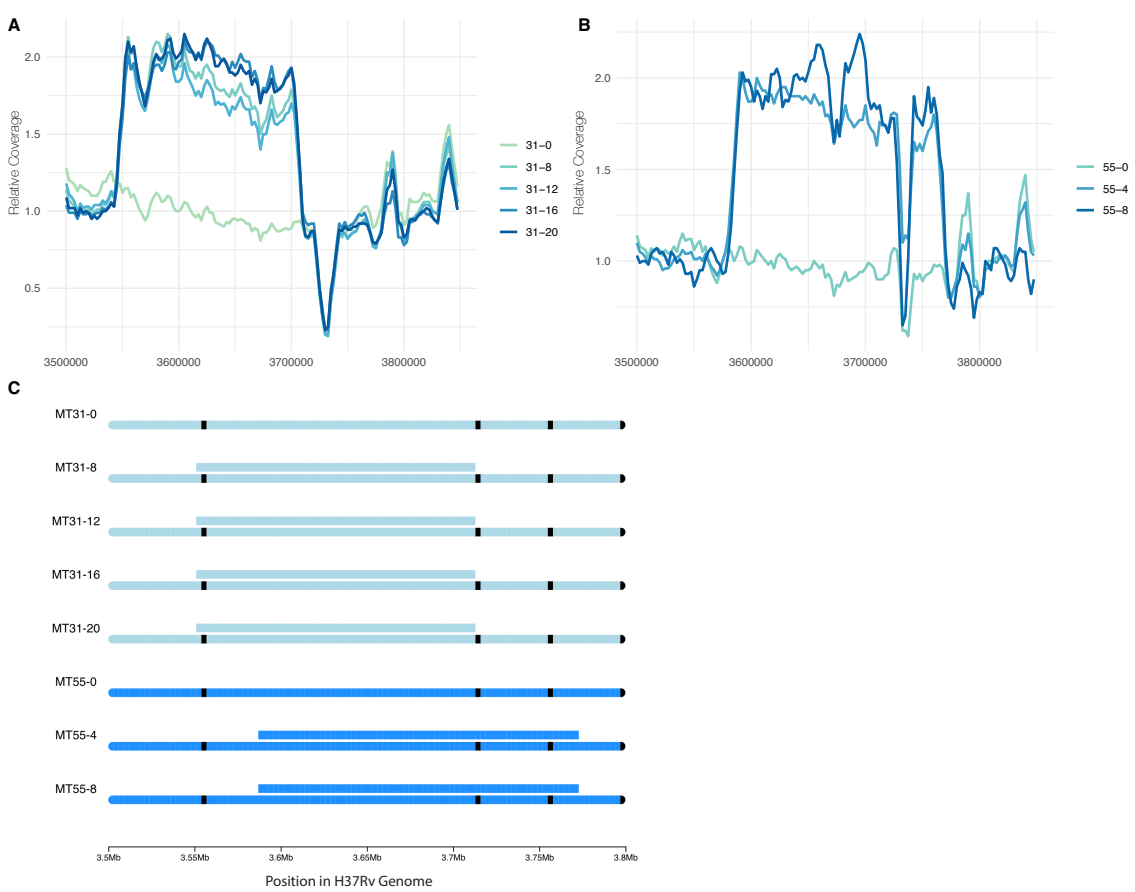


Figure 7: Convergent evolution of a large duplication in L4.9 strains MT31 and MT55, within the same genomic region flanked by mycobacterial IS6110 insertion sequences. Sliding window coverage plots show the increase in coverage in this region, and a chromosome plot shows the duplications' coordinates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

225 *Convergent evolution of large duplication*
226 We discovered a large (~120kb) duplication in pellicles of both strains from lineage L4.9 (MT31 and
227 MT55, Table 1) following pellicle passaging. The duplications were initially identified with Pilon (Walker
228 et al., 2014), and to confirm their coordinates and absence in the ancestral populations, we made
229 sliding window coverage plots in this region and confirmed ~ 2X coverage indicating the presence of a
230 duplication (Figure 7). The coordinates of the duplication appear to be slightly different in the two
231 populations with the duplication in MT55 starting about 3kb after the duplication in MT31 (Figure 7). We
232 determined that the duplication was fixed after the first sequenced passage in each population
233 (passage 4 and 8 for MT55 and MT31, respectively; Supplementary Table 1). After fixation, the
234 duplication appears to have remained stable in both populations, as far out as 20 passages in MT31, as
235 we do not see any reduction in coverage across or within the duplicated region at these later passages
236 (Figure 7). This duplication did not emerge in planktonically passaged isolates of MT31 and MT55,
237 suggesting it has some specificity for pellicle adaptation of these strains (Supplementary Figure 5).
238
239 There are a total of 172 genes within the duplicated regions in MT31 and MT55 (*metU* through *folD*). To
240 identify a possible functional enrichment within these genes we performed a GO term enrichment
241 analysis and compared our results to the results from randomly selected groups of contiguous genes
242 from throughout the genome (see Methods). While the initial analysis identified significant enrichment of
243 genes involved in nucleotide metabolism within the duplication (Supplementary Table 3), this
244 enrichment does not appear significant when compared with randomly selected gene sets
245 (Supplementary Figure 6).
246
247
248
249
250
251
252
253
254
255
256
257
258

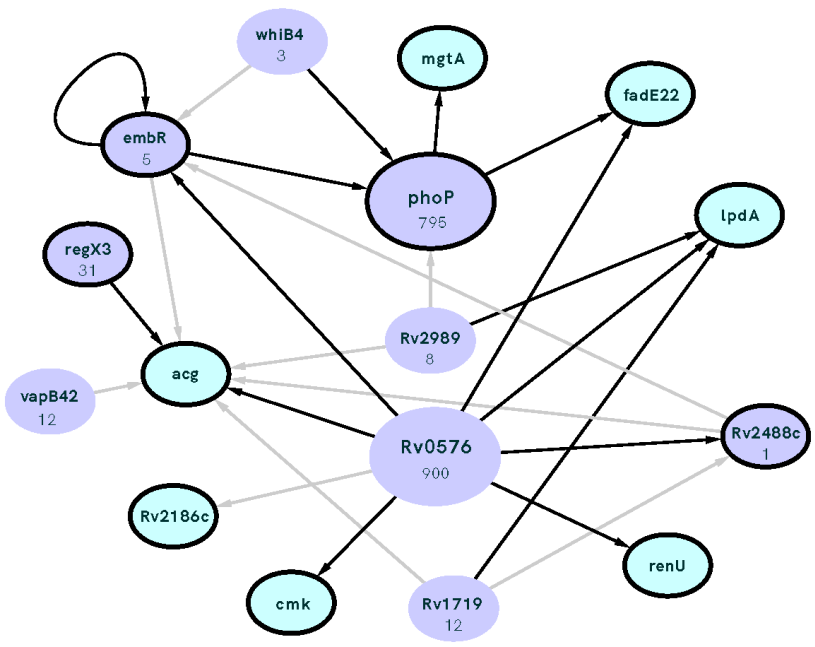


Figure 8: Interconnected regulation of our genes of interest (GOI; outlined in black) by common transcription factors (TF; purple) with the number of genes regulated by each TF listed under gene name. TF identified as having significant overlap with our GOI (i.e. the number of GOI regulated by the TF is higher than expected by chance) using the Mtb Network Portal (see Methods). Edges between nodes indicate >0.5 log2-fold absolute change in expression in a transcription factor overexpression (TFOE) strain, black lines show statistically significant (*P*-value <0.05) differential expression.

Regulatory pathways involved in biofilm formation

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. We identified fourteen genes of interest (GOI) in association with variants that rose to fixation during the passaging experiment (Table 1). Using the MTB Network Portal v2 database of transcription factor overexpression (TFOE) and ChIP-Seq data (http://networks.systemsbiology.net/mtb/) we identified regulatory pathways that contained GOI. In Figure 8 we show the transcription factors (TFs) that are predicted to affect expression of multiple GOI. Of our GOI, *acg*, is the most highly regulated, with seven TFs predicted to affect its expression. The function of *acg* is unknown: it is annotated as a conserved hypothetical, but according to *in vivo* studies may play an important role in infection (Hu and Coates, 2011; Singh et al., 2019). Additionally, we used these data to determine if GOI annotated as regulators (*regX3, phoP, Rv2488c* and *embR*) regulated other GOI (Figure 8). 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* (*M. tb*) 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. tb* 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. tb* 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). MT345 formed robust pellicles at baseline yet manifested a dramatic increase in wet weight following pellicle passaging, indicating that adaptive capacity remains even among high biofilm forming clinical isolates (Figure 3). Collectively, these results demonstrate 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.

313 During our experiment pellicle phenotypes stabilized between passage 2 and 8 (Figure 1); extended
314 passaging beyond 8 passages did not produce further phenotypic changes (Supplementary Figure 1).
315 Fixation of novel and previously segregating mutations occurred over the same time frame (Table 1;
316 Figure 5), and we did not observe any further sweeps over the course of the experiment. Within 8
317 passages we thus saw phenotypic and genotypic stabilization across all populations. This phenomenon
318 is frequently observed during experimental evolution, where rapid adaptation is followed by diminishing
319 returns after continued passaging (Barrick et al., 2009; Khan et al., 2011; Philippe et al., 2007).
320
321 We observed few mutations over the course of the experiment (Table 1), and some of these are likely
322 to be hitchhikers linked to advantageous alleles that swept to fixation (Table 1, Figure 5). These few
323 mutations appeared to have pleiotropic effects, as we observed changes in phenotypes other than the
324 one under selection. 'Off target' phenotypes, such as planktonic growth rate and cell length, varied
325 among strains. Experimentally evolved populations generally manifest first-pass, large-effect mutations
326 during initial adaptation; our results suggest that these mutations in *M. tb* are broad in their scope and
327 variable in their effects. The latter phenomenon may be due to a large genetic target for biofilm growth,
328 the impacts of hitchhiking mutations and/or genetic background of the strains.
329
330 Candidate single nucleotide polymorphisms and deletions mediating pellicle adaptation were either
331 exceedingly rare or absent in our search of a database of almost 40,000 *M. tb* genomes. This suggests
332 that the loci hit in our experiment are under strong purifying selection in *M. tb*'s natural environment:
333 that is, the mutations we observed are associated with substantial fitness costs to the bacterium during
334 human infection. This is not to suggest that biofilm growth is irrelevant to natural infection; as discussed
335 above, *M. tb* biofilms and biofilm biomarkers have been identified in clinical specimens (Canetti, 1955;
336 Nyka, 1977, 1967, 1963; Nyka and O'Neill, 1970; Chakraborty et al., 2021). Rather, we hypothesize
337 that the mutations we observed are selected against because they interfere with regulatory pathways
338 that normally constrain biofilm growth to specific environmental conditions. Regulatory constraints
339 would limit potential drawbacks of biofilm formation, such as the metabolic costs of matrix production
340 and immunostimulation. Another, not mutually exclusive, explanation is that fitness costs associated
341 with mutations observed here are due to effects on off-target traits such as planktonic growth rate, cell
342 size changes, and/or phenotypes unmeasured here. Experiments such as this one can illuminate
343 impermissible paths in *M. tb* adaptation and reveal specific vulnerabilities with the potential to be
344 exploited in TB therapeutics.
345

346 Two strains in our study evolved a large (~120kb) duplication in the same genomic region, which fixed
347 by passage 8 (Table 1, Figure 7). The duplication persisted without evidence of decay in extended
348 passaging in our study (up to passage 20 in MT31). Duplications in this region of the genome have
349 been identified previously in isolates from Lineages 2 and 4 (Domenech et al., 2010, 2014a; Weiner et
350 al., 2012). Duplications at this site are likely facilitated by flanking *IS6110* elements (Figure 7C) which
351 can induce unequal homologous recombination. Prior research has shown that the duplication confers
352 a fitness disadvantage *in vivo* (as measured in a mouse model) and can be selected during axenic
353 culture (Domenech et al., 2014a).
354
355 Here we find this lab adaptation to be associated with enhanced pellicle formation. Interestingly, while
356 the duplication was associated with increases in planktonic growth rate, we did not identify it in
357 planktonically passaged populations (Figure 4, Supplementary Figure 5). Domenech *et al* found
358 previously that emergence of a similar duplication was variable among strains and culture conditions
359 (Domenech et al., 2014b). Together, these data suggest that while the duplication is observed
360 commonly in the lab, it is selected on specific genetic backgrounds and in response to specific
361 environmental conditions. Large-scale genome rearrangements including duplications have previously
362 been identified as a rough adaptive 'first draft' followed by further refinement during experimental
363 evolution (Lynch and Conery, 2000; Zhang, 2003). The duplication was the only mutation that fixed in
364 MT55, which exhibited a dramatic increase in pellicle weight (Figure 3) and increased production of
365 ECM (Figure 2) following passaging, suggesting that it is indeed this seemingly multipurpose mutation
366 that is responsible for pellicle adaptation. The apparent impacts of the duplication were not limited to
367 pellicle growth, as MT55 also exhibited an increase in cell length (Figure 2) and a faster rate of growth
368 during planktonic culture, as did its sister strain MT31 also encoding the mutation (Figure 4). Taken
369 together, our findings suggest that the production of ECM and other aspects of biofilm growth are tightly
370 regulated during natural infection, and that first pass mutations outside of this environment result in de-
371 repression of biofilm production. Our findings further suggest that biofilm formation is co-regulated with
372 a range of phenotypes, consistent with prior research on biofilms, persistence and drug tolerance
373 (Richards et al., 2019).
374
375 *Epistasis*
376 The 120kb duplication occurred on two strains from the same sub-lineage (MT31 and MT49 from L4.9)
377 and did not occur in any other genetic background. A similar phenomenon emerged in association with
378 sub-lineage L4.4.1.1, where both strains underwent mutation at a transcription factor binding site
379 upstream of *lpdA*, with attendant impacts on gene expression. While our strains are all very closely

380 related, there are still distinct genetic variants for each sub-lineage that may affect the evolutionary
381 trajectory and lead to convergent adaptation of strains within (Supplementary Data 1). We observed
382 similar patterns in the planktonically passaged populations, for example in repeated mutations
383 occurring in an apparently sub-lineage- (*tuf*, L4.1.1.2) or strain-specific manner (Supplementary Table
384 2, Supplementary Figure 2). Repeated mutation at these loci provides strong support for their
385 involvement in the trait under selection. In addition, the association of candidate adaptive mutations
386 with specific sub-lineages suggests that *M. tb* strains' genetic backgrounds played a role in shaping the
387 adaptive path to enhanced biofilm growth.
388
389 Epistasis, which refers to the phenotypic impact of interactions among mutations at distinct loci, has
390 been observed previously during experimental evolution of microbes [e.g. (Fisher et al., 2019)]. Sign
391 epistasis is a form of epistasis in which a mutation is advantageous on one genetic background and
392 deleterious on another (Weinreich et al., 2005). The presence of sign epistasis can hinder access to a
393 fitness optimum, and thus it represents a constraint on adaptation (Weinreich et al., 2005; Poelwijk et
394 al., 2007). Our identification of mutations only in association with specific sub-lineages is consistent
395 with sign epistasis, as is the observation of parallel adaptation in this small sample of strains: it
396 suggests, for example, that mutations upstream of *lpdA* may be deleterious on some genetic
397 backgrounds, and/or that alternatives are deleterious on the L.4.1.1.1 background. A careful review of
398 the existing literature reveals additional evidence of sign epistasis in *M. tb*. For example, we and others
399 have found evidence suggesting that *M. tb* lineage affects the mutational path to drug resistance
400 (Mortimer et al., 2018; Farhat et al., 2019; Castro et al., 2019). Even more striking, Manson *et al*
401 identified a stereotypic order of resistance mutation acquisition in a large survey of drug resistant *M. tb*
402 isolates, a finding that was subsequently replicated (Manson et al., 2017; Ektefaie et al., 2021). This
403 suggests that the acquisition of an initial resistance mutation alters the fitness associated with
404 subsequent mutations, as has been observed in other settings (Silva et al., 2011). More direct evidence
405 for sign epistasis comes from mutagenesis experiments, in which the lethality of gene knockouts has
406 been shown to vary according to *M. tb* strain genetic background (Maksymiuk et al., 2015; Carey et al.,
407 2018).
408
409 *Upregulation of lpdA during biofilm growth*
410 As noted above, we identified convergent adaptation in the region upstream of *lpdA* (*Rv3303c*) on
411 L.4.1.1.1 backgrounds. *lpdA* is annotated as a NAD(P)H quinone reductase, which is known to be
412 expressed *in vivo* (Deb et al., 2002) and to contribute to virulence in mice (Akhtar et al., 2006). It has
413 not previously been linked with biofilm formation. The mechanism of its contribution to virulence *in vivo*

414 is thought to be by protection against oxidative stress (Akhtar et al., 2006). The strains in our study
415 evolved variants 5 bp apart, both of which lie within the TFBS for *Rv1719* (Figure 6A). These SNPs also
416 lie within ncRNA *ncRv13303c*, which was originally identified by RNA sequencing (Arnvig et al., 2011;
417 DeJesus et al., 2017) but is yet uncharacterized. Previous work characterizing variable-number tandem
418 repeats (VNTR) in this intergenic region have proposed a possible hairpin structure that lies within
419 *ncRv13303c* and could affect its function (Zheng et al., 2008). Differences in VNTR copy number and
420 promoter sequence in this region have been observed between different strains of mycobacteria, and
421 even linked to differences in expression of *lpdA* (Akhtar et al., 2009; Pérez-Lago et al., 2013; Zheng et
422 al., 2008).
423
424 We hypothesized that SNPs in the TFBS upstream of *lpdA* would affect TF binding and expression of
425 *lpdA*. This was indeed the case, as evolved strains had significantly higher expression of *lpdA* (Figure
426 6B). Additionally, we found expression of this gene to be higher during biofilm growth when compared
427 to planktonic growth. Further support for a link between increased expression of *lpdA* and enhanced
428 pellicle growth comes from our data showing an increase in pellicle wet weight following introduction of
429 a second copy of the gene via an integrative plasmid (Figure 6C). Interestingly, *lpdA* is within the
430 bounds of the large duplications we identified in strains MT31 and MT55, which may affect the dosage
431 of this gene in those evolved strains.
432
433 The two SNPs upstream of *lpdA* were exceedingly rare among clinical isolates of *M. tb*: in a search of
434 almost 40,000 *M. tb* genomes, we identified the MT49 SNP in just 61 genomes and the MT540 SNP
435 was found in 4 genomes (Table 1). Interestingly, the MT49 mutation was shown to sweep to fixation in
436 a deep sequencing study of serial sputum samples chronicling *M. tb* within host dynamics (Trauner et
437 al., 2017). This supports the hypothesis that biofilms are important at a specific phase of TB infection,
438 and that enhanced biofilm growth can also occur *in vivo* through increased expression of *lpdA* and
439 genes co-regulated with it. Taken together, these observations suggest that variants at the TFBS
440 upstream of *lpdA* can be transiently selected during infection but are otherwise deleterious, possibly
441 due to fitness costs associated with biofilm growth that is timed inappropriately.
442
443 *The regulatory system as a target of selection in M. tb*
444 The L4.9 duplication and SNPs upstream of *lpdA* presumably act via regulation/gene dosage.
445 Regulatory activity was also common among the other loci that accumulated mutations in our
446 experiment, including phosphorelay signal transduction and DNA binding activity (Table 1). We
447 identified fixed SNPs in four transcriptional regulators: *regX3, phoP, embR* and *Rv2488c* (Table 1).

448 These genes in turn regulate other candidate genes identified in our study (Figure 8). *regX3* and *phoP*
449 belong to two-component regulatory systems (2CRS), whereas *embR* and *Rv2488c* are of the ompB
450 and luxR families, respectively. Many of *M. tb*'s 2CRS have been implicated in virulence and survival in
451 host conditions (Li et al., 2019) and this is true of *regX3*, which with its partner *senX3* is required for
452 virulence in mice (Li et al., 2019; Parish et al., 2003). Of note, *regX3* regulates genes that are known to
453 be involved in *M. tb* biofilm development (Richards et al., 2019). Similarly, while *Rv2488c* has not
454 previously been directly implicated in *M. tb* biofilm growth, *luxR* regulators are involved in quorum
455 sensing in gram negative bacteria and have been hypothesized to play a similar role in mycobacteria
456 (Chen and Xie, 2011; Sharma et al., 2014). Our results provide support for further investigation of
457 quorum sensing in *M. tb* and its relationship to biofilm development.
458
459 Several studies have demonstrated that clinical isolates of *M. tb* vary in their methylation patterns, and
460 methylome phenotypes have been linked with allelic variants in cognate methyltransferases (Modlin et
461 al., 2020; Phelan et al., 2018; Shell et al., 2013; Zhu et al., 2015). Methyltransferase motifs are
462 associated with regulatory elements such as promoters, and variation in methylation patterns has been
463 associated with differences in gene expression in some cases (Chiner-Oms et al., 2019; Gomez-
464 Gonzalez et al., 2019; Modlin et al., 2020). We did not observe any mutations at loci known to encode
465 methyltransferase genes, but it is possible that adaptation to pellicle growth was accompanied by
466 epigenomic changes, or that epigenomic changes occurring by an unknown mechanism contributed to
467 the phenotypic changes we observed. Such epigenomic changes would have to be relatively stable, as
468 the phenotypic changes we observed persisted following growth under standard planktonic conditions.
469
470 Broadly speaking, our results show that selection in our system appears to have targeted genes that
471 regulate many other genes (hubs, e.g. *phoP*), and genes under regulation by many genes (e.g. *acg*),
472 whose expression is likely responsive to a broad range of conditions. This reflects general features of
473 *M. tb's* regulatory network, which exhibits a high degree of connectivity between regulatory pathways
474 (Chauhan et al., 2016; Galagan et al., 2013) as well as a hierarchical structure in which master
475 regulators can rapidly calibrate global patterns of gene expression (Chauhan et al., 2016; Parvati Sai
476 Arun et al., 2018). Similar findings have also been obtained in other systems, where experimentally
477 evolved populations have shown adaptation in transcription factors and other global regulators, with
478 global impacts on gene expression (Ali and Seshasayee, 2020; Conrad et al., 2010; Philippe et al.,
479 2007; Rodríguez-Verdugo et al., 2016; Saxer et al., 2014). These mutations occur quickly to enable
480 rapid adaptation to new environments, and it is hypothesized that secondary selection may act to refine
481 gene expression after the initial burst of adaptation (Ali and Seshasayee, 2020; Rodríguez-Verdugo et

482 al., 2016). Thorpe et al previously found evidence of positive selection at promoter sites in natural
483 populations of *M. tb*, suggesting that adaptation in natural as well as experimental populations is
484 facilitated by regulatory mutations (Thorpe et al., 2017). The broad potential impacts of regulators such
485 as *regX3, phoP, embR* and *Rv2488c* and large-scale duplications likely helps to explain the wide range
486 of phenotypic changes we observed. Further passaging of our isolates will reveal whether these
487 adaptations can be further refined or whether pellicle growth remains inextricably linked to a range of
488 phenotypes.
489
490 Of note, none of the genes identified in our study have previously been implicated in biofilm
491 development, except for *renU/mutT3*. Interestingly, previous research (Wolff et al., 2015) found that
492 *renU* knockouts were defective in pellicle growth, whereas, in our study, a frameshift mutation in *renU*
493 was associated with enhanced biofilm growth (MT72). The novelty of our findings highlights the
494 potential for experimental evolution to complement more traditional approaches of interrogating
495 important pathogen phenotypes, particularly in elucidating complex interactions and the phenotypic
496 impacts of subtle genetic changes.
497
498 From some perspectives, the adaptation of *M. tb* seems impossibly constrained: subject to genetic drift
499 imposed by its host's unpredictable behavior, subject to the limitations of a fully linked genome,
500 navigating a fitness landscape rendered complex by powerful epistatic interactions. Yet *M. tb* adapts
501 rapidly in both natural and experimental settings. *M. tb*'s complex regulatory architecture exhibits many
502 of the features that characterize robust systems – such as redundancy, modularity, and multiple
503 feedback mechanisms (Kitano, 2004) – and we hypothesize that it is among the reasons *M. tb*
504 continues to persist and thrive across the globe.
505

506 **Materials and Methods**
507
508 *Bacterial strains and growth conditions.*
509 Clinical strains were initially isolated from sputum
510 samples. Briefly, sputum samples were decontaminated
511 and struck on Löwenstein–Jensen slants. *Mycobacterium*
512 *tuberculosis* growth was inoculated into Middlebrook 7H9
513 broth (HiMedia) containing 0.2% w/v glycerol, 10% v/v
514 OADC supplement (oleic acid, albumin, D-glucose,
515 catalase; Becton Dickinson) and 0.05% w/v Tween-80
516 (7H9OTG) and incubated at 37^0C with 5% CO_2 for 3-5
517 weeks. Cultures were sub-cultured in 7H9OTG prior to
518 the start of this experiment. Six clinical strains were
519 selected for experimental passage (MT31, MT49, MT55,
520 MT72, MT345, and MT540). The terminal culture
521 (Passage 0) of each strain was grown to an OD_{600} ~1 in
522 planktonic culture and inoculated into media for biofilm growth (Passage 1; Figure 9). Aliquots of
523 passage 0 were frozen, and genomic DNA was isolated for sequencing. For planktonic growth, 250µL
524 was inoculated into 5 mL 7H9OTG, in duplicate, and incubated at 37^0C with shaking, until grown to an
525 OD_{600} ~1. For biofilm growth, 250µL of planktonic culture was inoculated into 25mL Sauton's medium
526 (for 1L: 0.5g KH_2PO_4, 0.5g $MgSO_4$, 4g L-Asparagine, 2g Citric acid, 0.05g Ammonium Iron (III) citrate,
527 60mL glycerol, adjust pH to 7.0 with NaOH) containing 0.1% w/v $ZnSO_4$, in duplicate, and incubated at
528 37^0C, with 5% CO_2, without shaking for 5-7 weeks. Biofilms were grown as previously described (Kulka
529 et al., 2012): 250mL bottles (Corning, 430281) were incubated with a tight-fitting cap for 3 weeks, and
530 then with a loose-fitting cap for the remainder of growth.

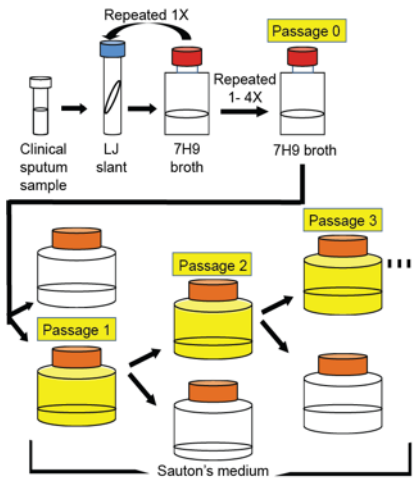


Figure 9: Protocol for serial passage of *M. tb* pellicles

531
532 *Experimental evolution of pellicles*
533 As illustrated in Figure 9, duplicate populations of 6 clinical strains were grown as a biofilm. After 5-7
534 weeks, the more robust of the two pellicles was selected for passage. 0.3 g (wet weight) of pellicle was
535 inoculated into 25mL Sauton's in duplicate and grown as previously described. The pellicle not selected
536 for passage was discarded. Between 8 and 20 total pellicle passages per strain were performed. Every
537 4 passages culture were frozen at -80^0C and gDNA was extracted for sequencing (excluding MT31 and
538 MT49, for whom the first passage that was frozen and had gDNA extracted was Passage 8;
539 Supplementary Table 1).
540

541 *Experimental evolution of planktonic cultures*
542 Starter cultures of ancestral strains were made from frozen cryovial stocks and sequenced to get an
543 ancestral baseline for the planktonic passages. Each strain was inoculated in triplicate to form 3
544 independently evolving populations: A, B and C. For each passage, 25mL of 7H9OTG was inoculated
545 with 500µL of culture at OD_{600} = 1. Cultures were grown at 37^0C with shaking until an OD_{600} ~1, before
546 freezing cryovials, extracting gDNA for sequencing, and seeding the next passage. A total of 15/18 of
547 the populations were passaged 4 times, while 3/18 were only passaged 3 times due to the presence of
548 contamination.
549
550 *Pellicle phenotypes and wet weights*
551 Pellicle growth was photo documented and phenotyped after 5-7 weeks of growth during the course of
552 the experiment. Wet weights were taken after regrowing both ancestral and Passage 8 strains from
553 cryovial stocks, as described above, inoculating first into 7H9OTG and then into Sauton's either in
554 duplicate or triplicate. Wet weight measurements were taken after 5 weeks of pellicle growth by
555 removing the spent media beneath the pellicle and weighing. Then, the tare weight of the empty bottle
556 was subtracted.
557
558 *SEM*
559 For scanning electron microscopy experiments, samples were taken from *M. tb* pellicle cultures after 5
560 weeks of growth and placed on poly-L-lysine-treated plastic coverslips (13 mm, Thermonax plastic for
561 cell culture). Samples were then fixed overnight in a 4% formaldehyde, 2% glutaraldehyde solution in
562 Dulbecco's phosphate-buffered saline (-calcium, -magnesium) (DPBS) (Hyclone Laboratories Inc.,
563 Logan, UT). Following fixation, samples were washed with DPBS, then treated with 1% osmium
564 tetroxide for 1 hour. Following osmium tetroxide treatment, samples were washed with DPBS. Next,
565 cells underwent ethanol dehydration, which was followed by critical point drying. Following that,
566 samples were then placed on aluminum stubs and sputter-coated with 20 nm platinum. The samples
567 were imaged at 3 kV by a Zeiss GeminiSEM 450 scanning electron microscope.
568
569 To measure cell length via scanning electron microscopy, SEM images taken at 10,000X were opened
570 on ImageJ. The pixel length of the scale bar was measured with the "straight" tool. Next, the image was
571 cropped to remove any area without *M. tb* cells. Six frames per image were selected using the
572 "unbiased frames" macro in ImageJ. The cell length in pixels of a fully exposed *M. tb* cells was
573 measured in each frame using the "straight" tool. A random number generator was used to select
574 frames for the remaining four cell lengths. After all cell lengths were measured, pixel length was

575 converted to µm using the scale bar prior to analysis. Cell lengths were measured across two biological
576 replicates (two independent cultures) for all strains except SK540-0, where only one replicate was
577 available.
578
579 *Planktonic growth curves*
580 To compare rates of growth between ancestral and evolved populations, we inoculated 5 mL 7H9OTG
581 with 250 µL thawed culture from a previously frozen cryovial for each of the ancestral strains and
582 passage 8 of each of the pellicle evolved strains. Cultures were incubated at 37^0C with shaking, until
583 grown to an OD_{600} ~ 1. 15mL fresh 7H9OTG was seeded at a starting OD_{600} of 0.02 in triplicate and
584 incubated at 37^0C with shaking. OD_{600} measurements were taken every day for 10 days (starting at day
585 2).
586
587 *DNA extraction*
588 gDNA for whole genome sequencing was isolated by following a modified Qiagen DNeasy Blood and
589 Tissue Kit protocol. Briefly, cultures were pelleted and resuspended in TE buffer in 2 mL screw-cap
590 tubes containing ~250µg of 0.1mm glass beads. The samples were heated to 80^0C for 50 minutes to
591 sterilize the suspensions, then frozen at -80^0C for at least 1 hour to aid cellular lysis. Tubes were
592 vortexed for 3 minutes for mechanical lysis. Suspensions were then incubated with 3% w/v mutanolysin
593 (Sigma-Aldrich, M9901) for 1 hour at 37^0C. Beads and cellular debris were pelleted, and supernatants
594 were transferred to tubes for DNA purification following the protocol detailed in the Qiagen DNeasy
595 Blood and Tissue Handbook, starting with step 4 of "Protocol: Pretreatment for Gram-Positive Bacteria."
596
597 *Library preparation and sequencing*
598 Library preparation was performed using a modified Nextera protocol as described by (Baym et al.,
599 2015) with a reconditioning PCR with fresh primers and polymerase for an additional 5 PCR cycles to
600 minimize chimeras and a two-step bead based size selection with target fragment size of 650 bp.
601 Libraries were submitted to the University of Wisconsin-Madison Biotechnology Center (UWBC).
602 Quality and quantity of the finished libraries were assessed using an Agilent DNA High Sensitivity chip
603 (Agilent Technologies, Santa Clara, CA) and Qubit dsDNA HS Assay Kit, respectively. Libraries were
604 standardized to 2 µM. Samples were sequenced generating paired end reads using Illumina HiSeq
605 2000. The majority of samples (87/101) were sequenced using the above protocol, however for a small
606 number of samples that needed resequencing, gDNA was submitted to the UWBC and to the Microbial
607 Genome Sequencing Center (MiGS) for both library prep and sequencing using Illumina NovaSeq 6000
608 and Illumina NextSeq 2000, respectively.

609 *Reference guided assembly and variant calling.*
610 Genome assemblies were performed using an in-house reference-guided assembly pipeline
611 (https://github.com/pepperell-lab/RGAPepPipe). Briefly, raw data was checked for quality with FastQC
612 v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed using TrimGalore
613 v0.6.4 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads were mapped to the
614 H37Rv reference genome using BWA-MEM v0.7.12 (Li, 2013), and Samtools v1.11 (Li et al., 2009)
615 view and sort were used to process SAM and BAM files. Picard v1.183
616 (http://broadinstitute.github.io/picard/) was used to remove duplicates and add read information. Reads
617 were locally re-aligned using GATK v3.5 (DePristo et al., 2011) and variants were identified using Pilon
618 v1.16 (Walker et al., 2014). Finally, assembly quality was determined using Qualimap BamQC
619 (Okonechnikov et al., 2016).
620
621 *Lineage Typing and Phylogenetic tree*
622 We used SNP-IT (Lipworth et al., 2019) to identify the lineages of our ancestral strains. To identify the
623 sub-lineage of our ancestral strains we obtained an alignment of known L4 isolates from a collaborator.
624 We created a SNP alignment of the assemblies from the RGAPepPipe using Snp-sites v 2.4.1 (Page et
625 al., 2016) and constructed a maximum likelihood phylogeny using RAxML v8.2.3 (Stamatakis, 2014).
626 Next, we visualized the phylogeny using Dendroscope (Huson and Scornavacca, 2012) and lineage
627 typed our isolates by comparing their location in the phylogeny to known L4 subtypes. The phylogenetic
628 tree in Figure 1 was plotted and colored in R using ggtree (Yu et al., 2017).
629
630 *Identification of fixed SNPs*
631 To identify genetic changes occurring after repeated passaging as a pellicle or planktonic culture, we
632 used Popoolation2 v1.201 (Kofler et al., 2011b) and breseq v0.35.0 (Deatherage and Barrick, 2014).
633 BAM files from RGAPepPipe (see above) were filtered for quality and used with Samtools mpileup (Li et
634 al., 2009) to tabulate mutation frequencies across sequenced timepoints. The mpileup file was
635 converted to a sync file using Popoolation2 and finally converted the sync file to a format suitable for
636 downstream analysis using Python (code available at https://github.com/myoungblom/mtb_ExpEvo).
637 Variants were filtered for quality by minimum coverage of 20X, minimum alternate allele count of 5 and
638 a minimum allele frequency of 5%. We also removed variants in repetitive families of genes (PE, PPE,
639 and PE-PGRS gene families) from analysis. Sequence data were also assembled and variants called
640 and annotated using breseq v0.35.0 (Deatherage and Barrick, 2014) to confirm variants called using
641 Popoolation2. Breseq was run in population mode for each sequenced timepoint, with all default
642 parameters except one: “polymorphism-minimum-total-coverage-each-strand” was set to 4. Then the

643 mutational frequencies of each timepoint were compared using gdtools (Deatherage and Barrick, 2014)
644 compare run with default parameters. To identify mutations that had significant frequency changes
645 throughout the experiment we looked for variants that were at a low starting frequency and rose to
646 fixation (>95% frequency) or conversely, started at intermediate frequency and disappeared from the
647 population (<5% frequency).
648
649 *Structural variant detection*
650 To identify potential insertions, deletions, and duplications of interest in our genomic sequence data, we
651 called structural variants using Pilon v1.16 as a part of RGAPepPipe. We removed all insertions,
652 deletions, and duplications that were located entirely in regions annotated as repetitive genes (PE,
653 PPE, and PE-PGRS gene families). We also removed variants that were identified in ancestral strains
654 and/or that did not remain fixed in subsequent passages. Additionally, we manually curated potential
655 variants by visually inspecting the alignments in IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013)
656 and removed variants that appeared to be mis-mappings or in regions of poor coverage. Sliding window
657 coverage plots to visualize duplications were made using Samtools bedcov (Li et al., 2009) using in-
658 house scripts (code available at https://github.com/myoungblom/mtb_ExpEvo). Relative coverage was
659 calculated by dividing each window coverage by the average coverage across the assembly – as given
660 by BamQC (Okonechnikov et al., 2016) – and plotted in R.
661
662 *GO term enrichment analysis*
663 To identify possible functional enrichments in the genes within the large duplication that evolved in
664 MT31 and MT55 pellicles (Table 1), we performed a GO term enrichment analysis. GO terms for all *M.*
665 *tb* genes were downloaded from Quick GO (https://www.ebi.ac.uk/QuickGO/). We used the R package
666 topGO (https://bioconductor.org/packages/release/bioc/html/topGO.html) to perform an enrichment
667 analysis of the genes present in one or both duplications (a total of 172 genes). We calculated
668 significance using a Fisher's exact test and filtering for *P*-values < 0.01 (Supplementary Table 3). We
669 performed the same enrichment analysis on 100 randomly selected, contiguous sets of 172 genes to
670 test the greater significance of our results (Supplementary Figure 6).
671
672 *Genotype plots*
673 To better understand the structure of the population over passaging we created genotypes by clustering
674 variants with similar trajectories throughout the experiment. Output from Popoolation2 as described
675 above, was filtered for based on >= 30% frequency change between the final passage and ancestral
676 populations, as well as filtering out all mutations within 1000bp of each other (we identified that these

677 mutations are often the result of repetitive regions and/or poor mapping). Lolipop
678 (https://github.com/cdeitrick/Lolipop) was used with default parameters to cluster mutations into
679 genotypes and create plots.
680
681 *Population genetics statistics*
682 Population genetics estimates (Tajima's D, pi and theta) were calculated as previously described
683 (O'Neill et al., 2015). Briefly, we used Popoolation (Kofler et al., 2011a) to estimate these parameters in
684 sliding windows across the genome. To reduce biases caused by variable coverage we randomly sub-
685 sampled read data without replacement to uniform coverage of 50X; we repeated this process 9 times
686 to further reduce biases. Genome-wide averages were calculated by averaging all windows across the
687 genome, across all replicate sub-sampled data. Scripts available at
688 https://github.com/myoungblom/mtb_ExpEvo.
689
690 *Network analysis*
691 Genes of interest (GOI) were defined as coding genes containing or just downstream of the fixed
692 mutations, we identified in populations passaged as pellicles (Table 1). Network analysis of GOI was
693 performed by accessing the transcription factor overexpression data available in the MTB Network
694 Portal (http://networks.systemsbiology.net/mtb/content/TFOE-Searchable-Data-File). Transcription
695 factors identified as significantly contributing to regulation of more than one of our GOIs were included
696 in Figure 8. We also used this same dataset to identify GOIs which are themselves regulators and
697 affect expression of other GOIs.
698
699 *Fixed SNPs in global datasets*
700 We searched for the presence of mutations identified in our study in publicly available sequencing data
701 using a searchable COBS index of bacterial genomes curated from the European Nucleotide Archive
702 (Blackwell et al., 2021). We used the Python interface to search the COBS index (Bingmann et al.,
703 2019). Our query sequences included 50bp in either direction of the SNP of interest, with a k-mer
704 matching threshold of 1. To determine the frequency of isolates in the COBS index with each SNP of
705 interest, we needed to determine the number of *M. tb* genomes in the database: we searched the
706 database using the full H37Rv 16S rRNA sequence (accession: NR_102810.2) which returned 39,941
707 results – this total number of *M. tb* isolates was used to calculate frequency of our SNPs of interest
708 (Table 1).
709
710

711 *RNA isolation*

712 Frozen cryovials of ancestral populations (MT540-0, MT49-0) and evolved populations (MT540-4,
713 MT49-12) were thawed, and 250 µl of each was inoculated into 25 ml 7H9OTG and incubated at 37^0C
714 with shaking and grown to an OD_{600} of ~1. 250 µl of culture was used to inoculate 25 ml Sauton's for
715 biofilm growth. The remaining culture was pelleted, resuspended in 3 ml RNAprotect Bacteria Reagent
716 (Qiagen), divided into 3 x 1 ml aliquots (3 technical replicates), and frozen at -80^0C until RNA
717 extraction. Biofilm cultures were grown as previously described, and pellicles were harvested at 5
718 weeks. Pellicles were broken up and suspended in 3 ml RNAprotect, divided into 3 x 1 ml aliquots (3
719 technical replicates), and frozen at -80^0C until RNA extraction. RNA was extracted from each 1 ml
720 aliquot (three technical replicates per sample), using the illustra RNAspin Mini RNA Isolation Kit (GE
721 Healthcare) following the "Protocol for total RNA purification from up to 10^9 bacterial cells," which
722 includes a 15-minute room-temperature incubation with RNase-free DNase I. RNA integrity was
723 assessed by bleach gel electrophoresis (O'Neill et al., 2018) and quantified using a NanoDrop ND-1000
724 spectrophotometer (NanoDrop Technologies Inc).

725

726 *RT-qPCR*

727 ~100 ng of each RNA sample was used for cDNA synthesis with random hexamer primers using the
728 RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) and included a 5-minute incubation at
729 65^0C and a reaction temperature of 44^0C for GC-rich templates. Real-time quantitative PCR was
730 performed in EU Fast Cycler PCR 96-well plates (BIOplastics) using the StepOnePlus Real-Time PCR
731 System (v. 2.3, Applied Biosystems). We used sets of primers previously published and verified (Akhtar
732 et al., 2009) to measure the expression of *lpdA* normalized to the expression of the endogenous
733 control, *sigA.* All real-time PCR assays were run in a total reaction volume of 20 µl comprised of 2X
734 Fast SYBR Green Master Mix (Applied Biosystems), 200 nM of both forward and reverse primers
735 (Integrated DNA Technologies, Coralville, IA, USA), and 3 µl of cDNA. RT-qPCR cycling parameters
736 were set as follows: an initial AmpliTaq Fast DNA Polymerase, UP Activation step of 20 seconds at
737 95^0C, followed by 40 cycles of 3 seconds at 95^0C and 30 seconds at 60^0C. Each reaction was repeated
738 three times with three independent cDNA samples. Negative controls consisting of no-template reaction
739 mixtures were run with all reactions. Melting-curve analysis was carried out to confirm the specificity of
740 the amplified product. After baseline correction and determination of the threshold settings, the Factor-
741 qPCR ratio method (Ruijter et al., 2015) was used to remove between-run variation due to the
742 experiment including multiple plates. The factor corrected ΔCT values of the three technical replicates
743 (MT540-EB has only 2 replicates) were then averaged and the $\Delta\Delta$CTs were calculated by subtracting a
744 reference value (MT540-AP) from each sample. The expression fold change was calculated using the

745 $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001), as PCR efficiencies had previously been found to be
746 similar. Low and high fold changes were calculated by adding or subtracting the standard deviation
747 from the fold change. The results are expressed as log2 of the fold changes. Kruskal-Wallis test was
748 used to determine the presence of significant log2 fold differences in the dataset (*P*-value = 1.89e-04).
749 Then, a Wilcox test was performed to compare log2 fold changes between sample types (AP, AB, EP,
750 EB) to identify significant differences between pairs. A *P*-value less than 0.05 was considered
751 significant.
752
753 *lpdA overexpression mutant*
754 The *lpdA* gene was overexpressed in *M. tb* H37Rv at the L5 site using the integrative, constitutive
755 overexpression plasmid pCT94 provided by Sarah Fortune's lab (Harvard T.H. Chan School of Public
756 Health). *lpdA* was amplified and modified to include *NdeI* and *HindIII* digestion sites using forward
757 primer CCGCATGCTTAATTAAGAAGGAGATATACA-CCGCCGTCTAAGCTTTCCCCC and reverse
758 primer GACCTCTAGGGTCCCCAATTAATTAGCTAA-TAGCCTGGGCGCTCGCGATAA. After
759 amplification with Platinum SuperFi Green PCR Master Mix (Invitrogen, Waltham, MA), the PCR
760 product was digested with *NdeI* and *HindIII* (NEB, Ipswich, MA). The digested product was then ligated
761 with a *NdeI* and *HindIII* cut pCT94 using Gibson assembly (NEB, Ipswich, MA). The assembly product
762 (CT94-*lpdA*) was then heat-shock transformed into DH5α *E. coli* cells and selected for on LB plates
763 with 50 ug/ml kanamycin overnight. Plasmid integration was verified via *NdeI* and *HindIII* digestion.
764 CT94-*lpdA* and an empty vector control (CT94) were electroporated into electrocompetent *M. tb* H37Rv
765 and transformants were recovered in 2 ml of fresh 7H9 for 24-48 h at 37°C, and then struck on 7H10
766 plates containing 25 ug/ml kanamycin. Both control and *lpdA* overexpression strains were grown as
767 pellicles in triplicate and wet weight measurements taken as described above.
768
769 **Data Availability**
770 Raw sequence data for ancestral and evolved isolates has been submitted to NCBI under the project
771 accession PRJNA720149.
772
773 **Author contributions**
774 TMS and CSP conceived of the study. TMS, MAY, JFK, SSF and TDM conducted experiments. TMS
775 and JFK performed SEM and image analysis. TMS, MAY, MAM, LLB, MBO and CSP performed data
776 analysis. TMS, MAY and CSP drafted the manuscript. All authors provided feedback and contributed to
777 manuscript revision.
778

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 thank the Fortune lab with their assistance in the design of our overexpression experiments. We would like to acknowledge Soleil Young for her help phenotyping pellicles. We also thank the University of Wisconsin Biotechnology Center and the Microbial Genome Sequencing 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

Akhtar, P., Srivastava, S., Srivastava, A., Srivastava, M., Srivastava, B.S., Srivastava, R., 2006. Rv3303c of Mycobacterium tuberculosis protects tubercle bacilli against oxidative stress in vivo and contributes to virulence in mice. Microbes Infect. 8, 2855–2862. https://doi.org/10.1016/j.micinf.2006.09.004

Ali, F., Seshasayee, A.S.N., 2020. Dynamics of genetic variation in transcription factors and its implications for the evolution of regulatory networks in Bacteria. Nucleic Acids Res. 48, 4100–4114. https://doi.org/10.1093/nar/gkaa162

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

817 Non-Coding RNA in the Total Transcriptome of Mycobacterium tuberculosis. PLoS Pathog. 7.
818 https://doi.org/10.1371/journal.ppat.1002342
819 Bacon, J., Alderwick, L.J., Allnutt, J.A., Gabasova, E., Watson, R., Hatch, K.A., Clark, S.O., Jeeves, R.E.,
820 Marriott, A., Rayner, E., Tolley, H., Pearson, G., Hall, G., Besra, G.S., Wernisch, L., Williams, A.,
821 Marsh, P.D., 2014. Non-Replicating Mycobacterium tuberculosis Elicits a Reduced Infectivity
822 Profile with Corresponding Modifications to the Cell Wall and Extracellular Matrix. PLOS ONE 9,
823 e87329. https://doi.org/10.1371/journal.pone.0087329
824 Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., Kim, J.F., 2009. Genome
825 evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461, 1243–
826 1247. https://doi.org/10.1038/nature08480
827 Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M., Kishony, R., 2015. Inexpensive
828 Multiplexed Library Preparation for Megabase-Sized Genomes (No. biorxiv;013771v2).
829 Bharti, S., Maurya, R.K., Venugopal, U., Singh, R., Akhtar, Md.S., Krishnan, M.Y., 2021. Rv1717 Is a Cell
830 Wall - Associated β-Galactosidase of Mycobacterium tuberculosis That Is Involved in Biofilm
831 Dispersion. Front. Microbiol. 11, 611122. https://doi.org/10.3389/fmicb.2020.611122
832 Bingmann, T., Bradley, P., Gauger, F., Iqbal, Z., 2019. COBS: A Compact Bit-Sliced Signature Index, in:
833 Brisaboa, N.R., Puglisi, S.J. (Eds.), String Processing and Information Retrieval, Lecture Notes in
834 Computer Science. Springer International Publishing, Cham, pp. 285–303.
835 https://doi.org/10.1007/978-3-030-32686-9_21
836 Blackwell, G.A., Hunt, M., Malone, K.M., Lima, L., Horesh, G., Alako, B.T.F., Thomson, N.R., Iqbal, Z.,
837 2021. Exploring bacterial diversity via a curated and searchable snapshot of archived DNA
838 sequences. bioRxiv 2021.03.02.433662. https://doi.org/10.1101/2021.03.02.433662
839 Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J.,
840 Alfsnes, K., Pettersson, J.O.-H., Kirkeleite, I., Fandinho, F., Silva, M.A. da, Perdigao, J., Portugal,
841 I., Viveiros, M., Clark, T., Caws, M., Dunstan, S., Thai, P.V.K., Lopez, B., Ritacco, V., Kitchen, A.,
842 Brown, T.S., Soolingen, D. van, O'Neill, M.B., Holt, K.E., Feil, E.J., Mathema, B., Balloux, F.,
843 Eldholm, V., 2018. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by
844 colonial migration and local adaptation. Sci. Adv. 4, eaat5869.
845 https://doi.org/10.1126/sciadv.aat5869
846 Canetti, G., 1955. The Tubercle Bacillus in the Pulmonary Lesion of Man. Springer, New York.
847 Carey, A.F., Rock, J.M., Krieger, I.V., Chase, M.R., Fernandez-Suarez, M., Gagneux, S., Sacchettini, J.C.,
848 Ioerger, T.R., Fortune, S.M., 2018. TnSeq of Mycobacterium tuberculosis clinical isolates reveals
849 strain-specific antibiotic liabilities. PLOS Pathog. 14, e1006939.
850 https://doi.org/10.1371/journal.ppat.1006939
851 Castro, R.A.D., Ross, A., Kamwela, L., Reinhard, M., Loiseau, C., Feldmann, J., Borrell, S., Trauner, A.,
852 Gagneux, S., 2019. The Genetic Background modulates the Evolution of Fluoroquinolone-
853 Resistance in Mycobacterium tuberculosis. Mol. Biol. Evol.
854 https://doi.org/10.1093/molbev/msz214
855 Chakraborty, P., Bajeli, S., Kaushal, D., Radotra, B.D., Kumar, A., 2021. Biofilm formation in the lung
856 contributes to virulence and drug tolerance of Mycobacterium tuberculosis. Nat. Commun. 12,
857 1606. https://doi.org/10.1038/s41467-021-21748-6
858 Chauhan, R., Ravi, J., Datta, P., Chen, T., Schnappinger, D., Bassler, K.E., Balázsi, G., Gennaro, M.L.,
859 2016. Reconstruction and topological characterization of the sigma factor regulatory network of
860 Mycobacterium tuberculosis. Nat. Commun. 7, 11062. https://doi.org/10.1038/ncomms11062

861 Chen, J., Xie, J., 2011. Role and regulation of bacterial LuxR-like regulators. J. Cell. Biochem. 112, 2694–
862 2702. https://doi.org/10.1002/jcb.23219
863 Chiner-Oms, Á., Berney, M., Boinett, C., González-Candelas, F., Young, D.B., Gagneux, S., Jacobs, W.R.,
864 Parkhill, J., Cortes, T., Comas, I., 2019. Genome-wide mutational biases fuel transcriptional
865 diversity in the Mycobacterium tuberculosis complex. Nat. Commun. 10, 3994.
866 https://doi.org/10.1038/s41467-019-11948-6
867 Conrad, T.M., Frazier, M., Joyce, A.R., Cho, B.-K., Knight, E.M., Lewis, N.E., Landick, R., Palsson, B.Ø.,
868 2010. RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli
869 for optimal growth in minimal media. Proc. Natl. Acad. Sci. 107, 20500–20505.
870 https://doi.org/10.1073/pnas.0911253107
871 Costerton, J.W., Stewart, P.S., Greenberg, E.P., 1999. Bacterial Biofilms: A Common Cause of Persistent
872 Infections. Science 284, 1318–1322. https://doi.org/10.1126/science.284.5418.1318
873 Deatherage, D.E., Barrick, J.E., 2014. Identification of mutations in laboratory evolved microbes from
874 next-generation sequencing data using breseq. Methods Mol. Biol. Clifton NJ 1151, 165–188.
875 https://doi.org/10.1007/978-1-4939-0554-6_12
876 Deb, D.K., Dahiya, P., Srivastava, K.K., Srivastava, R., Srivastava, B.S., 2002. Selective identification of
877 new therapeutic targets of Mycobacterium tuberculosis by IVIAT approach. Tuberc. Edinb.
878 Scotl. 82, 175–182. https://doi.org/10.1054/tube.2002.0337
879 DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D.,
880 Ehrt, S., Fortune, S.M., Sassetti, C.M., Ioerger, T.R., 2017. Comprehensive Essentiality Analysis of
881 the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. mBio 8.
882 https://doi.org/10.1128/mBio.02133-16
883 DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel,
884 G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y.,
885 Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery
886 and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498.
887 https://doi.org/10.1038/ng.806
888 Domenech, P., Kolly, G.S., Leon-Solis, L., Fallow, A., Reed, M.B., 2010. Massive Gene Duplication Event
889 among Clinical Isolates of the Mycobacterium tuberculosis W/Beijing Family. J. Bacteriol. 192,
890 4562–4570. https://doi.org/10.1128/JB.00536-10
891 Domenech, P., Rog, A., Moolji, J., Radomski, N., Fallow, A., Leon-Solis, L., Bowes, J., Behr, M.A., Reed,
892 M.B., 2014a. Origins of a 350-Kilobase Genomic Duplication in Mycobacterium tuberculosis and
893 Its Impact on Virulence. Infect. Immun. 82, 2902–2912. https://doi.org/10.1128/IAI.01791-14
894 Domenech, P., Rog, A., Moolji, J., Radomski, N., Fallow, A., Leon-Solis, L., Bowes, J., Behr, M.A., Reed,
895 M.B., 2014b. Origins of a 350-Kilobase Genomic Duplication in Mycobacterium tuberculosis and
896 Its Impact on Virulence. Infect. Immun. 82, 2902–2912. https://doi.org/10.1128/IAI.01791-14
897 Domenech, P., Zou, J., Averback, A., Syed, N., Curtis, D., Donato, S., Reed, M.B., 2017. Unique
898 Regulation of the DosR Regulon in the Beijing Lineage of Mycobacterium tuberculosis. J.
899 Bacteriol. 199. https://doi.org/10.1128/JB.00696-16
900 Dubos, R.J., Davis, B.D., 1946. FACTORS AFFECTING THE GROWTH OF TUBERCLE BACILLI IN LIQUID
901 MEDIA. J. Exp. Med. 83, 409–423.
902 Ektefaie, Y., Dixit, A., Freschi, L., Farhat, M.R., 2021. Globally diverse Mycobacterium tuberculosis
903 resistance acquisition: a retrospective geographical and temporal analysis of whole genome

904 sequences. Lancet Microbe S2666524720301956. https://doi.org/10.1016/S2666-
905 5247(20)30195-6
906 Eldholm, V., Balloux, F., 2016. Antimicrobial Resistance in Mycobacterium tuberculosis: The Odd One
907 Out. Trends Microbiol. 24, 637–648. https://doi.org/10.1016/j.tim.2016.03.007
908 Farhat, M.R., Freschi, L., Calderon, R., Ioerger, T., Snyder, M., Meehan, C.J., Jong, B. de, Rigouts, L.,
909 Sloutsky, A., Kaur, D., Sunyaev, S., Soolingen, D. van, Shendure, J., Sacchettini, J., Murray, M.,
910 2019. GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals
911 resistance genes and regulatory regions. Nat. Commun. 10, 2128.
912 https://doi.org/10.1038/s41467-019-10110-6
913 Fisher, K.J., Kryazhimskiy, S., Lang, G.I., 2019. Detecting genetic interactions using parallel evolution in
914 experimental populations. Philos. Trans. R. Soc. B Biol. Sci. 374, 20180237.
915 https://doi.org/10.1098/rstb.2018.0237
916 Galagan, J.E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T.,
917 Dolganov, G., Glotova, I., Abeel, T., Mahwinney, C., Kennedy, A.D., Allard, R., Brabant, W.,
918 Krueger, A., Jaini, S., Honda, B., Yu, W.-H., Hickey, M.J., Zucker, J., Garay, C., Weiner, B., Sisk, P.,
919 Stolte, C., Winkler, J.K., Van de Peer, Y., Iazzetti, P., Camacho, D., Dreyfuss, J., Liu, Y., Dorhoi, A.,
920 Mollenkopf, H.-J., Drogaris, P., Lamontagne, J., Zhou, Y., Piquenot, J., Park, S.T., Raman, S.,
921 Kaufmann, S.H.E., Mohney, R.P., Chelsky, D., Moody, D.B., Sherman, D.R., Schoolnik, G.K., 2013.
922 The Mycobacterium tuberculosis regulatory network and hypoxia. Nature 499, 178–183.
923 https://doi.org/10.1038/nature12337
924 Gerrick, E.R., Barbier, T., Chase, M.R., Xu, R., François, J., Lin, V.H., Szucs, M.J., Rock, J.M., Ahmad, R.,
925 Tjaden, B., Livny, J., Fortune, S.M., 2018. Small RNA profiling in Mycobacterium tuberculosis
926 identifies MrsI as necessary for an anticipatory iron sparing response. Proc. Natl. Acad. Sci. 115,
927 6464–6469. https://doi.org/10.1073/pnas.1718003115
928 Global Tuberculosis Report 2019 [WWW Document], n.d. URL https://www.who.int/teams/global-
929 tuberculosis-programme/tb-reports/global-report-2019 (accessed 3.9.21).
930 Gomez-Gonzalez, P.J., Andreu, N., Phelan, J.E., de Sessions, P.F., Glynn, J.R., Crampin, A.C., Campino, S.,
931 Butcher, P.D., Hibberd, M.L., Clark, T.G., 2019. An integrated whole genome analysis of
932 Mycobacterium tuberculosis reveals insights into relationship between its genome,
933 transcriptome and methylome. Sci. Rep. 9, 5204. https://doi.org/10.1038/s41598-019-41692-2
934 Hegde, S.R., 2020. Computational Identification of the Proteins Associated With Quorum Sensing and
935 Biofilm Formation in Mycobacterium tuberculosis. Front. Microbiol. 10.
936 https://doi.org/10.3389/fmicb.2019.03011
937 Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., Small, P.M., 2004. Stable association between
938 strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci U
939 A 101, 4871–6. https://doi.org/10.1073/pnas.0305627101
940 Hu, Y., Coates, A.R.M., 2011. Mycobacterium tuberculosis acg Gene Is Required for Growth and
941 Virulence In Vivo. PLOS ONE 6, e20958. https://doi.org/10.1371/journal.pone.0020958
942 Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees
943 and Networks. Syst. Biol. sys062. https://doi.org/10.1093/sysbio/sys062
944 Kerns, P.W., Ackhart, D.F., Basaraba, R.J., Leid, J.G., Shirtliff, M.E., 2014. Mycobacterium tuberculosis
945 pellicles express unique proteins recognized by the host humoral response. Pathog. Dis. 70,
946 347–358. https://doi.org/10.1111/2049-632X.12142

947 Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., Cooper, T.F., 2011. Negative Epistasis Between
948 Beneficial Mutations in an Evolving Bacterial Population. Science 332, 1193–1196.
949 https://doi.org/10.1126/science.1203801
950 Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826–837. https://doi.org/10.1038/nrg1471
951 Kobayashi, K., 2007. Bacillus subtilis Pellicle Formation Proceeds through Genetically Defined
952 Morphological Changes. J. Bacteriol. 189, 4920–4931. https://doi.org/10.1128/JB.00157-07
953 Kofler, R., Orozco-terWengel, P., Maio, N.D., Pandey, R.V., Nolte, V., Futschik, A., Kosiol, C., Schlötterer,
954 C., 2011a. PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation
955 Sequencing Data from Pooled Individuals. PLOS ONE 6, e15925.
956 https://doi.org/10.1371/journal.pone.0015925
957 Kofler, R., Pandey, R.V., Schlötterer, C., 2011b. PoPoolation2: identifying differentiation between
958 populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27, 3435–
959 3436. https://doi.org/10.1093/bioinformatics/btr589
960 Kulka, K., Hatfull, G., Ojha, A.K., 2012. Growth of Mycobacterium tuberculosis Biofilms. J. Vis. Exp. JoVE.
961 https://doi.org/10.3791/3820
962 Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
963 ArXiv13033997 Q-Bio.
964 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.,
965 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format
966 and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079.
967 https://doi.org/10.1093/bioinformatics/btp352
968 Li, X., Lv, X., Lin, Y., Zhen, J., Ruan, C., Duan, W., Li, Y., Xie, J., 2019. Role of two-component regulatory
969 systems in intracellular survival of Mycobacterium tuberculosis. J. Cell. Biochem. 120, 12197–
970 12207. https://doi.org/10.1002/jcb.28792
971 Lipworth, S., Jajou, R., de Neeling, A., Bradley, P., van der Hoek, W., Maphalala, G., Bonnet, M.,
972 Sanchez-Padilla, E., Diel, R., Niemann, S., Iqbal, Z., Smith, G., Peto, T., Crook, D., Walker, T., van
973 Soolingen, D., 2019. SNP-IT Tool for Identifying Subspecies and Associated Lineages of
974 *Mycobacterium tuberculosis* Complex. Emerg. Infect. Dis. 25, 482–488.
975 https://doi.org/10.3201/eid2503.180894
976 Liu, Q., Ma, A., Wei, L., Pang, Y., Wu, B., Luo, T., Zhou, Y., Zheng, H.-X., Jiang, Q., Gan, M., Zuo, T., Liu,
977 M., Yang, C., Jin, L., Comas, I., Gagneux, S., Zhao, Y., Pepperell, C.S., Gao, Q., 2018. China's
978 tuberculosis epidemic stems from historical expansion of four strains of Mycobacterium
979 tuberculosis. Nat. Ecol. Evol. 2, 1982. https://doi.org/10.1038/s41559-018-0680-6
980 Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative
981 PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif 25, 402–408.
982 https://doi.org/10.1006/meth.2001.1262
983 Lynch, M., Conery, J.S., 2000. The Evolutionary Fate and Consequences of Duplicate Genes. Science
984 290, 1151–1155. https://doi.org/10.1126/science.290.5494.1151
985 Maksymiuk, C., Ioerger, T., Balakrishnan, A., Bryk, R., Rhee, K., Sacchettini, J., Nathan, C., 2015.
986 Comparison of transposon and deletion mutants in Mycobacterium tuberculosis: The case of
987 rv1248c, encoding 2-hydroxy-3-oxoadipate synthase. Tuberculosis 95, 689–694.
988 https://doi.org/10.1016/j.tube.2015.08.009
989 Manson, A.L., Cohen, K.A., Abeel, T., Desjardins, C.A., Armstrong, D.T., Barry Iii, C.E., Brand, J., TBResist
990 Global Genome Consortium, Chapman, S.B., Cho, S.-N., Gabrielian, A., Gomez, J., Jodals, A.M.,

991 Joloba, M., Jureen, P., Lee, J.S., Malinga, L., Maiga, M., Nordenberg, D., Noroc, E., Romancenco,
992 E., Salazar, A., Ssengooba, W., Velayati, A.A., Winglee, K., Zalutskaya, A., Via, L.E., Cassell, G.H.,
993 Dorman, S.E., Ellner, J., Farnia, P., Galagan, J.E., Rosenthal, A., Crudu, V., Homorodean, D.,
994 Hsueh, P.-R., Narayanan, S., Pym, A.S., Skrahina, A., Swaminathan, S., Van der Walt, M., Alland,
995 D., Bishai, W.R., Cohen, T., Hoffner, S., Birren, B.W., Earl, A.M., 2017. Genomic analysis of
996 globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and
997 spread of multidrug resistance. Nat. Genet. advance online publication.
998 https://doi.org/10.1038/ng.3767
999 Modlin, S.J., Conkle-Gutierrez, D., Kim, C., Mitchell, S.N., Morrissey, C., Weinrick, B.C., Jacobs, W.R.,
1000 Ramirez-Busby, S.M., Hoffner, S.E., Valafar, F., 2020. Drivers and sites of diversity in the DNA
1001 adenine methylomes of 93 Mycobacterium tuberculosis complex clinical isolates. eLife 9,
1002 e58542. https://doi.org/10.7554/eLife.58542
1003 Mortimer, T.D., Weber, A.M., Pepperell, C.S., 2018. Signatures of Selection at Drug Resistance Loci in
1004 Mycobacterium tuberculosis. mSystems 3, e00108-17.
1005 https://doi.org/10.1128/mSystems.00108-17
1006 Mulholland, C.V., Shockey, A.C., Aung, H.L., Cursons, R.T., O'Toole, R.F., Gautam, S.S., Brites, D.,
1007 Gagneux, S., Roberts, S.A., Karalus, N., Cook, G.M., Pepperell, C.S., Arcus, V.L., 2019. Dispersal of
1008 Mycobacterium tuberculosis Driven by Historical European Trade in the South Pacific. Front.
1009 Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02778
1010 Nyka, W., 1977. The chromophobic tubercle bacilli and the problem of endogenous reactivation of
1011 tuberculosis. Mater. Medica Pol. Pol. J. Med. Pharm. 9, 175–185.
1012 Nyka, W., 1967. Method for staining both acid-fast and chromophobic tubercle bacilli with
1013 carbolfuschsin. J. Bacteriol. 93, 1458–1460.
1014 Nyka, W., 1963. STUDIES ON MYCOBACTERIUM TUBERCULOSIS IN LESIONS OF THE HUMAN LUNG. A
1015 NEW METHOD OF STAINING TUBERCLE BACILLI IN TISSUE SECTIONS. Am. Rev. Respir. Dis. 88,
1016 670–679. https://doi.org/10.1164/arrd.1963.88.5.670
1017 Nyka, W., O'Neill, E.F., 1970. A NEW APPROACH TO THE STUDY OF NON-ACID-FAST MYCOBACTERIA.
1018 Ann. N. Y. Acad. Sci. 174, 862–871. https://doi.org/10.1111/j.1749-6632.1970.tb45605.x
1019 Ojha, A.K., Baughn, A.D., Sambandan, D., Hsu, T., Trivelli, X., Guerardel, Y., Alahari, A., Kremer, L.,
1020 Jacobs, W.R., Hatfull, G.F., 2008. Growth of Mycobacterium tuberculosis biofilms containing
1021 free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. 69, 164–174.
1022 https://doi.org/10.1111/j.1365-2958.2008.06274.x
1023 Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality
1024 control for high-throughput sequencing data. Bioinformatics 32, 292–294.
1025 https://doi.org/10.1093/bioinformatics/btv566
1026 O'Neill, M.B., Mortimer, T.D., Pepperell, C.S., 2015. Diversity of Mycobacterium tuberculosis across
1027 Evolutionary Scales. PLOS Pathog. 11, e1005257. https://doi.org/10.1371/journal.ppat.1005257
1028 O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019.
1029 Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Mol.
1030 Ecol. mec.15120. https://doi.org/10.1111/mec.15120
1031 Øyås, O., Borrell, S., Trauner, A., Zimmermann, M., Feldmann, J., Liphardt, T., Gagneux, S., Stelling, J.,
1032 Sauer, U., Zampieri, M., 2020. Model-based integration of genomics and metabolomics reveals
1033 SNP functionality in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. 117, 8494–8502.
1034 https://doi.org/10.1073/pnas.1915551117

1035 Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites:
1036 rapid efficient extraction of SNPs from multi-FASTA alignments. Microb. Genomics 2.
1037 https://doi.org/10.1099/mgen.0.000056
1038 Pang, J.M., Layre, E., Sweet, L., Sherrid, A., Moody, D.B., Ojha, A., Sherman, D.R., 2012. The Polyketide
1039 Pks1 Contributes to Biofilm Formation in Mycobacterium tuberculosis. J. Bacteriol. 194, 715–
1040 721. https://doi.org/10.1128/JB.06304-11
1041 Parish, T., Smith, D.A., Kendall, S., Casali, N., Bancroft, G.J., Stoker, N.G., 2003. Deletion of Two-
1042 Component Regulatory Systems Increases the Virulence of Mycobacterium tuberculosis. Infect.
1043 Immun. https://doi.org/10.1128/IAI.71.3.1134-1140.2003
1044 Parvati Sai Arun, P.V., Miryala, S.K., Rana, A., Kurukuti, S., Akhter, Y., Yellaboina, S., 2018. System-wide
1045 coordinates of higher order functions in host-pathogen environment upon Mycobacterium
1046 tuberculosis infection. Sci. Rep. 8, 5079. https://doi.org/10.1038/s41598-018-22884-8
1047 Pepperell, C., Hoeppner, V.H., Lipatov, M., Wobeser, W., Schoolnik, G.K., Feldman, M.W., 2010.
1048 Bacterial genetic signatures of human social phenomena among M. tuberculosis from an
1049 Aboriginal Canadian population. Mol Biol Evol 27, 427–40.
1050 https://doi.org/10.1093/molbev/msp261
1051 Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Cornejo, O.E., Holmes, E.C., Birren, B., Galagan,
1052 J., Feldman, M.W., 2013. The Role of Selection in Shaping Diversity of Natural M. tuberculosis
1053 Populations. PLOS Pathog. 9, e1003543. https://doi.org/10.1371/journal.ppat.1003543
1054 Pepperell, C.S., Granka, J.M., Alexander, D.C., Behr, M.A., Chui, L., Gordon, J., Guthrie, J.L., Jamieson,
1055 F.B., Langlois-Klassen, D., Long, R., Nguyen, D., Wobeser, W., Feldman, M.W., 2011. Dispersal of
1056 Mycobacterium tuberculosis via the Canadian fur trade. Proc Natl Acad Sci U A 108, 6526–31.
1057 https://doi.org/10.1073/pnas.1016708108
1058 Pérez-Lago, L., Navarro, Y., Herranz, M., Bouza, E., García-de-Viedma, D., 2013. Differences in gene
1059 expression between clonal variants of Mycobacterium tuberculosis emerging as a result of
1060 microevolution. Int. J. Med. Microbiol. 303, 674–677.
1061 https://doi.org/10.1016/j.ijmm.2013.09.010
1062 Phelan, J., Sessions, P.F., Tientcheu, L., Perdigao, J., Machado, D., Hasan, R., Hasan, Z., Bergval, I.L.,
1063 Anthony, R., McNerney, R., Antonio, M., Portugal, I., Viveiros, M., Campino, S., Hibberd, M.L.,
1064 Clark, T.G., 2018. Methylation in Mycobacterium tuberculosis is lineage specific with associated
1065 mutations present globally. Sci. Rep. 8, 160. https://doi.org/10.1038/s41598-017-18188-y
1066 Philippe, N., Crozat, E., Lenski, R.E., Schneider, D., 2007. Evolution of global regulatory networks during
1067 a long-term experiment with Escherichia coli. BioEssays 29, 846–860.
1068 https://doi.org/10.1002/bies.20629
1069 Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J., 2007. Empirical fitness landscapes reveal
1070 accessible evolutionary paths. Nature 445, 383–386. https://doi.org/10.1038/nature05451
1071 Rastogi, S., Singh, A.K., Pant, G., Mitra, K., Sashidhara, K.V., Krishnan, M.Y., 2017. Down-regulation of
1072 PE11, a cell wall associated esterase, enhances the biofilm growth of Mycobacterium
1073 tuberculosis and reduces cell wall virulence lipid levels. Microbiology 163, 52–61.
1074 https://doi.org/10.1099/mic.0.000417
1075 Richards, J.P., Cai, W., Zill, N.A., Zhang, W., Ojha, A.K., 2019. Adaptation of Mycobacterium tuberculosis
1076 to Biofilm Growth Is Genetically Linked to Drug Tolerance. Antimicrob. Agents Chemother. 63.
1077 https://doi.org/10.1128/AAC.01213-19

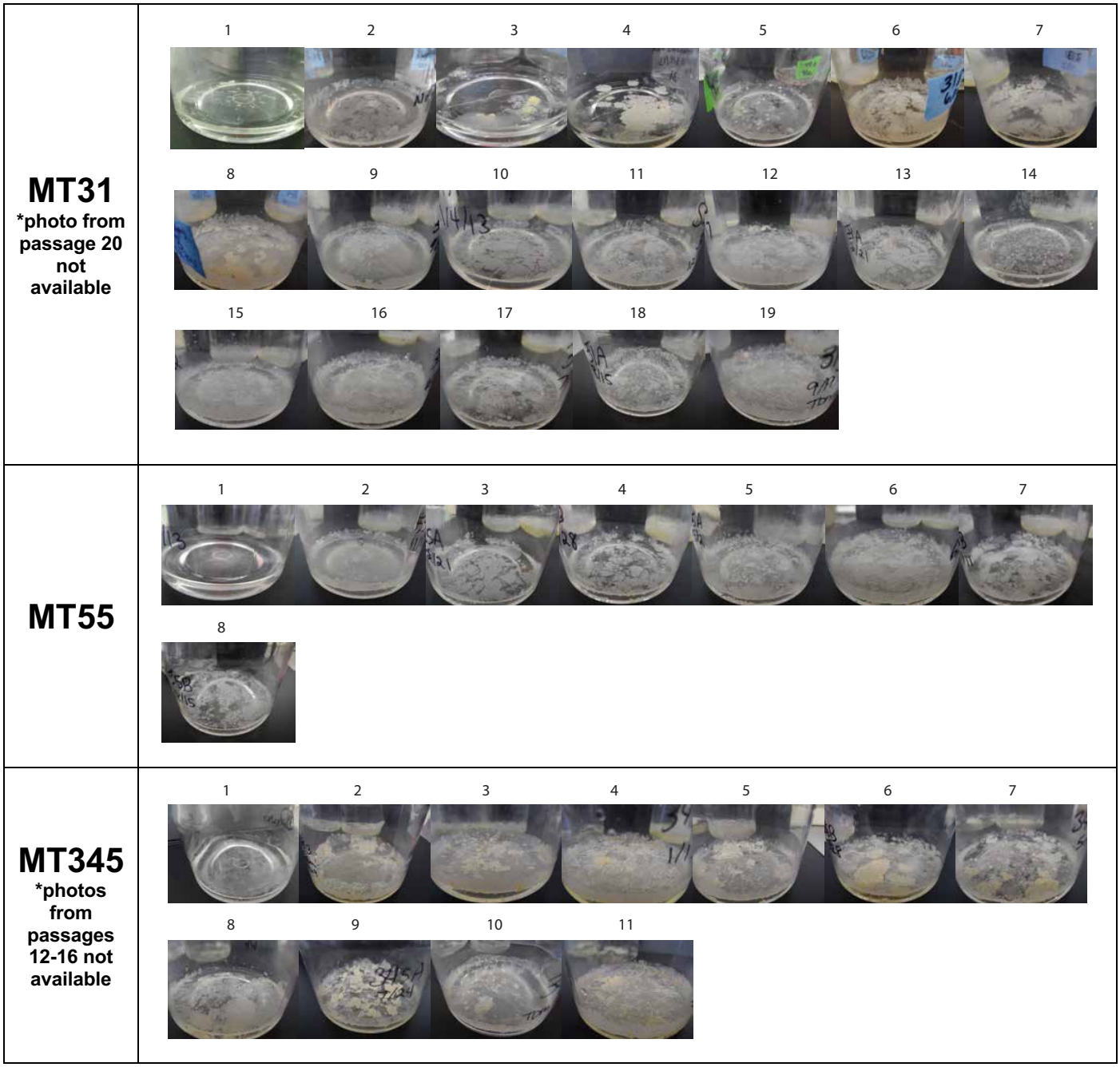
1078 Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P.,
1079 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.
1080 https://doi.org/10.1038/nbt.1754
1081 Rodríguez-Verdugo, A., Tenaillon, O., Gaut, B.S., 2016. First-Step Mutations during Adaptation Restore
1082 the Expression of Hundreds of Genes. Mol. Biol. Evol. 33, 25–39.
1083 https://doi.org/10.1093/molbev/msv228
1084 Ruijter, J.M., Ruiz Villalba, A., Hellemans, J., Untergasser, A., van den Hoff, M.J.B., 2015. Removal of
1085 between-run variation in a multi-plate qPCR experiment. Biomol. Detect. Quantif., Special Issue:
1086 Advanced Molecular Diagnostics for Biomarker Discovery – Part I 5, 10–14.
1087 https://doi.org/10.1016/j.bdq.2015.07.001
1088 Rustad, T.R., Minch, K.J., Ma, S., Winkler, J.K., Hobbs, S., Hickey, M., Brabant, W., Turkarslan, S., Price,
1089 N.D., Baliga, N.S., Sherman, D.R., 2014. Mapping and manipulating the Mycobacterium
1090 tuberculosis transcriptome using a transcription factor overexpression-derived regulatory
1091 network. Genome Biol. 15, 502. https://doi.org/10.1186/PREACCEPT-1701638048134699
1092 Sambandan, D., Dao, D.N., Weinrick, B.C., Vilchèze, C., Gurcha, S.S., Ojha, A., Kremer, L., Besra, G.S.,
1093 Hatfull, G.F., Jacobs, W.R., 2013. Keto-Mycolic Acid-Dependent Pellicle Formation Confers
1094 Tolerance to Drug-Sensitive Mycobacterium tuberculosis. mBio 4.
1095 https://doi.org/10.1128/mBio.00222-13
1096 Saxer, G., Krepps, M.D., Merkley, E.D., Ansong, C., Kaiser, B.L.D., Valovska, M.-T., Ristic, N., Yeh, P.T.,
1097 Prakash, V.P., Leiser, O.P., Nakhleh, L., Gibbons, H.S., Kreuzer, H.W., Shamoo, Y., 2014.
1098 Mutations in Global Regulators Lead to Metabolic Selection during Adaptation to Complex
1099 Environments. PLOS Genet. 10, e1004872. https://doi.org/10.1371/journal.pgen.1004872
1100 Sharma, I.M., Petchiappan, A., Chatterji, D., 2014. Quorum sensing and biofilm formation in
1101 mycobacteria: Role of c-di-GMP and methods to study this second messenger: QS, biofilm
1102 formation and c-di-GMP in mycobacteria. IUBMB Life 66, 823–834.
1103 https://doi.org/10.1002/iub.1339
1104 Shell, S.S., Prestwich, E.G., Baek, S.-H., Shah, R.R., Sassetti, C.M., Dedon, P.C., Fortune, S.M., 2013. DNA
1105 Methylation Impacts Gene Expression and Ensures Hypoxic Survival of Mycobacterium
1106 tuberculosis. PLOS Pathog. 9, e1003419. https://doi.org/10.1371/journal.ppat.1003419
1107 Silva, R.F., Mendonça, S.C.M., Carvalho, L.M., Reis, A.M., Gordo, I., Trindade, S., Dionisio, F., 2011.
1108 Pervasive Sign Epistasis between Conjugative Plasmids and Drug-Resistance Chromosomal
1109 Mutations. PLOS Genet. 7, e1002181. https://doi.org/10.1371/journal.pgen.1002181
1110 Singh, S., Sharma, M., Chaudhry, A., Sharma, S., 2019. Rv2626c and Rv2032 activate TH1 response and
1111 downregulate regulatory T cells in peripheral blood mononuclear cells of tuberculosis patients.
1112 Comp. Immunol. Microbiol. Infect. Dis. 62, 46–53. https://doi.org/10.1016/j.cimid.2018.11.016
1113 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1114 phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033
1115 Thorpe, H.A., Bayliss, S.C., Hurst, L.D., Feil, E.J., 2017. Comparative Analyses of Selection Operating on
1116 Nontranslated Intergenic Regions of Diverse Bacterial Species. Genetics 206, 363–376.
1117 https://doi.org/10.1534/genetics.116.195784
1118 Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-
1119 performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192.
1120 https://doi.org/10.1093/bib/bbs017

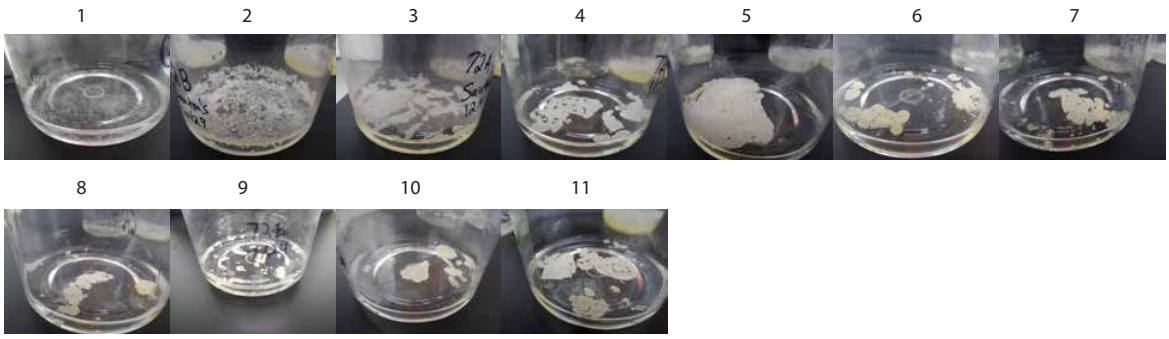
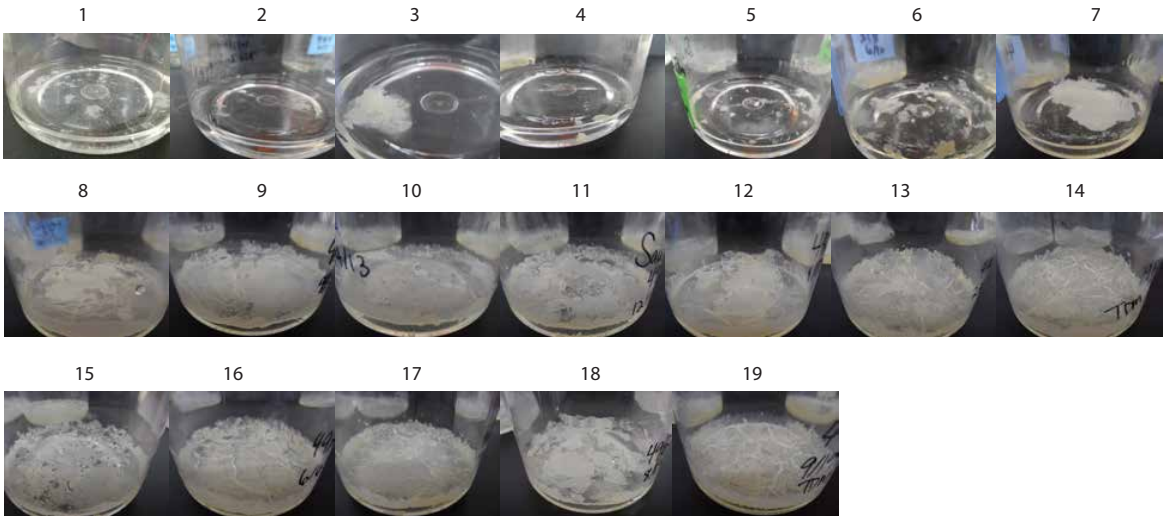
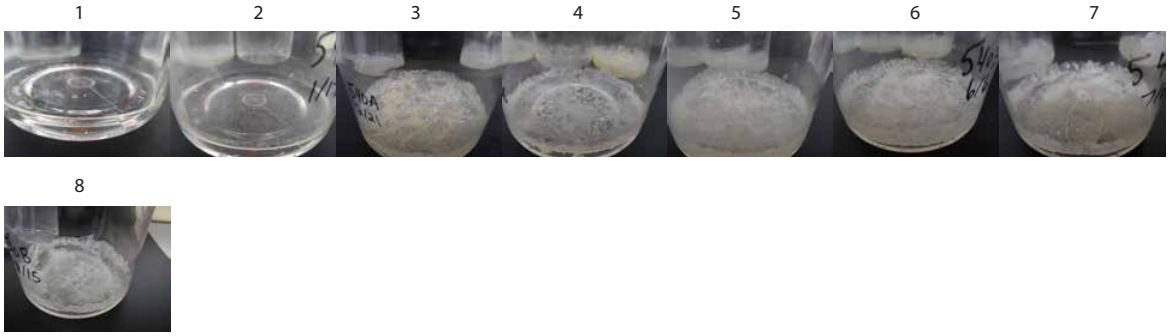
1121 Tizzano, B., Dallenga, T.K., Utpatel, C., Behrends, J., Homolka, S., Kohl, T.A., Niemann, S., 2021. Survival
1122 of hypoxia-induced dormancy is not a common feature of all strains of the Mycobacterium
1123 tuberculosis complex. Sci. Rep. 11, 2628. https://doi.org/10.1038/s41598-021-81223-6
1124 Trauner, A., Liu, Q., Via, L.E., Liu, X., Ruan, X., Liang, L., Shi, H., Chen, Y., Wang, Z., Liang, R., Zhang, W.,
1125 Wei, W., Gao, J., Sun, G., Brites, D., England, K., Zhang, G., Gagneux, S., Barry, C.E., Gao, Q.,
1126 2017. The within-host population dynamics of Mycobacterium tuberculosis vary with treatment
1127 efficacy. Genome Biol. 18, 71. https://doi.org/10.1186/s13059-017-1196-0
1128 Trivedi, A., Mavi, P.S., Bhatt, D., Kumar, A., 2016. Thiol reductive stress induces cellulose-anchored
1129 biofilm formation in Mycobacterium tuberculosis. Nat. Commun. 7, 11392.
1130 https://doi.org/10.1038/ncomms11392
1131 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.,
1132 Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: An Integrated Tool for Comprehensive
1133 Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE 9, e112963.
1134 https://doi.org/10.1371/journal.pone.0112963
1135 Wang, F., Sambandan, D., Halder, R., Wang, J., Batt, S.M., Weinrick, B., Ahmad, I., Yang, P., Zhang, Y.,
1136 Kim, J., Hassani, M., Huszar, S., Trefzer, C., Ma, Z., Kaneko, T., Mdluli, K.E., Franzblau, S.,
1137 Chatterjee, A.K., Johnsson, K., Mikusova, K., Besra, G.S., Fütterer, K., Robbins, S.H., Barnes, S.W.,
1138 Walker, J.R., Jacobs, W.R., Schultz, P.G., 2013. Identification of a small molecule with activity
1139 against drug-resistant and persistent tuberculosis. Proc. Natl. Acad. Sci. 110, E2510–E2517.
1140 https://doi.org/10.1073/pnas.1309171110
1141 Weiner, B., Gomez, J., Victor, T.C., Warren, R.M., Sloutsky, A., Plikaytis, B.B., Posey, J.E., Helden, P.D.
1142 van, Pittius, N.C.G. van, Koehrsen, M., Sisk, P., Stolte, C., White, J., Gagneux, S., Birren, B., Hung,
1143 D., Murray, M., Galagan, J., 2012. Independent Large Scale Duplications in Multiple M.
1144 tuberculosis Lineages Overlapping the Same Genomic Region. PLOS ONE 7, e26038.
1145 https://doi.org/10.1371/journal.pone.0026038
1146 Weinreich, D.M., Watson, R.A., Chao, L., 2005. Perspective: Sign Epistasis and Genetic Constraint on
1147 Evolutionary Trajectories. Evolution 59, 1165–1174.
1148 Wolff, K.A., Peña, A.H. de la, Nguyen, H.T., Pham, T.H., Amzel, L.M., Gabelli, S.B., Nguyen, L., 2015. A
1149 Redox Regulatory System Critical for Mycobacterial Survival in Macrophages and Biofilm
1150 Development. PLOS Pathog. 11, e1004839. https://doi.org/10.1371/journal.ppat.1004839
1151 Yang, D.C., Blair, K.M., Salama, N.R., 2016. Staying in Shape: the Impact of Cell Shape on Bacterial
1152 Survival in Diverse Environments. Microbiol. Mol. Biol. Rev. 80, 187–203.
1153 https://doi.org/10.1128/MMBR.00031-15
1154 Yang, Y., Thomas, J., Li, Y., Vilchèze, C., Derbyshire, K.M., Jacobs, W.R., Ojha, A.K., 2017. Defining a
1155 temporal order of genetic requirements for development of mycobacterial biofilms. Mol.
1156 Microbiol. https://doi.org/10.1111/mmi.13734
1157 Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y., 2017. ggtree: an r package for visualization and
1158 annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol.
1159 Evol. 8, 28–36. https://doi.org/10.1111/2041-210X.12628
1160 Zhang, J., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298.
1161 https://doi.org/10.1016/S0169-5347(03)00033-8
1162 Zheng, H., Lu, L., Wang, B., Pu, S., Zhang, X., Zhu, G., Shi, W., Zhang, L., Wang, H., Wang, S., Zhao, G.,
1163 Zhang, Y., 2008. Genetic Basis of Virulence Attenuation Revealed by Comparative Genomic

1164 Analysis of Mycobacterium tuberculosis Strain H37Ra versus H37Rv. PLOS ONE 3, e2375.
1165 https://doi.org/10.1371/journal.pone.0002375
1166 Zhu, L., Zhong, J., Jia, X., Liu, G., Kang, Y., Dong, M., Zhang, X., Li, Q., Yue, L., Li, C., Fu, J., Xiao, J., Yan, J.,
1167 Zhang, B., Lei, M., Chen, S., Lv, L., Zhu, B., Huang, H., Chen, F., 2015. Precision methylome
1168 characterization of Mycobacterium tuberculosis complex (MTBC) using PacBio single-molecule
1169 real-time (SMRT) technology. Nucleic Acids Res. gkv1498. https://doi.org/10.1093/nar/gkv1498
1170

Supplement

Supplementary Figure 1: All photos from pellicle passaging experiment. Photos taken at time of passaging, between 5 and 7 weeks of growth.



MT72 ***photos from passages 12-16 not available**	
MT49 ***photo from passage 20 not available**	
MT540	

1190 **Supplementary Table 1:** Outline of data collection for each strain at each passage point.
1191

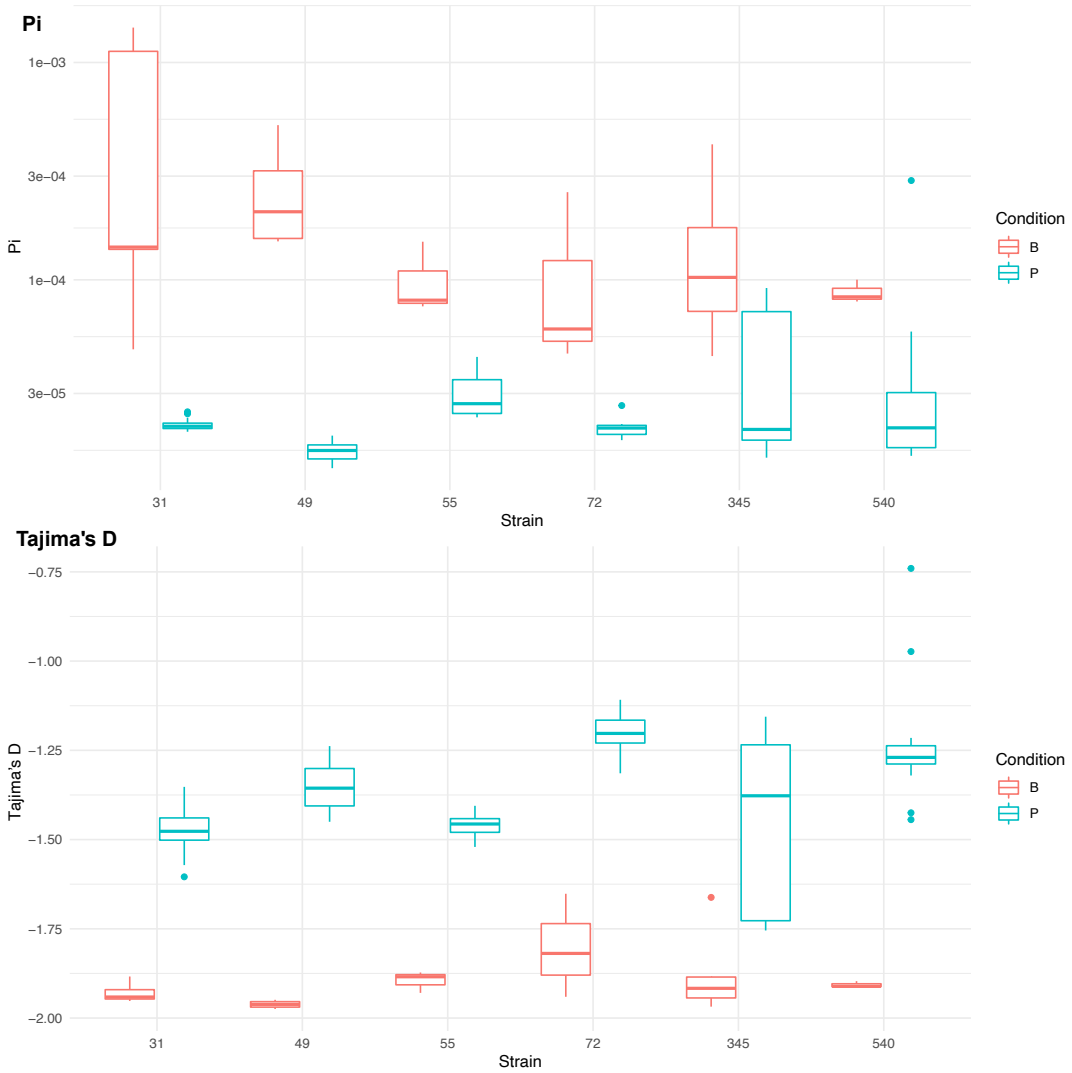
	P0	P4	P8	P12	P16	P20
MT31	sequencing wet weight growth curve SEM		sequencing wet weight growth curve SEM	sequencing	sequencing	sequencing
MT49	sequencing wet weight growth curve SEM		sequencing wet weight growth curve SEM	sequencing	sequencing	sequencing
MT72	sequencing wet weight growth curve SEM	sequencing	sequencing wet weight growth curve SEM	sequencing	sequencing	
MT345	sequencing wet weight growth curve SEM	sequencing	sequencing wet weight growth curve SEM	sequencing	sequencing	
MT55	sequencing wet weight growth curve SEM	sequencing	sequencing wet weight growth curve SEM			
MT540	sequencing wet weight growth curve SEM	sequencing	sequencing wet weight growth curve SEM			

1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204

1205 **Supplementary Table 2:** Variants with a >= 30% frequency change over the course of four planktonic passages.
1206 Grouped by mutations within the same gene and sorted by genes that acquired the most independent SNPs
1207 across replicates and strains.

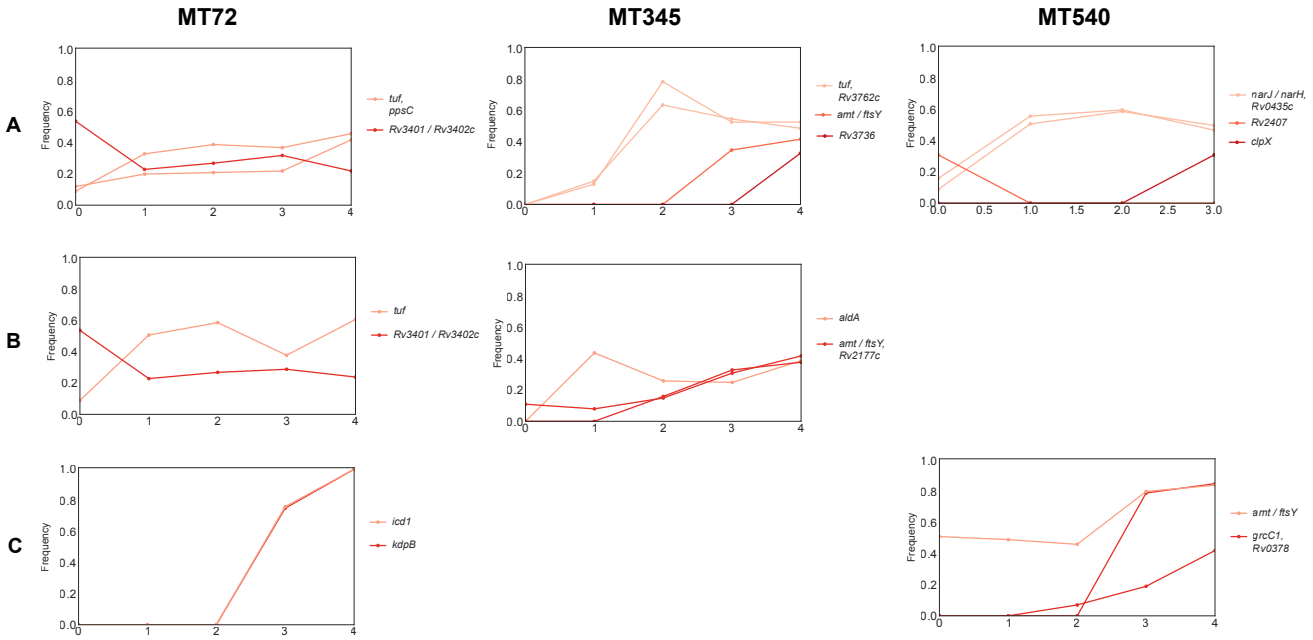
Gene	Strain(s)	Num. mutations	Position(s)	Frequency change(s)	Type(s)	Mutation(s)	Product	Proportion in global dataset
tuf	72, 345	3	784878, 785931	31, 49, 52%	Non-synonymous	H20D (CAC-GAC), M371V (ATG-GTG)	elongation factor Tu	0, 0
Rv2177c	55, 345	3	2439519	30, 31, 34%	Synonymous	R143R (CGC-CGA)	transposase	0.26
amt / ftsY	345, 540	3	3232703, 3232759	33, 38, 42%	Intergenic	-197 / +168 (G-A), -253 / +112 (G-A)	ammonium transport integral membrane protein / cell division protein signal recognition particle receptor	0.63, 0.63
grcC1	540	2	653223	41, 85%	Non-synonymous	F151L (TTC - TTA)	polyprenyl-diphosphate synthase	6.7E-05
Rv3401 / Rv3402c	72	2	3820545	30, 32%	Intergenic	+143 / +108 (A-G)	conserved enzyme / conserved hypothetical	0.77
Rv0378	540	1	455737	42%	Synonymous	A34G (GCC-GGC)	conserved hypothetical	0.00024
Rv0435c	540	1	524164	31%	Non-synonymous	G124S (GGC-AGC)	ATPase	0
aldA	345	1	860984	39%	Non-synonymous	F25V (TTC-GTC)	aldehyde dehydrogenase	0
kdpB	72	1	1154704	100%	Synonymous	G327G (GGC-GGT)	potassium transporting ATPase subunit B	0
narH / narJ	540	1	1292786	41%	Intergenic	+45 / -617 (A-G)	nitrate reductase subunit beta / nitrate reductase subunit delta	0
Rv2407	540	1	2704886	31%	Non-synonymous	G64S (GGC-AGC)	conserved hypothetical	0
clpX	540	1	2758927	31%	Non-synonymous	V188I (GTC-ATC)	ATP dependent CLP protease ATP-binding subunit	0
ppsC	72	1	3258368	30%	Non-synonymous	H895P (CAC-CCC)	Phenolpthiocerol synthesis type-I polyketide synthase	0.0018
icd1	72	1	3725179	100%	Nonsense	Y222* (TAC-TAA)	isocitrate dehydrogenase	4.4E-05
Rv3736	345	1	4186898	33%	Non-synonymous	A89P (GCC-CCC)	araC/xylS family transcriptional regulator	0.00027
Rv3762c	345	1	4207642	53%	Non-synonymous	P412R (CCG-CGC)	hydrolase	0

1208

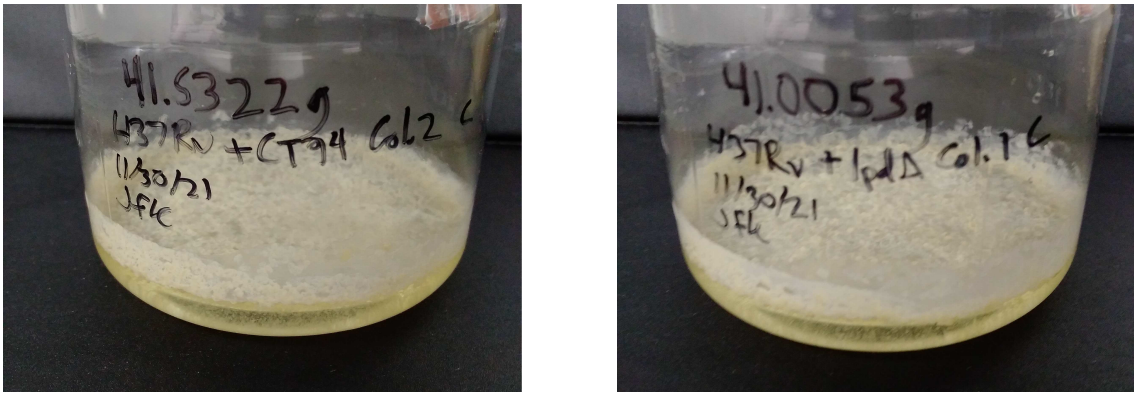


Supplementary Figure 2: Population genetics statistics calculated from Pool-seq data from all strains at all sequenced passage points. Range of values across all sequenced passage points given for each statistic. Boxplots are colored by strains passaged as biofilms (red) or passaged as planktonic cultures (blue).

SupplementaryData_1: Mutations present in ancestral pairs of strains which went on to develop convergent mutations in response to biofilm growth (MT31/MT55 and MT49/MT540). 'Ref' and 'alt' refer to the alleles present in the reference strain (H37Rv) and the clinical isolate, respectively. Mutations without an annotation exist in intergenic regions.

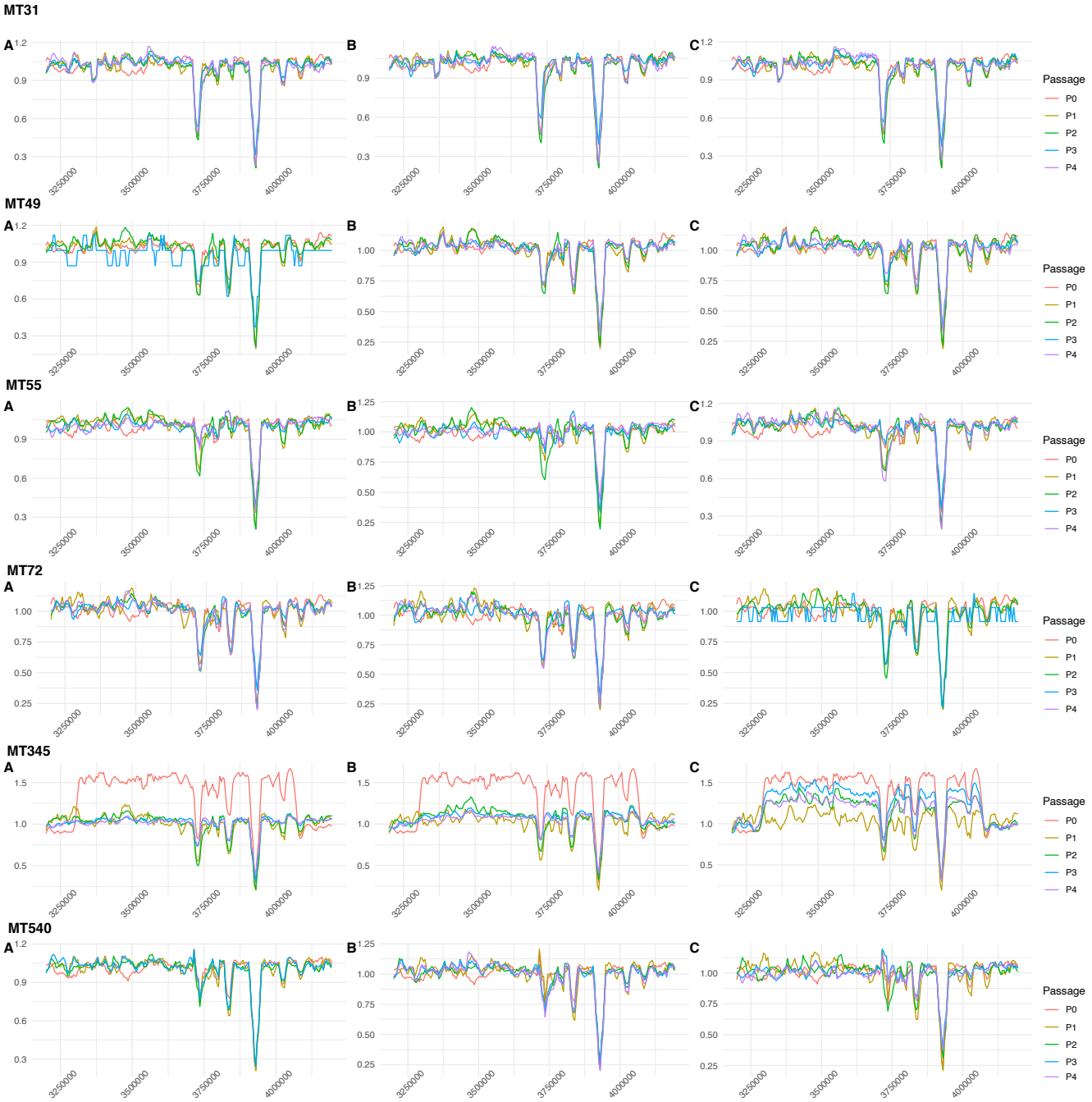


Supplementary Figure 3: Trajectories of mutations that changed =>30% over the course of planktonic passaging. In contrast to strains evolved under biofilm conditions, only 2 mutations became fixed (both in MT72-C).



H37Rv + CT94 H37Rv + CT94-lpdA

Supplementary Figure 4: H37Rv pellicles with empty vector (left) and extra copy of *lpdA* (right) introduced into the chromosome using pCT94. Qualitative pellicle phenotypes remain the same after introduction of additional *lpdA*.

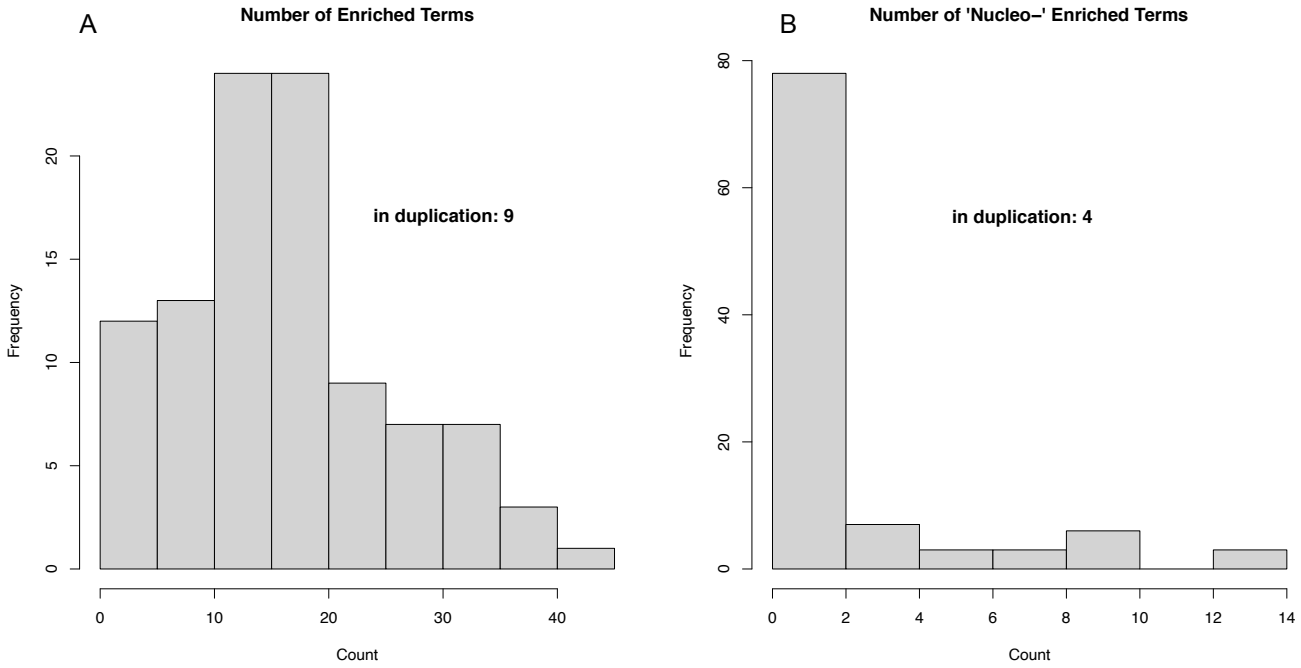


Supplementary Figure 5: Sliding-window plots showing relative sequencing coverage of planktonically passaged isolates (3 independently evolving replicates per strain). The region in which a convergent ~120kb duplication was identified in strains MT31 and MT55 after passaging as pellicles doesn't appear to have been replicated in strains evolved under these conditions. The one exception is MT345 passage 0, which acquired a duplication in this region during regular laboratory handling. This duplication is even larger than that in MT31 and MT55 (~700kb) and exists at intermediate frequency in the ancestral strain, before disappearing in 2/3 evolved replicates. In the final replicate, the duplication fluctuates but remains at intermediate frequency after four passages.

1249 **Supplementary Table 3**: Significantly (*P*-value <0.01) enriched GO terms within the large duplication that arose
1250 in L4.9 strains. Of note is enrichment of nucleotide biosynthesis/metabolic processes.
1251

GO.ID	Term	Annotated	Observed	Expected	classicFisher
GO:0009116	nucleoside metabolic process	31	6	1.29	0.0014
GO:1901657	glycosyl compound metabolic process	43	7	1.79	0.0017
GO:0034404	nucleobase-containing small molecule biosynthetic process	25	5	1.04	0.0031
GO:1901136	carbohydrate derivative catabolic process	26	5	1.08	0.0037
GO:0055086	nucleobase-containing small molecule metabolic process	170	15	7.08	0.0037
GO:0034654	nucleobase-containing compound biosynthetic process	352	25	14.66	0.0038
GO:0018130	heterocycle biosynthetic process	454	29	18.91	0.0083
GO:0000725	recombinational repair	11	3	0.46	0.0091
GO:0052646	alditol phosphate metabolic process	11	3	0.46	0.0091

1252
1253
1254



1255
1256 **Supplementary Figure 6**: Results of 100 permutations of random GO term enrichment analysis (see Methods).
1257 Histograms depicting the number of significantly (*P*-value <0.01) enriched GO terms (A) as well as the number of
1258 enriched terms involving nucleotide metabolism processes (B).