Parallel in vivo experimental evolution reveals that increased stress resistance was key for the emergence of persistent tuberculosis bacilli

Aideen C Allen, Wladimir Malaga, Cyril Gaudin, Arnaud Volle, Flavie Moreau, Ali Hassan, Catherine Astarie-Dequeker, Antonio Peixoto, Rudy Antoine, Alexandre Pawlik, et al.

Parallel experimental evolution in mice reveals that increased stress resistance was a key event for the emergence of persistent tuberculosis bacilli

Aideen C. Allen[1], Wladimir Malaga[1], Cyril Gaudin[2,3], Arnaud Volle[1], Flavie Moreau[1], Ali Hassan[1], Catherine Astarie-Dequeker[1], Antonio Peixoto[1], Rudy Antoine[2], Alexandre Pawlik[4], Wafa Frigui[4], Céline Berrone[1], Roland Brosch[4], Philip Supply[2], and Christophe Guilhot[1]*

[1] Institut de Pharmacologie et de Biologie Structurale (IPBS), Université de Toulouse, CNRS, UPS, Toulouse, France

[2] Inserm U1019, CNRS UMR8204, Université de Lille, Institut Pasteur de Lille, Center for Infection and Immunity, Lille 59000, France

[3] Genoscreen, Lille, France

[4] Institut Pasteur, Unit for Integrated Mycobacterial Pathogenomics, CNRS UMR3525, Paris 75015, France

*Corresponding author: Christophe.Guilhot@ipbs.fr

Abstract

Pathogenomic evidence suggests that *Mycobacterium tuberculosis* evolved from an environmental ancestor similar to *Mycobacterium canettii*, a rare human pathogen. However, the adaptations responsible for the transition from an environmental mycobacterium to an obligate human pathogen are poorly characterised. The ability to persist in the human host appears to be one important trait. Here we set out to identify the adaptations that contribute to the evolution of persistence in *M. tuberculosis.* By using experimental evolution of 8 *M. canettii* populations in mice, we selected mutants with enhanced persistence *in vivo* when compared to parental strains, which were thus phenotypically closer to *M. tuberculosis*. Genome sequencing of 140 *M. canettii* mutants and complementation analysis revealed that mutations in two loci were responsible for the enhanced persistence phenotypes. Most of the tested mutants were more resistant than their parental strains to nitric oxide, which is an important effector of immunity against *M. tuberculosis* infection. This resistance was common to modern *M. tuberculosis* strains but not to *M. canettii* strains. Our findings demonstrate phenotypic convergence during the experimental evolution of *M. canettii*, which mirrors natural evolution of *M. tuberculosis.* Furthermore, they indicate that the ability to withstand host-induced stresses, such as nitric oxide, was key for the emergence of persistent *M. tuberculosis*.

Introduction

The success of *Mycobacterium tuberculosis* (*MTB*), the primary aetiological agent of human tuberculosis (TB), is largely attributed to its ability to evade host defences and thus persist for months within the lungs. Indeed, it is estimated that one quarter of the world's population has been infected with this pathogen and a substantial fraction still carries viable bacilli[1,2]. The adaptations that allowed the ancestor of *MTB* to colonise humans and evolve into the persistent pathogen that now causes an estimated 1.5 million deaths each year remain mostly uncharacterised[2,3].

In contrast to *MTB*, *Mycobacterium canettii* are TB bacilli of only marginal epidemiological importance. These rare human isolates can cause disease clinically indistinguishable from classical TB, however epidemiological data indicate that they are not fully adapted for colonisation of the human host[4-6]. Consistently, *M. canettii* strains are less able to persist and cause disease in animal models than *MTB* strains[7-9]. During evolution, these strains diverged early from *MTB* and they are now believed to have an environmental reservoir[5,10].

The current model is that *MTB* evolved by clonal expansion from an *M. canettii*-like progenitor through adaptations that rendered them persistent in mammalian hosts[9,11,12]. However, virtually all identified *MTB* factors important for establishing chronic infection are also present in *M. canettii*[9], and few genes and loci are specific to *MTB*. It is therefore likely that the strict pathogenic phenotype differentiating *MTB* from facultative human pathogens like *M. canettii* and other environmental mycobacteria is due to modification of shared pathways. The *M. canettii* strains are therefore an outstanding resource for investigating the molecular mechanisms driving enhanced persistence in tuberculosis bacilli, and for gaining new insights into the evolutionary events involved in *MTB* emergence.

Following aerosol or intravenous (i.v.) infection of mice, *M. canettii* and *MTB* strains multiply in the lungs similarly during the acute phase of infection (first 4 weeks). In the chronic phase of infection, *MTB* persists and the bacterial load is comparable to that achieved during the acute phase. In sharp contrast, the lung bacterial load for *M. canettii* strains declines substantially during the chronic

infection phase, and for STB-K, the most attenuated *M. canettii* isolate in this model, becomes undetectable after 30 weeks of infection. This lower persistence is associated with less pathology[9], possibly explaining why there is no direct transmission in humans reported for *M. canettii*[4-6].

Our objective for this study was to select *M. canettii* mutants with enhanced persistence *in vivo* in order to identify what adaptations are associated with this phenotypic gain. We first implemented a positive selection strategy based on unbiased experimental evolution of 8 *M. canettii* populations in a mouse model. Then we used a combination of genome sequencing and phenotypic characterisation of multiple evolved clones to identify and describe the genomic and functional adaptations associated with enhanced persistence of *M. canettii*. Here we show that increased resistance to host-derived stresses, and especially nitric oxide (NO), is a key adaptation required by *M. canettii* to survive for several months *in vivo*.

Results

Unbiased *in vivo* experimental evolution of *M. canettii* selects mutants with enhanced persistence

It has previously been established that *M. canettii* are less persistent than *MTB* in different mouse models[8,9]. We therefore set up an experimental evolution procedure in mice to select for *M. canettii* mutants with an improved ability to persist, thus resembling *MTB* strains. These mutants should outcompete parental *M. canettii* strains in competitive experiments in mice.

Our experimental evolution strategy consisted of serial passages in BALB/c mice (**Fig. 1a**) starting from two distinct *M. canettii* isolates; the STB-K and STB-D strains which are phylogenetically farthest from and closest to *MTB* respectively[9] (**Supplementary Fig. 1**). For both STB-K and STB-D, we generated four evolution lineages (A, B, C and D) that evolved independently and were never mixed. We performed 15 cycles for STB-K and 6 cycles for STB-D. At each cycle, we evaluated the lung and spleen bacterial loads at days 1, 28 and 56 post-infection, and we calculated a persistence index corresponding to the

log ratio of lung bacterial burden at day 56 *vs* day 28 (**Fig. 1b**). For both STB-K and STB-D, we found that the persistence index increased throughout the cycles, suggesting that mutants with enhanced abilities to persist emerged. To confirm this, we randomly collected clones from agar plates used to evaluate the bacterial burden in the lungs during the experimental evolution cycles, and intranasally (i.n.) infected BALB/c mice with 24 individual clones recovered after cycles 6, 9, 12 and 15 for strains derived from STB-K, or cycle 6 for STB-D. We found that all the tested clones caused an increased bacterial burden in the lungs and spleen relative to their parental strains (**Fig. 1c,d** and **Supplementary Fig. 2**). This was most pronounced at later time points, indicating that mutants with enhanced persistence were selected during experimental evolution of both *M. canettii* STB-K and STB-D.

Small numbers of mutations were fixed in the populations during the experimental evolution

To identify the mutations that arose in STB-K and STB-D during experimental evolution in mice, we sequenced the whole genomes of 100 individual clones derived from STB-K and 40 clones derived from STB-D, and searched for differences relative to their parental strains.

In the STB-K series, we found that mutations in two loci were fixed independently in the four evolution lineages. An identical nucleotide substitution C>T at nucleotide position 1582889, corresponding to the gene *BN42_21248,* was found in all clones recovered after cycle 6 (**Supplementary Table 1**). This mutation changed a single amino acid P96L in the protein it encodes, which is the orthologue of Rv1339, a *MTB* H37Rv protein, which was recently shown to encode a phosphodiesterase degrading cyclic-AMP (cAMP)[13] (**Fig. 2a**). In the second locus, three mutations became additionally fixed independently in evolution lineages A, B and C: two single nucleotide insertion/deletions (indels) and a 3073 bp deletion. The two small indels at positions 2112495 and 2113560 induce frameshifts in two adjacent genes, *pe18* and *ppe26* respectively. The large deletion resulted in removal of the genes *pe18* and *ppe26,* and fusion of the 5'-terminal part of *ppe25* with the 3'-terminal part of *ppe27*, probably due to homologous recombination (**Fig. 2b**).

In the STB-D series, the analyses revealed a more complex situation. Indeed, we identified single nucleotide polymorphisms (SNPs) present at relatively high frequency in the inoculum used for the first cycle, suggesting three genetic profiles (**Supplementary Table 2**). Other SNPs were fixed independently in these three different genetic backgrounds in orthologues of the genes *rv0200*, *rv0246* and *rv2990c* in evolution lineages A, B and C respectively (**Supplementary Table 2**). No mutation was clearly selected in the evolution lineage D.

These findings indicated parallel evolution in the different lineages derived from STB-K, whereas in the STB-D background, such patterns were not visible in the framework of our experimental setting. We decided to focus our studies on STB-K clones, given the limited number of mutations identified and their apparent strong selective advantage. We first analysed the appearance and kinetics of mutation selection in the *rv1339* orthologue and *pe18-ppe26* locus. Whole-genome sequencing (WGS) of the bacterial populations from each evolution lineage and each cycle revealed that the *rv1339* orthologue mutation was fixed very early, before cycle 4, in the four evolution lineages (**Fig. 2c**). In contrast, mutations at the *pe18-ppe26* locus became fixed in the population later, between cycles 9 and 13 (**Fig. 2d**). These later mutations were only selected after fixation of the *rv1339* orthologue mutation. Rapid emergence and parallel fixation of an identical SNP in the *rv1339* orthologue in the four evolution lineages suggested that it may have pre-existed at low frequency in the initial STB-K culture. However, deep sequencing did not detect the mutation above background variant frequency (below 0.1%) in the starting STB-K inoculum (**Fig. 2c**). As our experimental evolution strategy included a 14 day *in vitro* culture between each cycle, we hypothesised that the *rv1339* orthologue mutation might have provided a selective advantage for growth *in vitro*. However, WGS of cultures subjected to 6 cycles of *in vitro* experimental evolution showed that this mutation was not enriched in the population during these passages (**Fig. 2e**). Thus we concluded that the mutation in the *rv1339* orthologue provides a strong *in vivo* selective advantage in the STB-K genetic background, whereas mutations at the *pe18-ppe26* locus may provide an additional advantage only when *rv1339* is mutated.

The persistence phenotypes of *rv1339* and *pe18-ppe26* STB-K mutants are not specific to the mouse strain or route of infection

To further investigate the effects of the selected mutations in the STB-K background, we chose two STB-K-derived variants isolated after cycle 9: KC9A1 which only has the mutation in the *rv1339* orthologue (hereafter referred to as the *rv1339* mutation), and KC9C1 which has both the *pe18-ppe26* deletion and the *rv1339* mutation. To examine whether the persistence phenotypes of these strains were specific to the BALB/c mouse genetic background, we infected C3HeB/FeJ mice, which are very susceptible to TB infection[14,15], with either the parental STB-K strain, KC9A1, KC9C1, or *MTB* H37Rv. The bacterial burden in the lungs and spleen of C3HeB/FeJ mice was consistent with an enhanced persistence phenotype for KC9A1 and KC9C1 when compared to STB-K 77 days after i.n. infection (**Fig. 3a,b**), indicating that this phenotype was not restricted to a single mouse genetic background. Furthermore, we found that this phenotype was also not specific to the route of infection: C3HeB/FeJ mice infected i.v. with KC9A1 and KC9C1 also had substantially more bacteria in the lungs and spleen at 76 days post-infection than had mice infected with STB-K (**Fig. 3c,d**), despite exhibiting similar bacterial loads at 21 days post-infection (**Supplementary Fig. 3**). C3HeB/FeJ mice have been reported to develop well-organised and necrotic granulomas in the lungs following aerosol infection with *MTB* H37Rv[16,17]. Therefore, we examined if there was a difference in granuloma formation 76 days after i.n. infection with STB-K, KC9A1, KC9C1, and H37Rv. We found that H37Rv-infected mice had twice as much lung inflammation as had mice infected with the *M. canettii* strains (**Fig. 3e**), and only H37Rv-infected mice had mature granulomas in the lungs (**Supplementary Fig. 4**). There was, however, no significant difference in lung inflammation between the three *M. canettii* strains (**Fig. 3e**). Analyses of the cell types present in granulomas showed no difference between KC9A1 and STB-K, but KC9C1 exhibited a pattern more similar to H37Rv, i.e. higher percentages of $CD3^+$ cells and neutrophils, and lower percentages of $B220^+$ cells than for STB-K (**Supplementary Fig. 5**). Similar concentrations of IFN-

γ, TNF and IL-10 were found in the organ homogenates from mice 21 and 76 days after i.v. infection with the different strains (**Supplementary Fig. 6**). These results showed a convergence in the immune response induced by the more persistent STB-K-derived mutant, KC9C1, and H37Rv. In contrast, the persistence gain of KC9A1 was not associated with a change in inflammatory response.

The *pe18-ppe26* and *rv1339* mutations confer enhanced persistence

To demonstrate that the *pe18–ppe26* deletion is directly responsible for the difference in persistence between KC9A1 and KC9C1, we complemented KC9C1 (which has both the *pe18–ppe26* deletion and the *rv1339* mutation) with a plasmid containing the *ppe25-ppe27* region of STB-K (KC9C1::K), thus rendering the strain genetically similar to KC9A1. We also constructed strains containing empty vectors (EV), KC9A1::EV and KC9C1::EV, to control for potential effects of the plasmid on strain fitness. We infected BALB/c mice i.n. with these strains and found that at 70 days post-infection, KC9C1::K persisted in the lungs to the same extent as KC9A1 and KC9A1::EV (**Fig. 4a**). In contrast, mice infected with KC9C1 or KC9C1::EV had approximately 10-fold more bacteria in the lungs (**Fig. 4a**). This trend was also reflected in the spleen (**Fig. 4b**). These data indicate that the *pe18-ppe26* deletion is responsible for the enhanced persistence phenotype of KC9C1 relative to KC9A1.

To extend these findings, we examined if the *pe18–ppe26* deletion selected after experimental evolution of STB-K could also confer enhanced persistence to another *M. canettii* strain. To do so, we deleted the *ppe25-ppe27* region in STB-D and complemented this strain with the *ppe25-ppe27* regions from either STB-D (::D), STB-K (::K), or KC9C1 (::C1) (**Supplementary Fig. 7**). BALB/c mice infected i.n. with these strains had very similar levels of bacteria in the lungs 70 days post-infection (**Supplementary Fig. 8**), indicating that either the *pe18-ppe26* deletion confers a persistence advantage in STB-K only, or that it requires an existing mutation in *rv1339* to have a significant impact.

To demonstrate that the *rv1339* mutation is directly responsible for the gain of persistence in KC9A1 and to examine if the gain associated with this mutation is specific to STB-K, we used the STB-D strain in which the *rv1339* orthologue is 100% identical in sequence to that in STB-K. First we knocked out the *rv1339* orthologue of STB-D (DΔrv1339), and then we complemented this strain by introducing a plasmid containing either the wild-type *rv1339* orthologue (::rv1339K), or the mutated *rv1339* identified in KC9A1 (::rv1339A1) (**Supplementary Fig. 9**). BALB/c mice infected i.n. with DΔrv1339 or ::rv1339A1 had substantially more bacteria in their lungs 77 days post-infection than had mice infected with STB-D or ::rv1339K (**Fig. 4c**). A similar phenomenon was also seen for ::rv1339A1 in the spleen (**Fig. 4d**). These data confirm that the *rv1339* mutation is responsible for the enhanced persistence phenotype of KC9A1 relative to STB-K, and suggest that it is a loss-of-function mutation. They also demonstrate that although the *rv1339* mutation arose during experimental evolution of STB-K, it can enhance persistence of the STB-D strain and is therefore possibly a general mechanism for increasing the persistence of *M. canettii* strains.

The *rv1339* mutation confers enhanced resistance to host-derived stresses

As seen above, the enhanced persistence of KC9A1 relative to STB-K does not seem to relate to a profound alteration of the immune response. Due to lymphocyte recruitment and phagocyte activation with IFN-γ, bacilli must withstand various stresses[18-21]. These include low pH, in the range of pH 4.5 – 5.5[19,22,23], nutrient restriction[24-26], metal intoxication[24,27,28], or physical attacks by oxygen or nitrogen radicals[29-31] and antimicrobial peptides[32-35]. Therefore, we examined if the selected mutants have an enhanced capacity to survive these various bactericidal mechanisms. Growth of STB-K, KC9A1 and KC9C1 at various neutral or acidic pH and in phosphate-limited media was similar for the three strains (**Supplementary Fig. 10**). H37Rv was more resistant to H_2O_2-induced stress than the three *M. canettii* strains, which resembled each other in terms of resistance (**Supplementary Fig. 10**). In contrast, both KC9A1 and KC9C1 were more tolerant than STB-K to high concentrations of copper and

to a lesser extent, zinc, two metals identified as important effectors of nutritional immunity against *MTB*[24,27,28] (**Fig. 5a, b**). Furthermore, both KC9A1 and KC9C1 were significantly more resistant to the NO donor DETA/NO than was STB-K (**Fig. 5c**).

To link increased resistance to stress with enhanced persistence *in vivo*, we examined the persistence of STB-K and KC9A1 in $NOS2^{-/-}$ mice which are deficient for NO production, and in bone-marrow derived macrophages (BMDMs) stimulated with IFN-γ and LPS and treated with L-NAME to inhibit NO production (**Supplementary Fig. 11**). We found that, at 56 days post-infection, KC9A1 had higher bacterial loads in the lungs and spleens of $NOS2^{-/-}$ mice than in control mice. A similar trend was observed for STB-K in the spleen. Consistently in BMDMs, treatment with L-NAME substantially increased the survival of both STB-K and the two mutants KC9A1 and KC9C1. Together these findings demonstrate that NO is an important factor reducing STB-K and KC9A1 persistence.

As KC9A1 and KC9C1 had similar levels of resistance to NO and tolerance to copper and zinc *in vitro*, we hypothesised that these phenotypes were associated with their common mutation in the *rv1339* orthologue. To test this, we treated STB-D, DΔrv1339, ::rv1339K, and ::rv1339A1 with DETA/NO for 48 hours. We found that DΔrv1339 and ::rv1339A1 were significantly more resistant to killing by NO than ::rv1339K and STB-D (**Fig. 5d**), demonstrating that the *rv1339* mutation is responsible for this phenotype.

As the first barrier to exogenous toxic compounds is the bacterial cell envelope, we examined whether the *rv1339* mutation impacts its permeability. To address this question, we used the non-pathogenic *Mycobacterium smegmatis* model, which carries a *rv1339* orthologue (*MSMEG_4902*). First, we disrupted this gene to generate smegΔrv1339 (**Supplementary Fig. 12**). Then we complemented this mutant strain by introducing a plasmid containing either the wild-type *rv1339* orthologue from STB-K (smeg::rv1339K) or the mutated *rv1339* from KC9A1 (smeg::rv1339A1). Next, we performed a permeability assay using ethidium bromide (EtBr) as a reporter molecule and found that both the

smegΔrv1339 mutant and the smeg::rv1339A1 complemented strain incorporated less EtBr in comparison to the wild-type and smeg::rv1339K strains (**Fig. 5e**).

Together, these findings indicate that the *rv1339* mutation decreases the cell envelope permeability and confers increased resistance to several stresses encountered by the pathogen during infection.

MTB exhibit higher NO resistance than *M. canettii* and this gain is important for persistence

Persistent bacteria arising from experimental evolution of STB-K had mutations in *rv1339*, whereas the persistent bacteria that evolved from the STB-D strain did not. As tuberculosis bacilli have acquired multiple mechanisms to resist host-derived stress[18-20,36-38], we examined the sensitivities of several persistent STB-D mutants selected after cycle 6 to NO as a representative stress. Two out of four persistent mutants tested, DC6C1 and DC6D1, were substantially more resistant to DETA/NO treatment than STB-D, and closely resembled *MTB* H37Rv (**Fig. 6a**). This indicated that phenotypic convergence occurred during experimental evolution of the genomically distant STB-K and STB-D strains, and highlighted that increased resistance to NO, and possibly to other stresses, are key mechanisms driving persistence *in vivo*.

To assess the ability of phylogenetically diverse *M. canettii* strains (**Supplementary Fig. 1**) to resist NO, we exposed five strains, STB-K, -D, -A, -J and -L, to DETA/NO for 48 hours. All the strains tested were significantly more sensitive to killing by NO than *MTB* H37Rv (**Fig. 6b**). We also tested the susceptibilities of three other *MTB* strains from various lineages to DETA/NO and found that there was no significant difference in the resistance of *MTB* strains to NO with the exception of the lineage 1 strain IO367, which was slightly more susceptible to killing than H37Rv only (**Fig. 6c**). These results thus demonstrate that *MTB* are more resistant than *M. canettii* to stresses, such as NO and indicate that increased NO resistance is associated with enhanced persistence *in vivo*.

Discussion

Parallel molecular evolution and phenotypic convergence are canonical signatures of genes and processes under strong adaptive selection, such as those related to pathogen adaptation to host environments[39-43]. Here, the identification of parallel and convergent evolution during experimental evolution of ancestor-like *MTB* strains in mice highlights that a key patho-adaptation differentiating modern *MTB* from their environmental ancestor is enhanced resistance to stress, especially NO. NO and other reactive nitrogen intermediates are produced by macrophages upon activation by T cell-derived IFN-γ, and are major effectors in the immune response to TB[29,31,44,45]. These radicals are believed to protect TB-infected hosts through direct bacterial killing[46,47], and inhibition of granulocytic inflammation and immunopathology[48,49]. Enhancing resistance to these compounds, and possibly other toxic molecules generated during the immune response such as metal ions, appears to be required to increase persistence of ancestor-like *MTB* strains in the lungs, and was probably a major step in the emergence of *MTB*.

In this study, we found that diverse mutations acquired by *M. canettii* STB-K and STB-D through experimental evolution produce this phenotype. One of these mutations was in the *rv1339* orthologue, and it conferred both STB-K and STB-D strains with enhanced resistance to NO. The gene *rv1339* is highly conserved in TB bacilli (**Supplementary Fig. 13**) and other mycobacteria and a recent study revealed that the Rv1339 protein degrades cAMP, an important signalling molecule controlling peptidoglycan metabolism[13]. Inactivation of *rv1339* or expression of the mutated *rv1339* allele from KC9A1 increased stress resistance in *M. canettii* STB-D and STB-K and reduced envelope permeability in *M. smegmatis*. Consistently *MTB* strains containing point mutations or transposon insertions in *rv1339* have increased resistance to antitubercular compounds[50,51] and a reduced capacity to import fatty acids during macrophage infection[52]. Therefore, the likely mechanism of enhanced persistence associated with the *rv1339* mutation is a decrease in envelope permeability allowing bacilli to withstand toxic stresses such as NO or metals. The other mutations selected in the STB-K-derived

278 persistent clones concerned *pe18* and *ppe26* genes located within the *esx-5* locus, which encodes one
279 of the five type VII protein secretion systems of *MTB*. Interestingly, clinical *MTB* isolates successfully
280 spreading in Tunisia harbour a single *pe-ppe* gene pair in the *esx-5* locus that likely arose by
281 homologous recombination between *ppe25-ppe27* genes[53] as in the mutant KC9C1. Furthermore,
282 *MTB* strains from the Beijing family exhibit *esx-5* locus mutations disrupting secretion of a large
283 number of PE/PPE proteins and enhancing their virulence[54]. This convergence between experimental
284 evolution of ancestor-like *MTB* strains and the natural evolution of modern TB strains suggests that
285 the loss of certain PE/PPE proteins is another method exploited by TB bacilli to better survive and
286 enhance their virulence in mammalian hosts. The precise function of PE18-PPE26 is unknown but
287 PE/PPE proteins were recently found to mediate solute-specific import across the *MTB* outer
288 membrane[52,55]. Therefore, *pe18-ppe26* mutations in STB-K might also limit the penetration of toxic
289 molecules allowing the *M. canettii* persistent mutants to better survive the immune response.
290 However, we found no evidence of enhanced resistance to stress associated with the *pe18-ppe26*
291 mutation. Therefore, an alternative explanation might relate to the difference in immune cell
292 recruitment observed in mice infected with the KC9C1 mutant that more closely resembles the
293 situation in H37Rv-infected mice.

294 In this study, we identified a major driving force for patho-evolution of TB bacilli and illustrated that
295 the required adaptations can be acquired by exploiting diverse bacterial pathways. This supports the
296 concept that phenotypic convergence can be achieved through diverse genetic modification due to
297 historical contingency. The associated loss-of-function mutational patterns apparent in STB-K further
298 highlight the key role of gene loss in driving evolutionary adaptation[56], in this case towards enhanced
299 persistence of a major pathogen. The rapid selection of *M. canettii* mutants with similar persistence
300 characteristics to *MTB* and diversity of mutations leading to a similar gain-of-function is surprising,
301 considering the accepted model that *MTB* emerged only once from an environmental ancestor[9]. Long-
302 term association with the human host requires both the ability to persist, and to transmit to
303 uninfected people. *M. canettii* strains have never been reported to be transmitted between humans[4].

304 Therefore, another major adaptation may be the gain of function(s) favouring aerosol transmission,
305 which remains to be identified.

Materials and methods

Mycobacterial strains and culture conditions

The *M. canettii, MTB* and *M. smegmatis* strains used in this study are listed in **Supplementary Table 3**. All strains were cultured in Middlebrook 7H9 broth (BD Difco) supplemented with 0.2% glycerol (w/v) and 10% Middlebrook albumin-dextrose-catalase enrichments (ADC; BD BBL), with 0.05% Tween 80 (Sigma) (7H9/ADC/Tween) unless otherwise described. To grow recombinant strains, the antibiotics hygromycin (Euromedex) and kanamycin (Euromedex) were added as required at final concentrations of 50 µg/ml and 40 µg/ml respectively. Cultures were agitated approximately twice per week and maintained for 3-4 weeks at 37°C before use. During the experimental evolution cycles, the antibiotic mixture PANTA (BD BBL) was added at a dilution of 25 µl for 10 ml 7H9/ADC/Tween for *in vitro* cultures. For growth on solid medium, 7H11 agar base (BD BBL) was supplemented with 0.5% glycerol and 10% Middlebrook oleic acid-albumin-dextrose-catalase (OADC; BD BBL). Counting of CFUs on 7H11 agar plates was performed after 2, 3 and 4 weeks of incubation at 37°C.

Experimental evolution strategy

The experimental evolution procedure consisted of independent serial passages of strain STB-K (phylogenetically farthest from *MTB*) and strain STB-D (phylogenetically closest to *MTB*) in BALB/c mice. For both STB-K and STB-D, four independent cultures were passaged corresponding to evolution lineages A, B, C and D, and they were never mixed. At cycle 1, 12 mice were infected i.n. with 10^4 CFU of STB-K or STB-D. STB-infected mice were then euthanised after 1 and 28 days of infection (4 mice per time point) and the CFU in the lungs and spleens were enumerated to evaluate the initial bacterial load and the multiplication during the acute phase of the infection. The 4 remaining STB-infected mice were kept until 56 days post-infection and then sacrificed to determine CFU at the chronic infection stage. The bacterial load in the organs was evaluated by plating part of the lung and spleen

330 homogenates on solid medium in order to get information on the entire population. Part of the lung
331 homogenates recovered from the 4 mice 56 days post-infection was also used to initiate four separate
332 cultures (evolution lineages A, B, C and D) which were allowed to grow for 14 days at 37°C before
333 being used as the new inocula for the next cycle. Each subsequent cycle was initiated by i.n. infecting
334 12 mice with approximately 10^4 CFU of the new STB-K or STB-D inocula prepared from cultures of
335 bacterial populations obtained 56 days post-infection in the previous cycle. This process of infection
336 and CFU monitoring was repeated for 15 cycles for STB-K and 6 cycles for STB-D. At cycles 3, 6, 9, 12
337 and 15 of STB-K, 20 individual colonies were picked up, briefly cultivated and stored at -80°C
338 (corresponding to 100 clones total). 20 of these individual clones (8 selected at cycle 6 and 4 selected
339 at cycles 9, 12 and 15) were characterised individually for their persistence in animal models. At cycles
340 3 and 6 of STB-D, 20 individual colonies were similarly picked up, briefly cultivated and stored at -80°C
341 (corresponding to 40 clones total). 4 of these individual clones (selected at cycle 6) were characterised
342 for their persistence in animal models.

343

344 **Sequencing and sequence analyses**

345 We performed Illumina-based WGS on 100 individual clones derived from STB-K (recovered at cycles
346 3, 6, 9, 12 and 15) and 40 clones derived from STB-D (recovered at cycles 3 and 6). Genomic DNA from
347 each clone was extracted from stationary phase culture using either the DNeasy Blood & Tissue kit
348 (Qiagen) or the Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research) according to the
349 manufacturer protocols.

350 For whole genome sequence analyses, CLC Genomics Workbench, version 20.0 (QIAGEN, Aarrhus) was
351 utilised for sequence mapping, using the global alignment option and ignoring reads with more than
352 one match, and variant calling, using the basic variant calling mode with a minimum frequency filter
353 set to 85%, and other parameters set to default. Mapping was performed against previously
354 assembled reference genomes of STB-K and STB-D (accession number FO203509.1 and NC_019950.1

355 respectively), and a newly generated dataset obtained by PacBio for STB-K. WGS data generated from
356 STB-K and STB-D parental strain batches, used for the first cycle of experimental evolution, were used
357 to correct for SNPs that were pre-existing in the genome assemblies (i.e. before experimental
358 evolution). Manual curation was subsequently performed to discard probable false positive variants
359 typically identified in repetitive regions (e.g. PE/PPE protein-encoding genes), including variants found
360 within a 12 bp-distance range. The false positive variant calling rate with this analysing pipeline was
361 verified by mapping sequence reads from *MTB* H37Rv genomic DNA against a previously assembled
362 reference genome obtained after PacBio sequencing of the same genomic DNA preparation, resulting
363 in absence of any detectable SNP or indel.

364 For the large deletion identified in the *pe18-ppe26* gene region of STB-K evolution lineage C, a specific
365 read mapping analysis was performed against the parental reference sequence using bowtie2 version
366 2.3.5[57]. Proportions of bacterial wild-type versus mutant (i.e. carrying this deletion) subpopulations
367 over the different cycles were determined after normalisation of the coverage depth detected around
368 the deleted region, by dividing this coverage depth by the mean genome wide coverage depth
369 obtained for the cycle considered. The proportion detected in the parental strain was also set at 100%,
370 to better visualise the progressive fixation of the deletion, starting from the initial population.

371 To search for *rv1339* variants among TB bacilli isolates, we analysed a collection of 6167 genome
372 assemblies available as of January 2020 from the NCBI by BLASTN search-based screening and we
373 looked at the distribution of mutations fixed in the *rv1339* gene in isolates belonging to lineages 1–8
374 or the animal lineages of the *MTB* complex.

375

376 **Strain and plasmid construction**

377 The plasmids and primers used in this study are listed in **Supplementary Tables 4** and **5**, respectively.

For the construction of the *BN44_11505* (*rv1339* orthologue) mutant in *M. canettii* STB-D, a 2kb DNA fragment overlapping the *BN44_11505* gene was amplified by PCR using primers 39a and 39b and cloned into the plasmid pJET (ThermoFisher Scientific) to give plasmid pWM355. A 0.8kb PCR fragment containing the kanamycin (*km*) gene was amplified from plasmid pET26b (Addgene) using primers km1 and km2 and inserted into the EcoRV site of plasmid pWM355, which is located inside the *BN44_11505* gene, to generate plasmid pWM359. The substrate for allelic exchange was produced by PCR amplification of a 2.8kb fragment, carrying the *BN44_11505* gene disrupted by the *km* cassette, from pWM359. This DNA fragment was electroporated in a recombinant *M. canettii* STB-D expressing the recombineering system[58] from plasmid pJV53H and transformants were selected on km-containing plates. Ten km-resistant clones were randomly picked and analysed by PCR using primers 39c / kmF, 39d / kmR and 39e / 39f. A clone giving the expected PCR profile (**Supplementary Fig. 9**) was selected for further experiments and named PMM305 (or DΔrv1339).

For the construction of the DΔppe25-ppe27 mutant (also named PMM304), three PCR fragments were amplified and fused. First, a 1kb fragment upstream of *ppe25* was generated from STB-D genomic DNA by PCR amplification using primers esx1 and esx2 which carry DraIII restriction site at their 5'- end. A second 1kb fragment downstream of *ppe27* was produced using primers esx5 and esx6 which also carry DraIII restriction sites at their 5'- end. Finally, the third fragment corresponding to the *km* gene was amplified using primers k1 and k2 which add DraIII sites on both side of the *km* gene. The three PCR fragments were digested using the restriction enzyme DraIII for 1h at 37°C. The three fragments were then ligated and the allelic exchange substrate corresponding to the ligated PCR fragments was amplified by PCR as a 2.8kb product using primers esx1 and esx6. This DNA product was transferred by electroporation into STB-D:pJV53H after induction of phage recombinase using acetamide as previously described[58]. Transformants were selected on km-containing plates. Ten colonies were randomly picked and analysed by PCR using primers esx7 / kmR, esx8 / kmF and 26a / 26b. A clone giving the expected PCR profile (**Supplementary Fig. 7**) was selected for further experimentation and named PMM304 (or DΔppe25-ppe27).

For the construction of the *MSMEG_4902* (*rv1339* orthologue) mutant in *M. smegmatis* mc^{2}155, two 1kb DNA fragments located upstream and downstream of *MSMEG_4902* gene were amplified by PCR using either primers 4902A and 4902B, or primers 4902C and 4902D, and then purified. The two fragments were ligated before amplification using primers 4902E and 4902F to generate a 2 kb DNA fragment. This fragment was cloned into the plasmid pJET. A 0.8kb PCR fragment containing the *km* gene was amplified from plasmid pET26b using primers km1 and km2 and inserted into the EcoRV site located at the junction of the fused 1kb fragments to generate plasmid pAH06. The substrate for allelic exchange was produced by PCR amplification of a 2.8kb fragment, carrying the *MSMEG_4902* gene disrupted by the *km* cassette, from pAH06. This DNA fragment was ligated into the EcoRV site of plasmid pJQ200[59] to generate plasmid pAH10. This plasmid was electroporated into *M. smegmatis* mc^{2}155 and transformants were selected on km-containing plates. One transformant was used to grow a culture and then plated on solid medium containing km and sucrose 5%. Ten km-resistant clones were analysed by PCR using primers 4902D / kmF, 4902A / kmR, and 4902A / 4902G. One clone giving the expected PCR profile (**Supplementary Fig. 12**) was selected for further experiments and named PMM280 (or smeg Δrv1339).

The *BN42_21248* gene from *M. canettii* STB-K (orthologue of *rv1339* in *MTB* H37Rv) or the mutated allele from KC9A1, were amplified from the corresponding strains using primers 1339G and 1339H. They were then inserted between the Nde1 and Spe1 sites of a derivative of the integrative plasmid pMV361[60] carrying a hygromycin resistance gene (*hyg*) and the mycobacterial promotor *pBlaF**[61], to give the complementation plasmids pAH01H (*BN42_21248* mutated) and pAH02H (*BN42_21248* wild-type). In pAH01H and pAH02H, the wild-type *BN42_21248* or the mutated variant were expressed under the control of the heterologous promotor *pBlaF**. These plasmids were transferred independently into the strain PMM305 by electroporation to give PMM305::pAH01H (abbreviated ::rv1339A1) and PMM305::pAH02H (abbreviated ::rv1339K). Plasmid pAH01 (carrying the gene *BN42_21248* mutated from KC9A1) or pAH02 (carrying the wild-type gene *BN42_21248* mutated from STB-K) were also transferred independently into the strain PMM280 by electroporation to give

PMM280::pAH01H (abbreviated smeg::rv1339A1) and PMM280::pAH02H (abbreviated smeg::rv1339K).

Similarly, the *ppe25-ppe27* genes from *M. canettii* STB-K or the corresponding allele from KC9C1 (carrying the *pe18-ppe26* deletion) were produced by PCR amplification of the corresponding genomic region using primers esxA and esxB. These regions were then inserted between the XbaI/HindIII restriction sites from the pMV361 derivative carrying the *hyg* gene to give the complementation plasmids pWM361 (*ppe25-ppe27* region from STB-K) and pWM386 (*ppe25-ppe27* region from KC9C1). The corresponding genomic regions from STB-D were amplified using primers esxC and esxD and cloned in the same vector to generate pWM385 (*ppe25-ppe27* region from STB-D). These three complementation plasmids, or the corresponding empty vectors (EV), were transformed into strains KC9A1, KC9C1, PMM304, or STB-D.

Animal infection studies

The persistence capacity of the *M. canettii* mutants and recombinant strains was assessed in four different mouse models. 6-8 week old female BALB/c mice (Janvier; 3/4 mice per group per time point) were infected i.n. with approximately 10^3 CFU per strain. 6-10 week old female C3HeB/FeJ mice (bred in-house in the IPBS animal facility; 4/5 mice per group per time point) were infected either i.n. with approximately 200 CFU per strain or i.v. with approximately 10^3 CFU per strain. All mice were anaesthetised prior to infection: BALB/c mice were injected intraperitoneally (i.p.) with a combination of valium, atropine and imalgene, and C3HeB/FeJ mice were injected i.p. with a ketamine-xylazine mixture for i.n. infection, or were exposed to isoflurane for i.v. infection. For experiments with $NOS2^{-/-}$ mice, 7-15 week old $NOS2^{-/-}$ mice (gift from E. Russo and P. Bousso, Institut Pasteur) or 6-8 week old C57BL/6 control mice (Janvier) were infected via aerosol with approximately 100 CFU per strain. Mice were housed under specific pathogen-free conditions. Mice displaying signs of health deterioration were euthanised by cervical dislocation. At the indicated time points, mice were

sacrificed by cervical dislocation and the lungs and spleens were isolated and homogenised using a GentleMACS Dissociator (Miltenyi Biotec) and gentlemacs M tubes. Bacterial load in the organs was estimated by plating serial dilutions of the lung and spleen homogenates on solid medium. For inoculum preparation, 3-4 week old bacterial cultures were re-cultured for 10 days in 7H9/ADC/Tween. 10 day old cultures were washed twice in PBS/0.05% Tween 80 and adjusted to the final desired concentration in PBS/0.05% Tween 80 for infection. Inoculum accuracy was verified by plating serial dilutions on 7H11 agar plates supplemented with 10% OADC.

Histopathology

For histopathological evaluation, C3HeB/FeJ mice were sacrificed by i.p. injection of Dolethal at 11 weeks post-infection, and lungs were isolated and fixed in 10% neutral buffered formalin for 48 hours. Following fixation, lungs were transferred into 70% ethanol and then delivered to the Experimental Histopathology service in Purpan Hospital, Toulouse (Anexplo). The lungs were embedded in paraffin for 5 µm sectioning and sections were stained with hematoxylin and eosin. Stained samples (one entire lung section for each mouse) were analysed using Pannoramic Viewer software to quantify the percentage lung inflammation per sample. For immunohistochemistry analyses, the epitopes were unmasked by HIER treatment for 20 minutes at 95°C, either in citrate buffer pH6 for B cell and neutrophil labelling or in citrate buffer pH9 for T-cell labelling. For macrophages, regeneration was performed by PIER treatment for 10 minutes at room temperature. Endogenous peroxidase and endogenous biotin were blocked by treatment with blocking solutions (S2023 Agilent for peroxidase and NB306 Clinisciences for biotin) for 10 minutes and 20 minutes respectively. Then the tissue sections were incubated with goat serum for 30 minutes. Specific labelling was performed by incubation for 1 hour with primary antibodies for macrophages (F4/80, MCA497G, Serotec; 1/100 dilution), B lymphocytes (B220/CD45r, MCA1258G, Serotec; 1/400 dilution) and T lymphocytes (CD3, ab16669, Abcam; 1/100 dilution) as well as overnight at 4°C for neutrophils (Neutrophil marker, sc-

71674, Santa Cruz biotechnology; 1/200 dilution). The lung sections were then incubated for 30 minutes with either biotinylated goat anti-rat secondary antibody for neutrophils, macrophages and B cells or goat anti-rabbit biotinylated antibody for T cells and then incubated with horseradish peroxidase (R.T.U., PK-7100, Vector Laboratories) for 30 minutes. The labelling was revealed in Diaminobenzidine (DAB) with a counter-stain of modified hematoxylin. Quantification of positive cells for F4/80, CD3, B220 and neutrophils markers per lesion was performed using the QuPath software[62]. More precisely, regions of interest were drawn for each individual lung lesion using the wand annotation tool. Next, cell detections were performed within each region of interest using hematoxylin staining to detect the nucleus of individual cells combined with cell expansion option to segment the whole cell. Finally, the mean intensity of DAB staining of cells from infected lung sections stained with biotinylated secondary antibody only was used to define DAB intensity thresholds for each marker and applied to quantify the percentage of positive cells within each lesion.

Cytokine quantification in organ homogenates

The lungs and spleens of infected mice were isolated and homogenised as described above. Samples were centrifuged at 1800 rpm for 10 minutes at 4°C to pellet the remaining tissue, and supernatants were filtered twice using 0.2 µm Millex-GP polyethersulfone filters (Merck Millipore). The cytokine concentrations in the lung and spleen supernatants were quantified using standardised ELISA kits according to the manufacturer's instructions; IL-10 (R&D), TNF (Invitrogen) and IFN-γ (BD).

Bone-marrow-derived macrophage studies

Bone-marrow derived macrophages were obtained from the femur, tibia, and hips from bones from BALB/c mice. Bone marrow cells were cultured at 37°C and 5% CO_2 for 6 days in BMM medium to obtain BMDMs. BMM medium consisted of high glucose DMEM medium with no glutamine or phenol

red (ThermoFisher) supplemented with 10% heat-inactivated fetal calf serum (Sigma), 20% M-CSF (Generous gift from Dr Etienne Meunier, Toulouse), 1% Hepes (Gibco), 200 mM Glutamine (ThermoFisher), 100 mM Sodium pyruvate (Gibco) and 0.5% Penicillin/Streptomycin (Sigma), which was filtered before use. After 6 or 7 days of culture, BMDMs were washed, counted and plated in flat-bottomed 24 well cell culture plates at 2.5 x 10^5 cells per well in BMM containing no antibiotics. BMDMs were then either left unstimulated, or activated by treatment for 16 hours with IFN-γ (100 U/ml, Biolegend) + LPS (100ng/ml, Sigma) to induce the production of NO, or activated for 16h with IFN-γ (100 U/ml) + LPS (100ng/ml) and treated L-NAME (4mM, Sigma) to inhibit iNOS. In these conditions, we detected no NO in the supernatant of unstimulated cells, between 10 and 20 µM NO in the supernatant of activated cells, and less than 5 µM NO in activated cells treated with L-NAME, using the Griess assay (Sigma). Following stimulation, BMDMs were washed and then incubated with STB-K, KC9A1, KC9C1 or *MTB* H37Rv at MOI 0.5 bacteria per cell for 2 hours before washing again. To evaluate bacterial counts just after infection on day 0 or 3 days after infection, BMDMs were lysed using 0.1% Triton X-100 in H_2O and cell lysates were plated on 7H11 agar plates.

In vitro sensitivity testing

Before performing susceptibility testing to NO, 3-4 week old bacterial cultures were re-cultured for 10 days in 7H9/ADC/Tween to ensure bacteria were in the exponential phase. 10 day old cultures were washed twice in PBS/0.05% Tween 80 and diluted into 7H9/ADC/Tween at a calculated final OD_{600nm} of 0.0005, corresponding to 10^5 CFU/ml, and incubated in the presence of 5 mM DETA/NO (Sigma). The number of viable bacteria in treated and non-treated control cultures was determined by CFU plating at the times indicated in the figure legends.

To evaluate bacterial growth at neutral pH, 3-4 week old cultures were diluted 1 in 100 into medium composed of 7H9/ADC/Tween at pH 6.8. For the acidic pH growth assay, 3-4 week old cultures were diluted 1 in 100 into 7H9 supplemented with ADC and 0.05% Tyloxapol (Merck), which had been

adjusted to pH 5, 5.5, 6 or 6.6. 100 mM MOPS buffer (Sigma) was used to keep the pH stable in medium adjusted to pH6.6, and 100 mM MES buffer (Sigma) was used for pH5, 5.5 and 6. Tyloxapol was used as the dispersal agent for this experiment as Tween 80 was previously shown to be toxic to *MTB* at acidic pH[19]. Growth was monitored by measuring OD_{600nm} at various intervals. Results are represented as the $\log_{10} OD_{600nm}$ and are representative of three independent experiments.

To evaluate bacterial growth in phosphate-limiting conditions, strains were grown for 3-4 weeks in 7H9/ADC/Tween, pelleted, washed three times in Sauton's medium minus phosphate and then diluted to approximately OD_{600nm} 0.05 in Sauton's medium containing the standard concentration of phosphate (2870 µM) or low-phosphate concentrations (0 µM, 50 µM). The composition of Sauton's medium minus phosphate was 4g/L L-asparagine, 2g/L citric acid monohydrate, 0.05g/L ferric ammonium citrate, 1% zinc sulfate, 0.5g/L magnesium sulfate heptahydrate and 60ml/L glycerol in H_2O. Phosphate was added as required in the form of potassium phosphate dibasic trihydrate. Strain growth was evaluated by monitoring the OD_{600nm}. Data are represented as the mean and standard deviation from three independent experiments.

To evaluate bacterial tolerance to zinc and copper, 3-4 week old bacterial cultures were re-cultured for 7 days in 7H9/ADC/Tween to ensure bacteria were in the exponential phase. 7 day old cultures were diluted into 4 ml 7H9/ADC/Tween at a calculated final OD_{600nm} of 0.005 in McFarland tubes, and incubated in the presence of various concentrations of zinc ($ZnSO_4$, Sigma), or copper ($CuCl_2$, Sigma). At the indicated time points, each tube was vortexed before evaluation of turbidity using a McFarland densitometer. Each measure was performed in triplicate and the experiment was performed four times independently. Data are represented as the mean and standard deviation of the triplicates from one experiment.

Permeability assay

To assess membrane permeability of the wild-type *M. smegmatis* (WT), PMM280 (smegΔrv1339), and the complemented strains PMM280::pAH01H (smeg::rv1339A1) and PMM305::pAH02H (smeg::rv1339K), an EtBr uptake assay on intact mycobacteria was performed, as previously described[63]. Briefly, logarithmic-phase strains grown in 7H9/ADC/Tween were centrifuged, normalised to OD_{600nm} of 0.8 in PBS buffer containing glucose 0.4%. 100 µl of bacterial suspension was added per well into 96-well microplates. 10 µl EtBr (25 µM) was added to the bacterial suspension and the mix was incubated at 37°C. Fluorescence, associated with accumulation of EtBr, was measured for 1 hour at 3 minute intervals with an excitation wavelength of 525 nm and an emission wavelength of 615 nm. The experiment was performed twice in quintuplicate for the WT and smegΔrv1339 strains and once in quintuplicate for the smeg::rv1339A1 and smeg::rv1339K strains.

Statistical analysis

Statistical analysis was carried out using GraphPad Prism 5 and 7. One or two-way analysis of variance (ANOVA) followed by the Bonferroni post hoc test was used to analyse the statistical significance between three or more groups. P values less than 0.05 were considered to be statistically significant. When bacteria were undetectable, the CFU value was set at the limit of detection (log_{10} 1.7 for *in vivo* studies and log_{10} 1.3 for *in vitro* studies , with the exception of the study involving C57BL/6 and $NOS2^{-/-}$ mice where the limit was log_{10} 2) for the purpose of statistical analysis. For comparison of cell types in the lesions, the Brown-Forsythe and Welch ANOVA test with the Games-Howell post hoc test was used.

Ethics statement

All studies with *M. canettii* and *M. tuberculosis* strains were performed in biosafety level 3 (BSL3) conditions, strictly following all relevant biosafety guidelines for experimental work with wild-type and

577 genetically modified tuberculosis bacilli (Authorization N°6406 from the French Ministry for Higher
578 Education and Research).
579 Animal studies were performed in agreement with European and French guidelines (EC Directive
580 2010/63/UE and French Law 2013-118 issued on February 1, 2013) after approval by the relevant
581 Ethics Committee (Comité d'éthique en experimentation animale N°01 and N°89) and the French
582 Ministry for Higher Education and Research after ethical evaluation (number 201508271122464 v2
583 (APAFIS#1535) and number 2018060717283847 v1 (APAFIS#15409)).

584

585 **Data availability**

586 The sequence data were deposited in the NCBI repository under project PRJNA662472, PRJNA664307
587 and PRJNA665366 accession codes for Illumina-derived genome sequences of STB-K- and STB-D-
588 derived clones and populations and PacBio-derived genome sequence of STB-K respectively. All other
589 data that support the findings of this study are available from the corresponding author upon request.

590

591

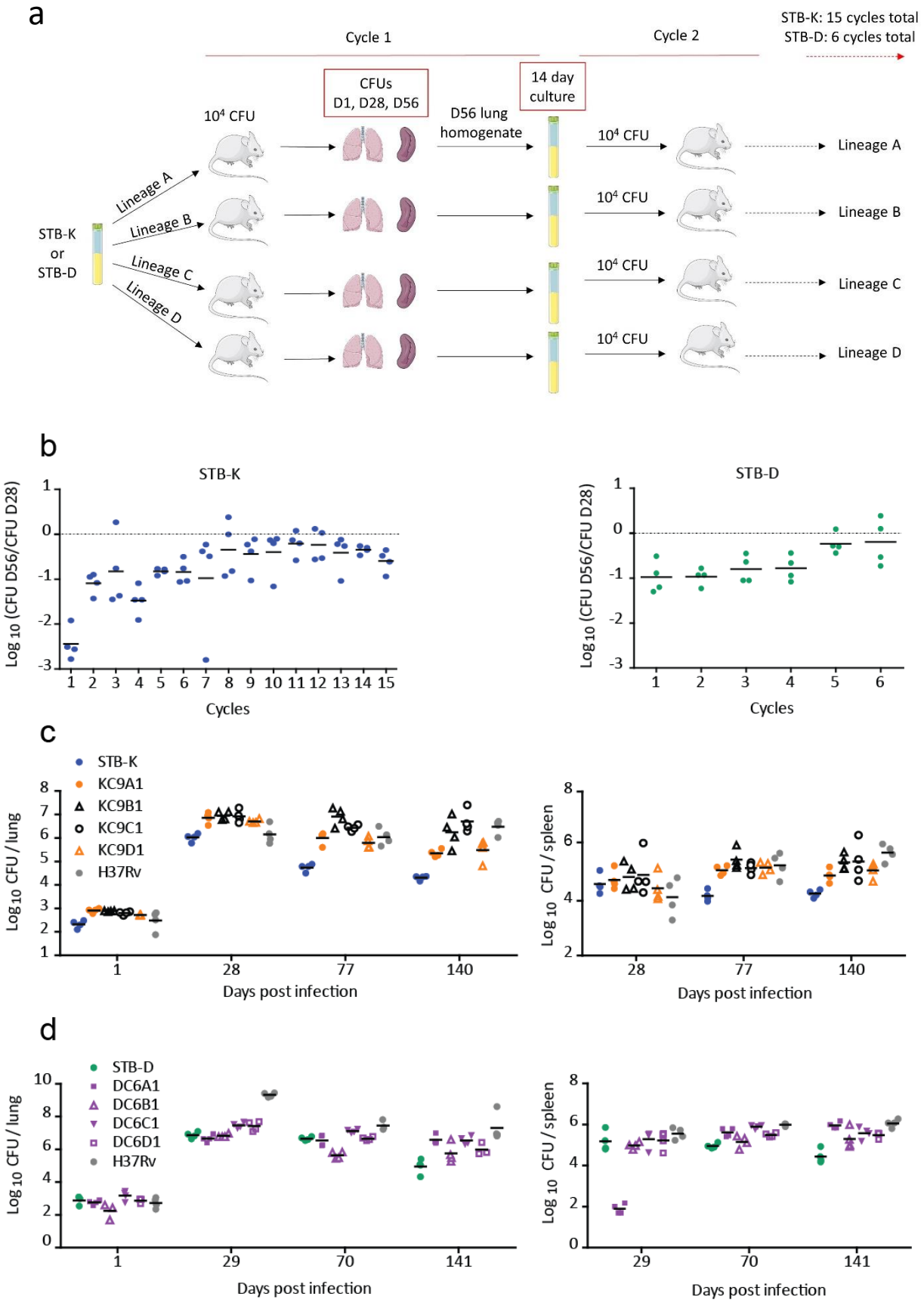

a
Cycle 1
Cycle 2
STB-K: 15 cycles total
STB-D: 6 cycles total
CFUs
D1, D28, D56
14 day
culture
D56 lung
homogenate
10⁴ CFU
STB-K
or
STB-D
Lineage A
Lineage B
Lineage C
Lineage D
b
STB-K
STB-D
Log 10 (CFU D56/CFU D28)
Cycles
c
STB-K
KC9A1
KC9B1
KC9C1
KC9D1
H37Rv
Log 10 CFU / lung
Log 10 CFU / spleen
Days post infection
d
STB-D
DC6A1
DC6B1
DC6C1
DC6D1
H37Rv
Log 10 CFU / lung
Log 10 CFU / spleen
Days post infection

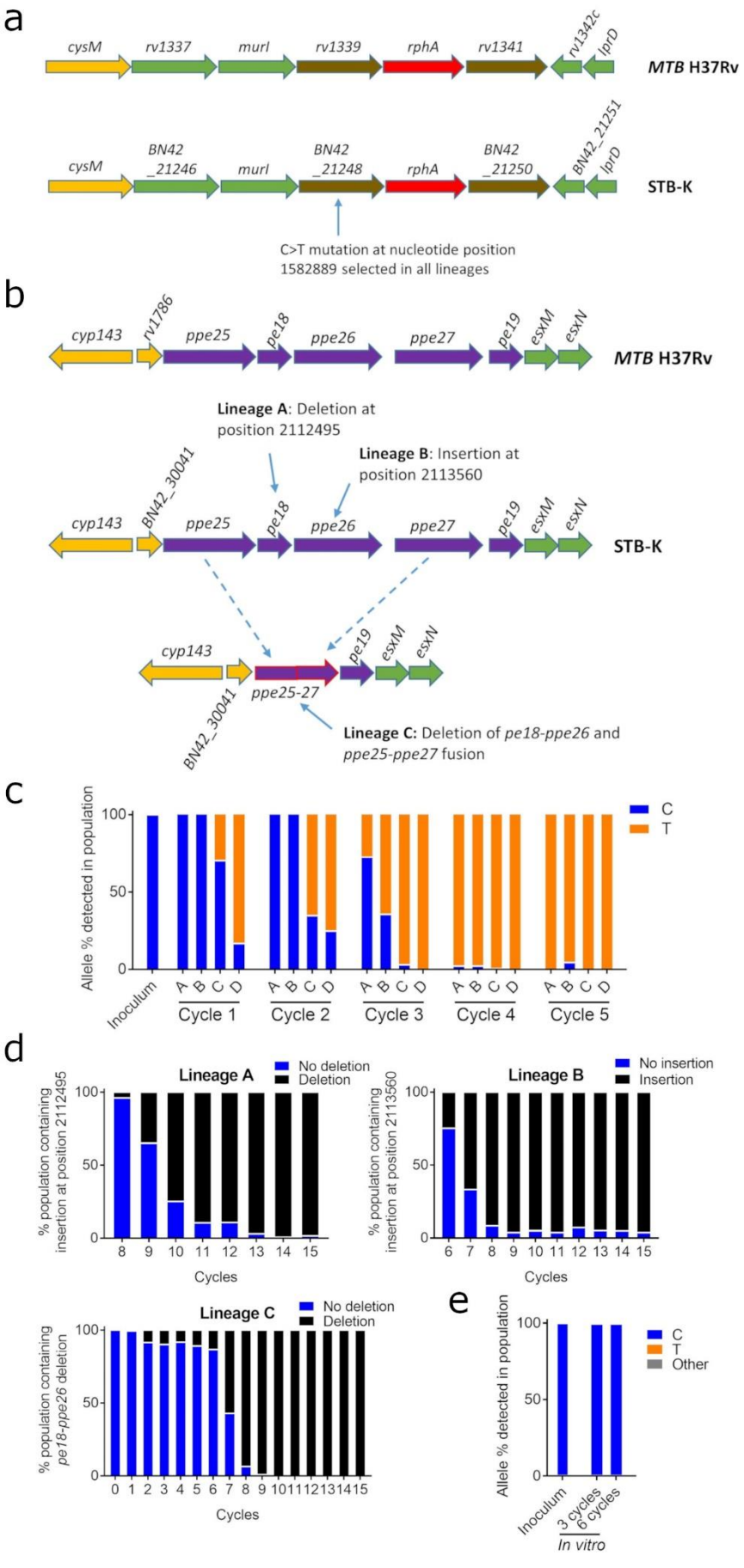
Figure 1: Experimental evolution in mice selects *M. canettii* clones with enhanced persistence.

a) Two experimental evolutions were performed in BALB/c mice using *M. canettii* STB-K and STB-D. For both strains, 4 independent cultures were passaged corresponding to evolution lineages A, B, C and D, and were never mixed. At the beginning of each cycle, 12 mice were infected i.n. with 10^4 CFU. CFU numbers in the lungs and spleen were evaluated at 1, 28 and 56 days post-infection (4 mice per time point) by plating. Part of the lung homogenates recovered from the mice sacrificed at 56 days post-infection was also used to initiate 4 separate cultures. These cultures were grown for 14 days at 37°C and were then used as the new inocula for the next cycle. 15 cycles of experimental evolution were performed for STB-K and 6 cycles were performed for STB-D. This diagram was created using images from Servier Medical Art.

b) The persistence index was evaluated for each cycle in the experimental evolution by calculating the log ratio of bacterial burden in the lung at day 56 *vs* day 28. Data are shown as $\log_{10}$ CFU D56/D28 for each lineage and the mean.

c) Persistence characteristics of 4 individual clones recovered after cycle 9 from STB-K evolution lineages A, B, C and D. Mice were i.n. infected with 10^3 CFU of the parental STB-K strain, mutants KC9A1, KC9B1, KC9C1, KC9D1, or *MTB* H37Rv and the bacterial burden in the lungs and spleen was assessed at days 1, 28, 77 and 140 post-infection. Data are shown as $\log_{10}$ CFU for individual mice and the mean from an experiment performed with 3/4 mice per group.

d) Persistence phenotypes of 4 individual clones recovered after cycle 6 from STB-D evolution lineages A, B, C and D. Mice were i.n. infected with 10^3 CFU of the parental STB-D strain, mutants DC6A1, DC6B1, DC6C1, DC6D1 or *MTB* H37Rv and the CFU numbers in the lungs and spleen were assessed at days 1, 29, 70 and 141 post-infection. Data are shown as $\log_{10}$ CFU for individual mice and the mean from an experiment performed with 3/4 mice per group. When bacteria were undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.7).


a
cysM
rv1337
murI
rv1339
rphA
rv1341
rv1342c
lprD
MTB H37Rv
cysM
BN42_21246
murI
BN42_21248
rphA
BN42_21250
BN42_21251
lprD
STB-K
C>T mutation at nucleotide position 1582889 selected in all lineages
b
cyp143
rv1786
ppe25
pe18
ppe26
ppe27
pe19
esxM
esxN
MTB H37Rv
Lineage A: Deletion at position 2112495
Lineage B: Insertion at position 2113560
cyp143
BN42_30041
ppe25
pe18
ppe26
ppe27
pe19
esxM
esxN
STB-K
cyp143
pe19
esxM
esxN
ppe25-27
BN42_30041
Lineage C: Deletion of pe18-ppe26 and ppe25-ppe27 fusion
c
Allele % detected in population
100
50
0
C
T
Inoculum
A B C D
Cycle 1
Cycle 2
Cycle 3
Cycle 4
Cycle 5
d
Lineage A
No deletion
Deletion
% population containing insertion at position 2112495
Cycles
Lineage B
No insertion
Insertion
% population containing insertion at position 2113560
Cycles
Lineage C
No deletion
Deletion
% population containing pe18-ppe26 deletion
Cycles
e
Allele % detected in population
C
T
Other
Inoculum
3 cycles
6 cycles
In vitro

616

Figure 2: Mutations in the *rv1339* orthologue and the *ppe25-ppe27* region of STB-K mutants were identified by whole-genome sequencing.

a) Genetic region containing the *rv1339* gene in *MTB* H37Rv and the orthologue in *M. canettii* STB-K. The blue arrow indicates the location of the point mutation identified in STB-K mutants following experimental evolution.

b) The *ppe25-ppe27* locus of *MTB* H37Rv and STB-K. The blue arrows indicate the positions of the mutations identified in the various STB-K evolution lineages between cycles 6 and 9 of the experimental evolution.

c) Fixation of the C>T mutation in the *rv1339* orthologue occurred early during the experimental evolution.

d) The mutations in the *ppe25-ppe27* region of the STB-K mutants became fixed in evolution lineages A, B and C after 9-13 cycles of experimental evolution.

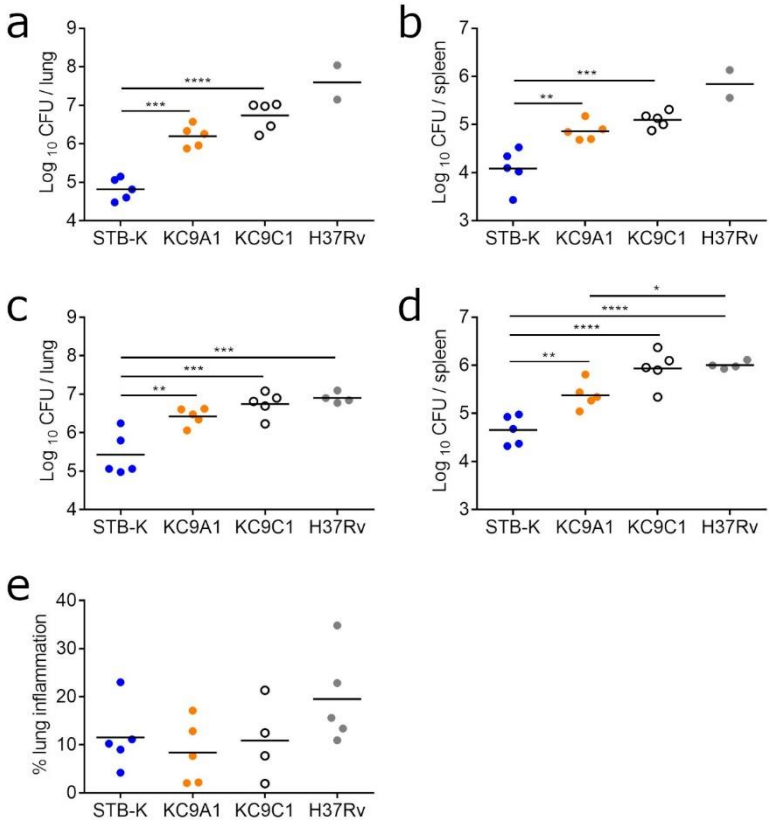
e) WGS demonstrating that the C>T mutation was not enriched following serial passaging of STB-K *in vitro*.

For **c**), and **d**) whole -genome sequence data were generated from the bacterial populations corresponding to the inoculum and to each evolution lineage recovered after cycles 1 to 15. Positions corresponding to the mutation in the *rv1339* orthologue or the deletions in the *ppe25-ppe27* regions were analysed to determine the distribution of wild-type and mutant alleles in the population.

For **e**), we performed 6 cycles of growth of STB-K in 7H9/ADC/Tween supplemented with panta (as for the *in vitro* cultures between each mice infection cycle). Whole -genome sequence data were generated from the bacterial populations corresponding to the inoculum and bacterial populations recovered after cycles 3 and 6. For the position corresponding to the C>T in the *rv1339* orthologue, the coverage depth was x98, x124 and x159 for the inoculum, cycle 3 and cycle 6 populations respectively.

641 For **c**), **d**), and **e**), data are shown as the percentages of the wild-type and mutated alleles detected in

642 the STB-K-derived populations from each evolution lineage after each cycle by WGS.



643

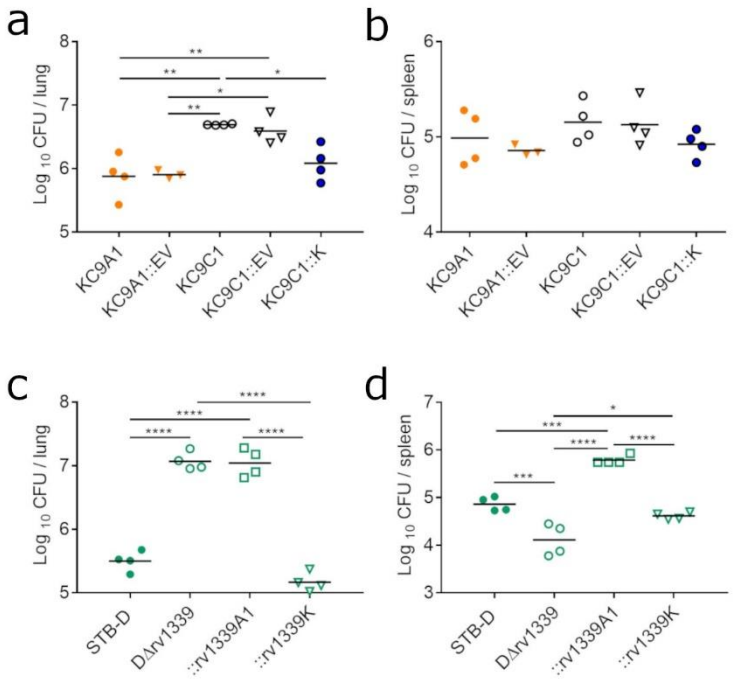
644 **Figure 3: The persistence phenotypes of *rv1339* and *pe18-ppe26* STB-K mutants are not specific to**
645 **the mouse strain or route of infection.**

646 C3HeB/FeJ mice were i.n. infected with approximately 200 CFU of STB-K, KC9A1, KC9C1 or *MTB* H37Rv
647 and bacterial burden in the **a**) lung and **b**) spleen was quantified at 77 days post-infection by CFU
648 plating on solid medium. Results are shown as the $\log_{10}$ CFU for individual mice and the mean (5 mice
649 per group except for H37Rv because 3 mice had to be euthanised before the designated end point
650 due to health deterioration), and are representative of two experiments. **P < 0.01, ***P < 0.001,
651 ****P < 0.0001 by one-way ANOVA with the Bonferroni post hoc test.

652 C3HeB/FeJ mice were i.v. infected with approximately 10^3 CFU of STB-K, KC9A1, KC9C1 or H37Rv and
653 the CFU numbers in the **c**) lung and **d**) spleen were evaluated at 77 days post-infection. Results are
654 shown as the $\log_{10}$ CFU for individual mice (4/5 per group) and the mean, and are representative of

two independent experiments. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by one-way ANOVA with the Bonferroni post hoc test.

e) C3HeB/FeJ mice were i.n. infected with approximately 200 CFU of STB-K, KC9A1, KC9C1 or *MTB* H37Rv. After 76 days of infection, lungs were isolated for histopathology. Lung samples were fixed and stained with hematoxylin and eosin. Percentage lung inflammation was quantified by dividing the total inflamed area by the total lung area, and multiplying by 100. Data are represented as the percentage lung inflammation for individual mice (one entire lung section per mouse) and the mean from an experiment performed with 4/5 mice per group. Non-significant by one-way ANOVA with the Bonferroni post hoc test.

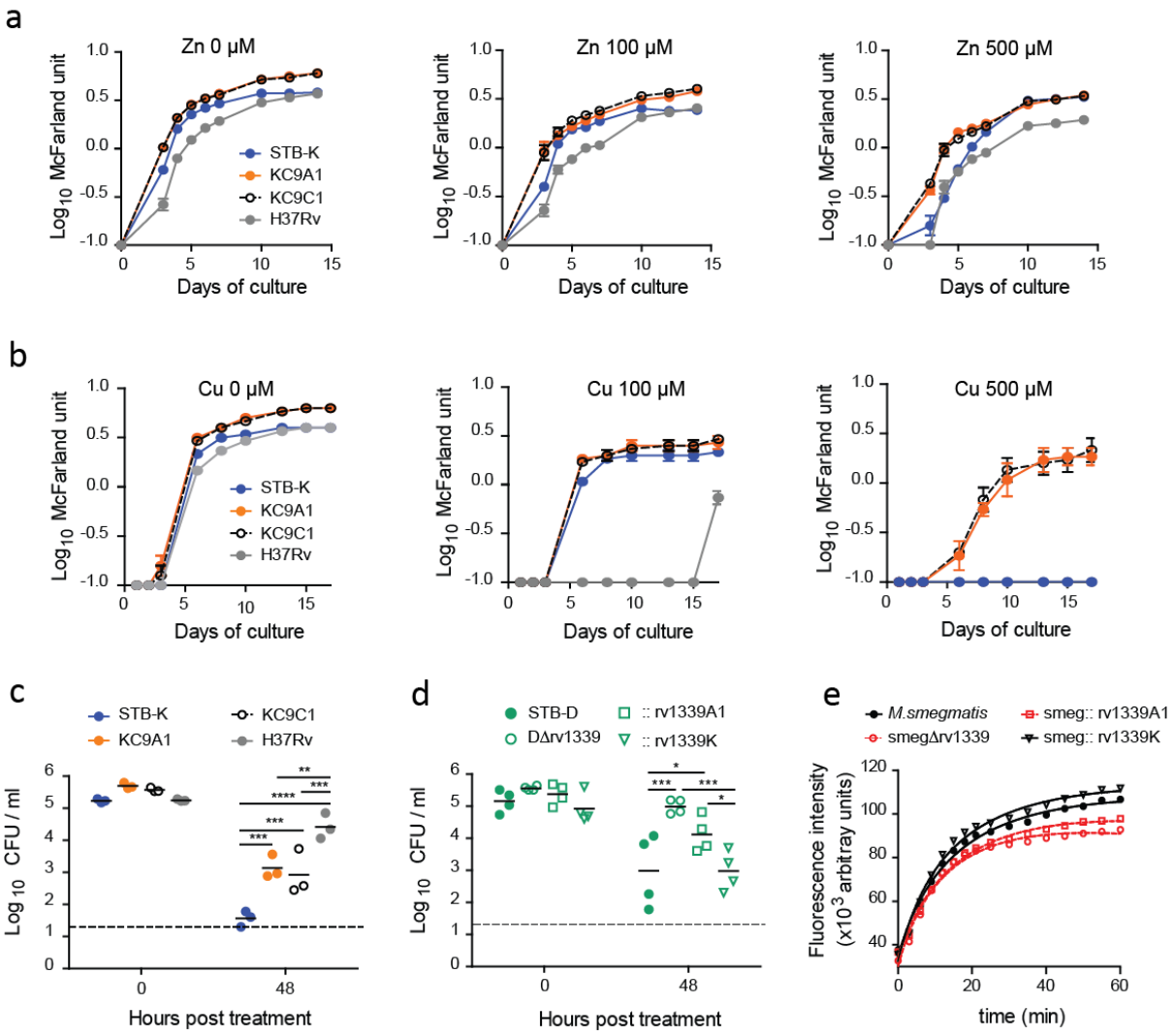


664

665 **Figure 4: The *pe18-ppe26* and *rv1339* mutations confer enhanced persistence.**

666 BALB/c mice were i.n. infected with approximately 10^3 CFU of *M. canettii* KC9A1, KC9A1::EV, KC9C1,
667 KC9C1::EV, or KC9C1::K, and CFU counts in the **a**) lung and **b**) spleen were quantified at 70 days post-
668 infection by plating serial dilutions of the organ homogenate on 7H11 agar plates. Results are shown
669 as the $\log_{10}$ CFU for individual mice (3/4 per group) and the mean, and are representative of two
670 independent experiments. *P < 0.05, **P < 0.01, by one-way ANOVA with the Bonferroni post hoc
671 test.

672 BALB/c mice were i.n. infected with approximately 10^3 CFU of STB-D, DΔrv1339, ::rv1339A1 or
673 ::rv1339K. **c**) Lung and **d**) spleen bacterial burden was evaluated at 77 days post-infection by CFU
674 plating on solid medium. Results are shown as the $\log_{10}$ CFU for individual mice (4 per group) and the
675 mean, and are representative of two independent experiments. *P < 0.05, ***P < 0.001, ****P <
676 0.0001 by one-way ANOVA with the Bonferroni post hoc test.



677

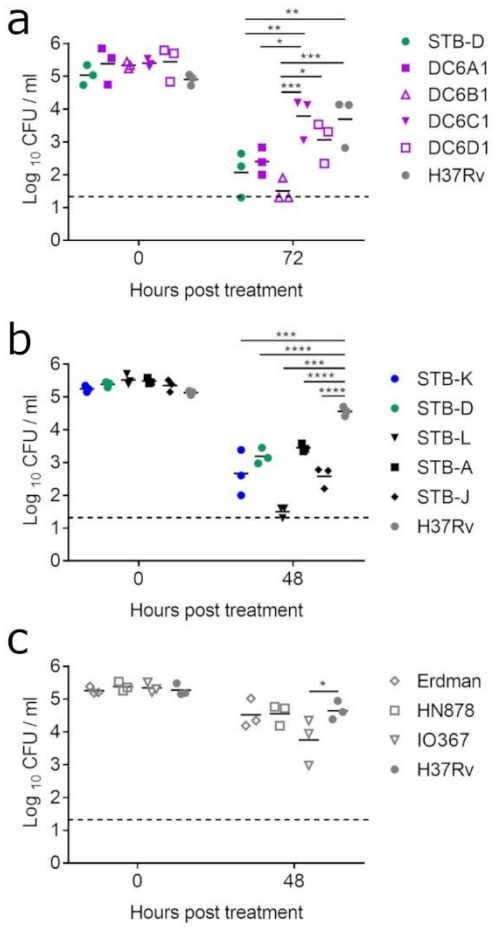
678 **Figure 5: The *rv1339* mutation contributes to enhanced resistance to host-derived stresses.**

679 Tubes of 7H9/ADC/Tween containing various concentration of **a**) $ZnSO_4$, or **b**) $CuCl_2$ were inoculated
680 with STB-K, KC9A1, KC9C1, and *MTB* H37Rv at a final OD_{600nm} of approximately 0.005 and growth was
681 evaluated by monitoring the density (evaluated as McFarland units). Data are represented as the
682 mean and standard deviation from triplicates and are representative of four independent
683 experiments.

684 Flasks of 7H9/ADC/Tween +/- 5 mM DETA/NO were inoculated with 10^5 CFU of **c**) STB-K, KC9A1, KC9C1
685 and H37Rv, or **d**) STB-D, DΔrv1339, ::rv1339A1 and ::rv1339K. Bacterial survival was monitored by
686 evaluating CFU numbers at 0 and 48 hours post treatment with DETA/NO. Results are shown as log_{10}
687 CFU/ml for individual data points from **c**) three and **d**) four independent experiments and the mean.

When bacteria were undetectable, the CFU value was set at the limit of detection (log_{10} 1.3) for the purposes of statistical analysis. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by two-way ANOVA with the Bonferroni post hoc test.

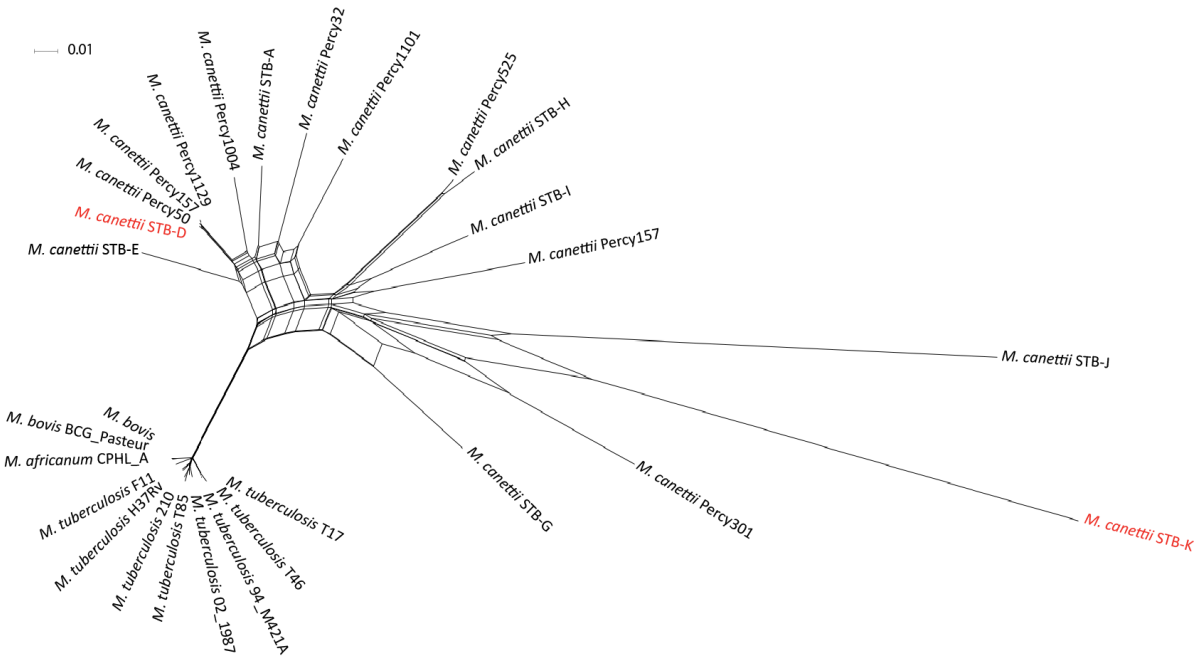
e) Exponentially growing *M. smegmatis*, smegΔrv1339, smeg::rv1339A1 and smeg::rv1339K were incubated with EtBr and intracellular accumulation was followed over 60 minutes by measuring fluorescence intensity. Data are represented as the mean and standard deviation from quintuplicates. The experiment was performed twice in quintuplicate for the WT and smegΔrv1339 strains and once in quintuplicate for the smeg::rv1339A1 and smeg::rv1339K strains.



698

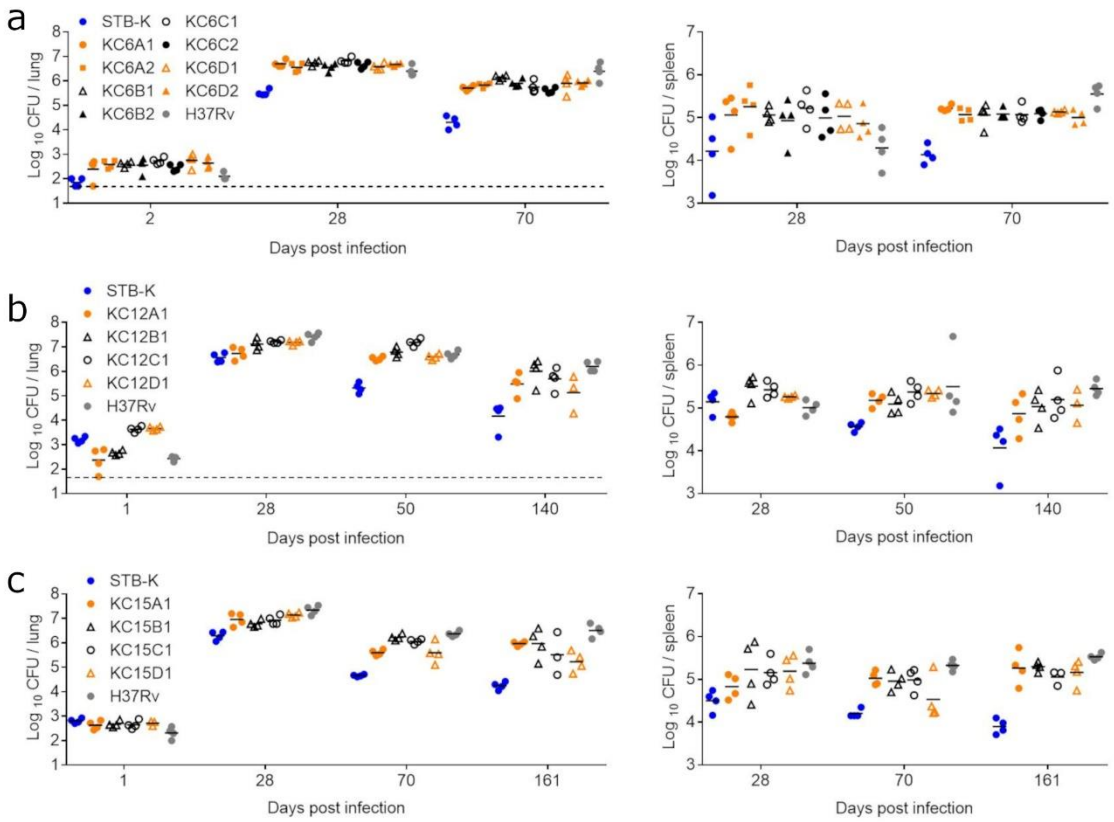
699 **Figure 6: *MTB* exhibit higher nitric oxide resistance than *M. canettii*.**

700 Flasks of 7H9/ADC/Tween +/- 5 mM DETA/NO were inoculated with 10^5 CFU of **a**) STB-D, DC6A1,
701 DC6B1, DC6C1, DC6D1, and *MTB* H37Rv, **b**) *M. canettii* STB-K, -L, -A, -D, -J, and *MTB* H37Rv, or **c**) *MTB*
702 Erdman, HN878, IO367, and H37Rv. Bacterial survival was monitored by evaluating CFU numbers at
703 time 0 and 48 or 72 hours post treatment with DETA/NO. Results are shown as the $\log_{10}$ CFU/ml for
704 individual data points from three independent experiments and the mean. When bacteria were
705 undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.3) for the purposes of statistical
706 analysis. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by two-way ANOVA with the Bonferroni
707 post hoc test.



Supplementary Figure 1: **Phylogenetic tree showing the distribution of *M. canettii* and *MTB* strains.**

Network phylogeny inferred among 28 genomes by NeighborNet analysis, based on pairwise alignments of whole genome SNP data from a selection of strains (17 *M. canettii*, 8 *MTB* from various phylogenetic lineages and 3 strains from the *M. africanum* and *M. bovis* lineages), depicted by the SplitsTree software[64]. The whole SNP dataset was originally generated for Supplementary Figure 1 of reference[10], using genome data described by[4,9,65]. Strains used in experimental evolution experiments are coloured in red.



716

717 **Supplementary Figure 2: Experimental evolution in mice selects *M. canettii* clones with enhanced**

718 **persistence.**

719 Persistence phenotypes of individual STB-K clones recovered from STB-K evolution lineages A-D after

720 cycles **a**) 6, **b**) 12 and **c**) 15. Mice were i.n. infected with 10^3 CFU of **a**) KC6A1, KC6A2, KC6B1, KC6B2,

721 KC6C1, KC6C2, KC6D1, and KC6D2, **b**) KC12A1, KC12B1, KC12C1, and KC12D1, or **c**) KC15A1, KC15B1,

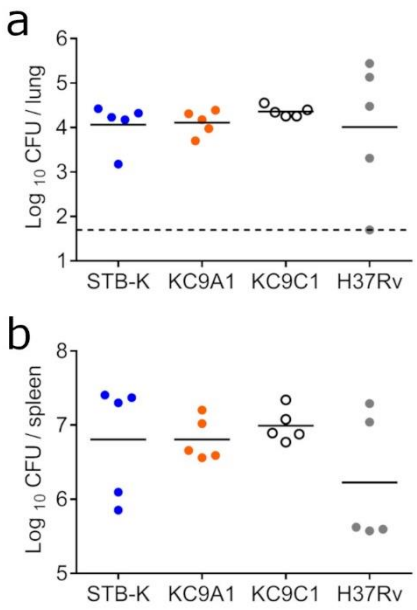
722 KC15C1, and KC15D1. Parental STB-K and control *MTB* H37Rv strains were included in all three

723 experiments. The bacterial burden in the lungs and spleen was assessed at various intervals post-

724 infection by plating serial dilutions of organ homogenates on 7H11 agar plates. Data are shown as

725 $\log_{10}$ CFU for individual mice and the mean, from an experiment performed with 3/4 mice per group.

726 When bacteria were undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.7).



727

728 **Supplementary Figure 3: KC9A1 and KC9C1 display similar bacterial burden to STB-K during the acute**

729 **phase of infection.**

730 C3HeB/FeJ mice were i.v. infected with approximately 10^3 CFU of STB-K, KC9A1, KC9C1, or *MTB* H37Rv,

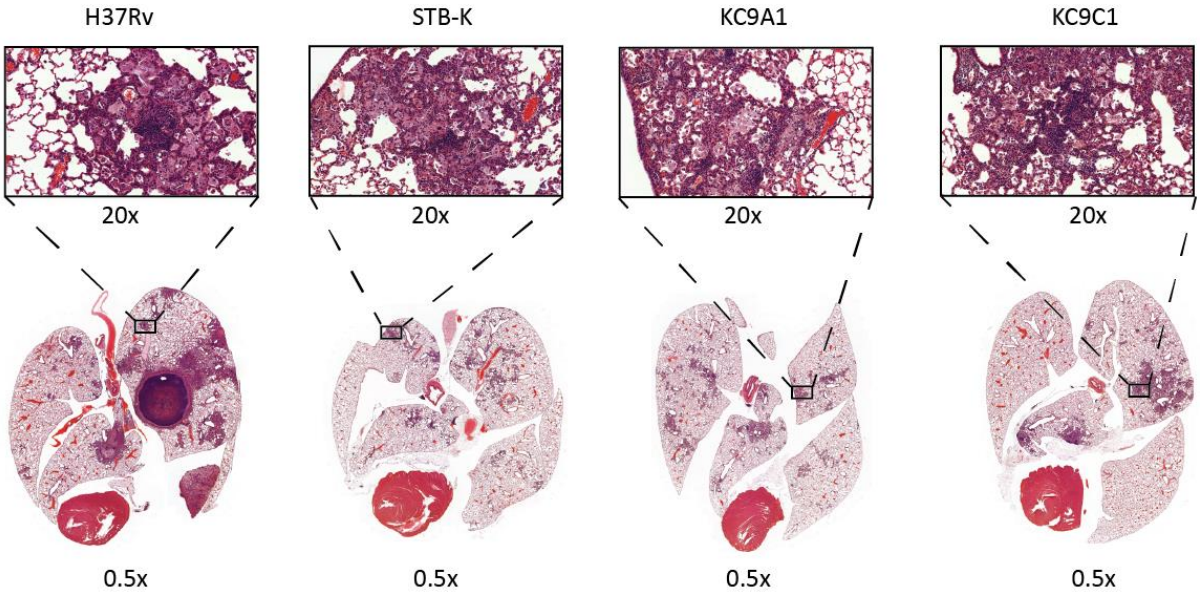
731 and CFU counts in the **a**) lung and **b**) spleen were quantified at 21 days post-infection by plating

732 dilutions of the organ homogenate on 7H11 agar plates. Results are shown as the $\log_{10}$ CFU for

733 individual mice (5 per group) and the mean, and are representative of two independent experiments.

734 Non-significant by one-way ANOVA with the Bonferroni post hoc test. When bacteria were

735 undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.7).



736

737 **Supplementary Figure 4: KC9A1 and KC9C1 do not induce a greater inflammatory response in the**

738 **lungs of mice than STB-K.**

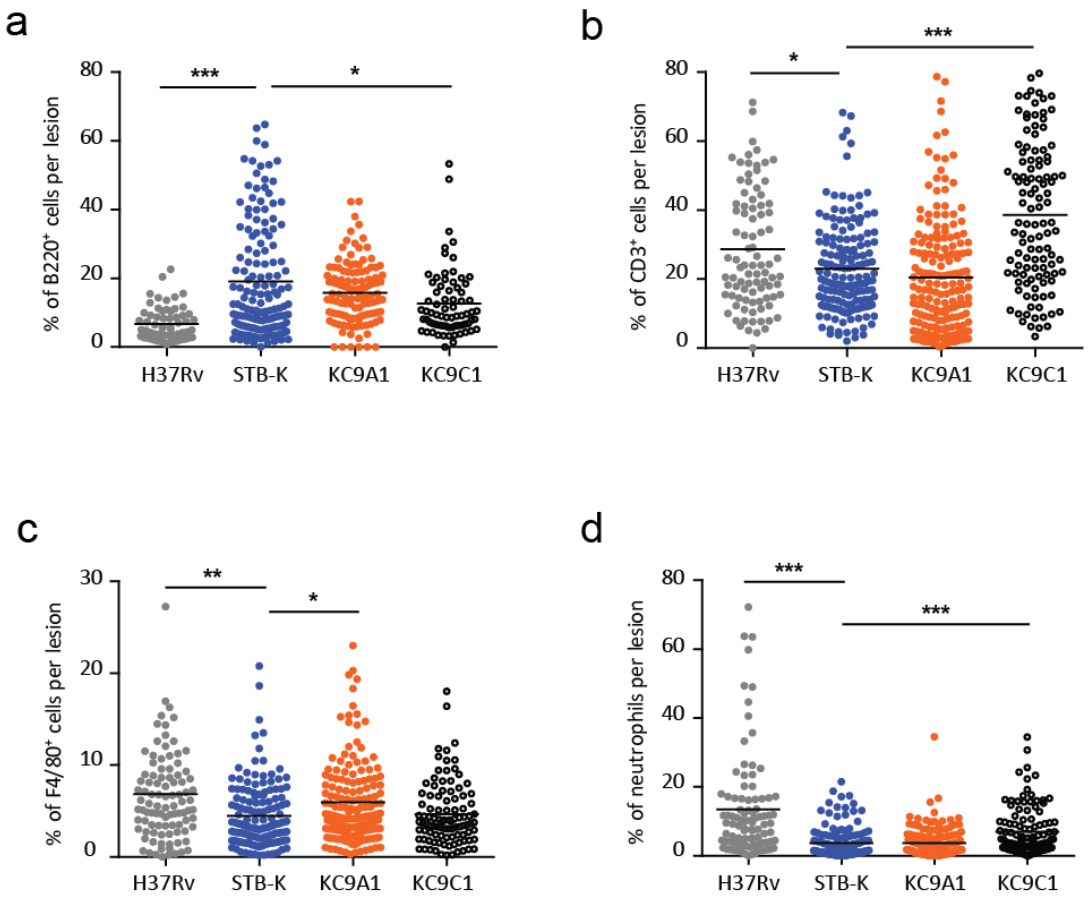
739 Lungs were isolated from C3HeB/FeJ mice 76 days after i.n. infection with approximately 200 CFU of

740 STB-K, KC9A1, KC9C1, or H37Rv for histopathology. Lung samples were fixed and stained with

741 hematoxylin and eosin, and analysed using Pannoramic Viewer. For each mouse (4/5 mice per group),

742 one entire lung section was analysed and the images shown are representative.

743



744

745 **Supplementary Figure 5: STB-K and KC9A1 induce a similar recruitment of immune cells to the lungs**

746 **whereas the pattern for KC9C1 is closer to that of H37Rv.**

747 Lung samples from C3HeB/FeJ mice 76 days after i.n. infection with approximately 200 CFU of STB-K,

748 KC9A1, KC9C1, or H37Rv were fixed, and analysed by immunohistochemistry. Serial lung sections, one

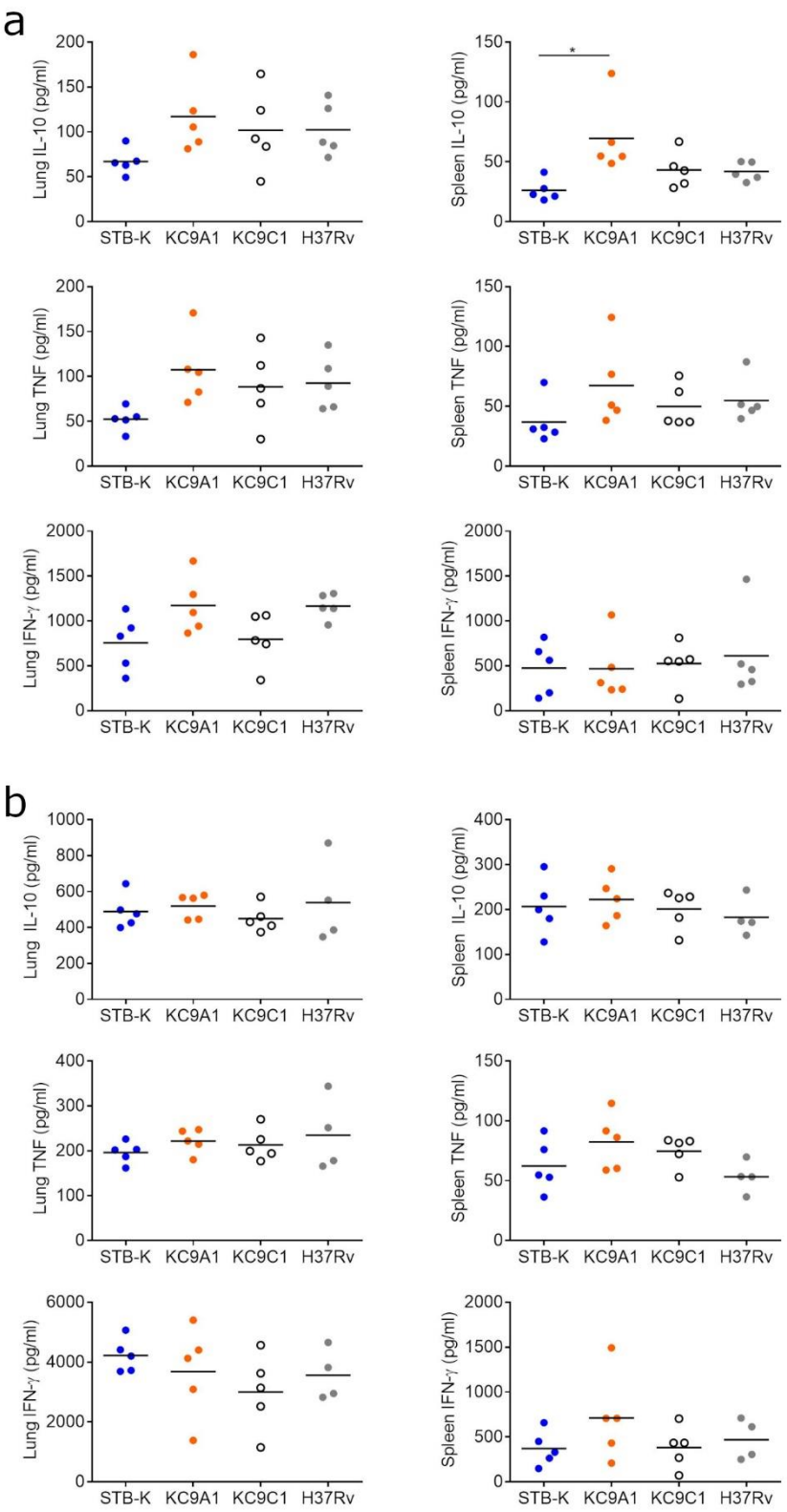
749 per infected mouse and per primary antibody, were labelled with antibodies against **a**) B lymphocytes

750 (B220), **b**) T lymphocytes (CD3), **c**) macrophages (F4/80), and **d**) neutrophils (neutrophil marker). At

751 least four sections from 4 mice were analysed with each antibody. The percentage of each cell type in

752 the various lesions was quantified. *P < 0.05, **P < 0.01, ***P < 0.001, by Brown-Forsythe and Welch

753 ANOVA test with the Games-Howell post hoc test.

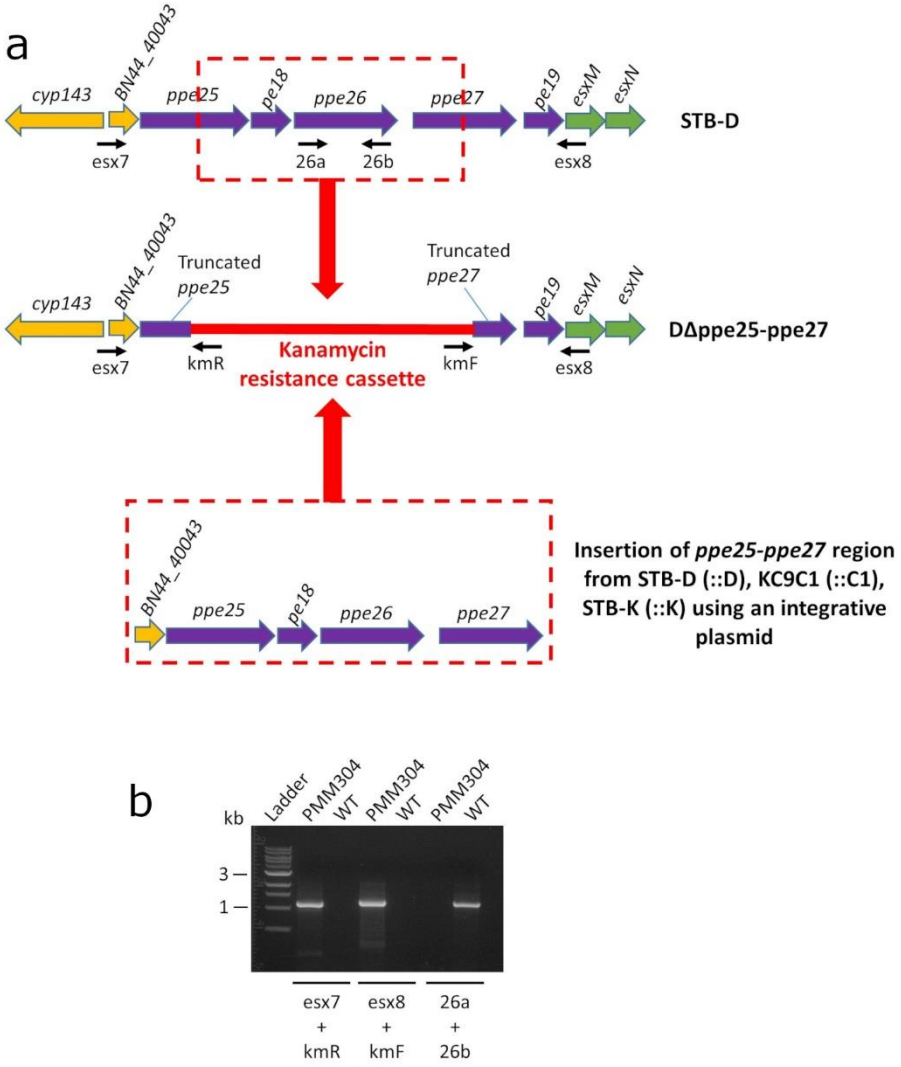


754

755 **Supplementary Figure 6: The persistent phenotypes of KC9A1 and KC9C1 are not associated with a**

756 **change in cytokine production.**

C3HeB/FeJ mice were i.v. infected with approximately 10^3 CFU of STB-K, KC9A1, KC9C1, or *MTB* H37Rv, and IL-10, IFN-γ and TNF concentrations in the lung and spleen homogenates were quantified at **a**) 21 days and **b**) 76 days post-infection by ELISA. Mean cytokine concentrations are shown, in addition to the concentrations for individual mice (4/5 per group). *P < 0.05 by one-way ANOVA with the Bonferroni post hoc test.

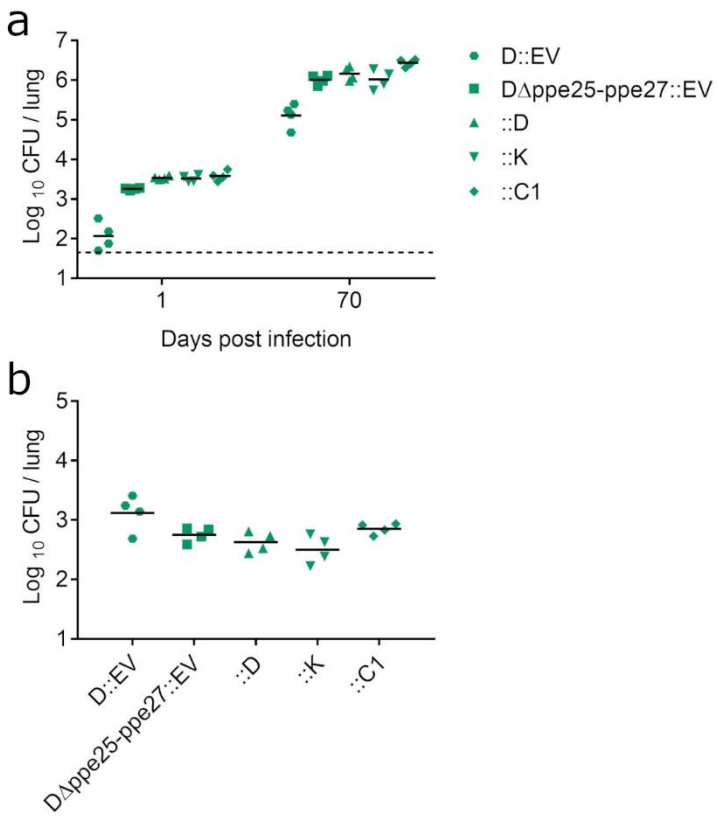


762

763 **Supplementary Figure 7: Insertion of the *ppe25-ppe27* regions from KC9C1 and STB-K into STB-D.**

764 **a**) The *ppe25-ppe27* region of STB-D was replaced with a *km* resistance cassette by allelic exchange to
765 create DΔppe25-ppe27 (PMM304). This knockout strain was then complemented with an integrative
766 plasmid containing the *ppe25-ppe27* regions from STB-D (::D), KC9C1 (::C1), or STB-K (::K). Only the
767 region from STB-D is represented in this figure. The black arrows indicate the primers used to amplify
768 the corresponding regions.

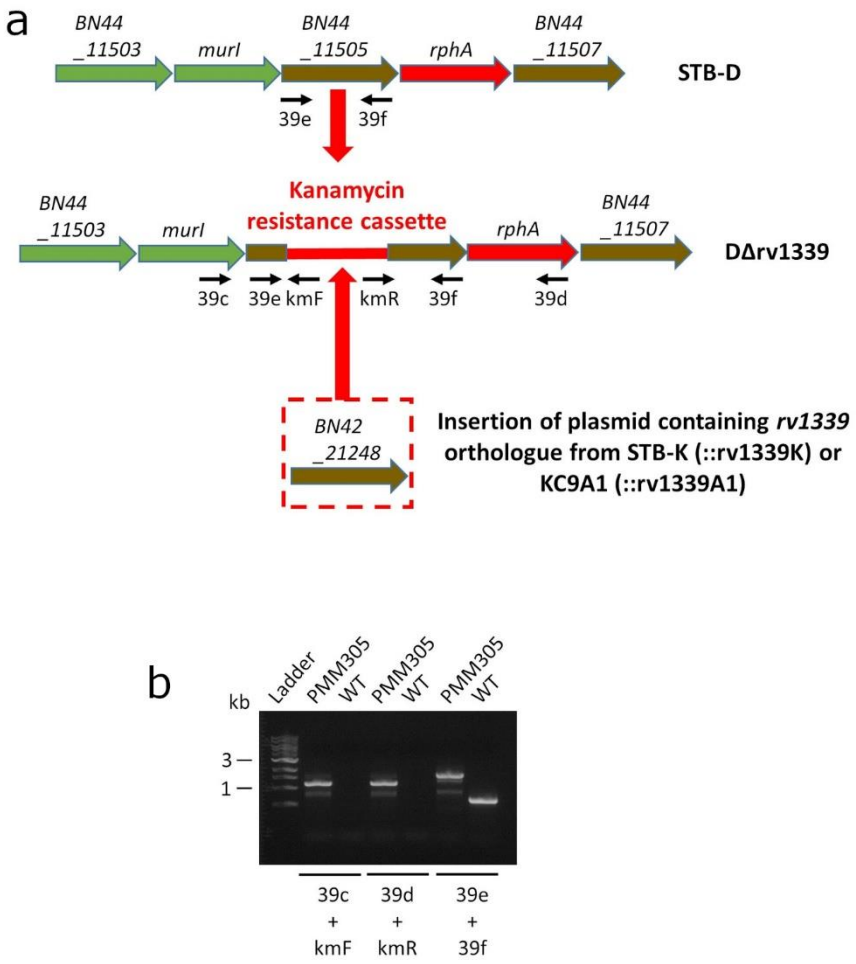
769 **b**) PCR profiles corresponding to a mutant deleted for *pe18-ppe26* (PMM304), or the wild-type STB-D
770 strain (WT), when the indicated primer pairs were used for amplification.



771

772 **Supplementary Figure 8: The *pe18-ppe26* deletion does not confer enhanced persistence to STB-D.**

773 BALB/c mice were i.n. infected with approximately 10^3 CFU of *M. canettii* D::EV, DΔppe25-ppe27::EV,
774 ::D, ::K, or ::C1, and CFU counts in the lung were quantified at 1 and 70 days post-infection by plating
775 dilutions of the organ homogenate on 7H11 agar plates. Results are shown as **a**) the $\log_{10}$ CFU per lung
776 for individual mice and the mean, and **b**) normalised values ($\log_{10}$ CFU D77 – $\log_{10}$ CFU D1), from an
777 experiment performed with 4 mice per group. Differences between DΔppe25-ppe27::EV, ::D, ::K and
778 ::C1 were deemed non-significant by **a**) two-way and **b**) one-way ANOVA with the Bonferroni post hoc
779 test. The statistical difference between D::EV and the other groups is not shown. When bacteria were
780 undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.7).

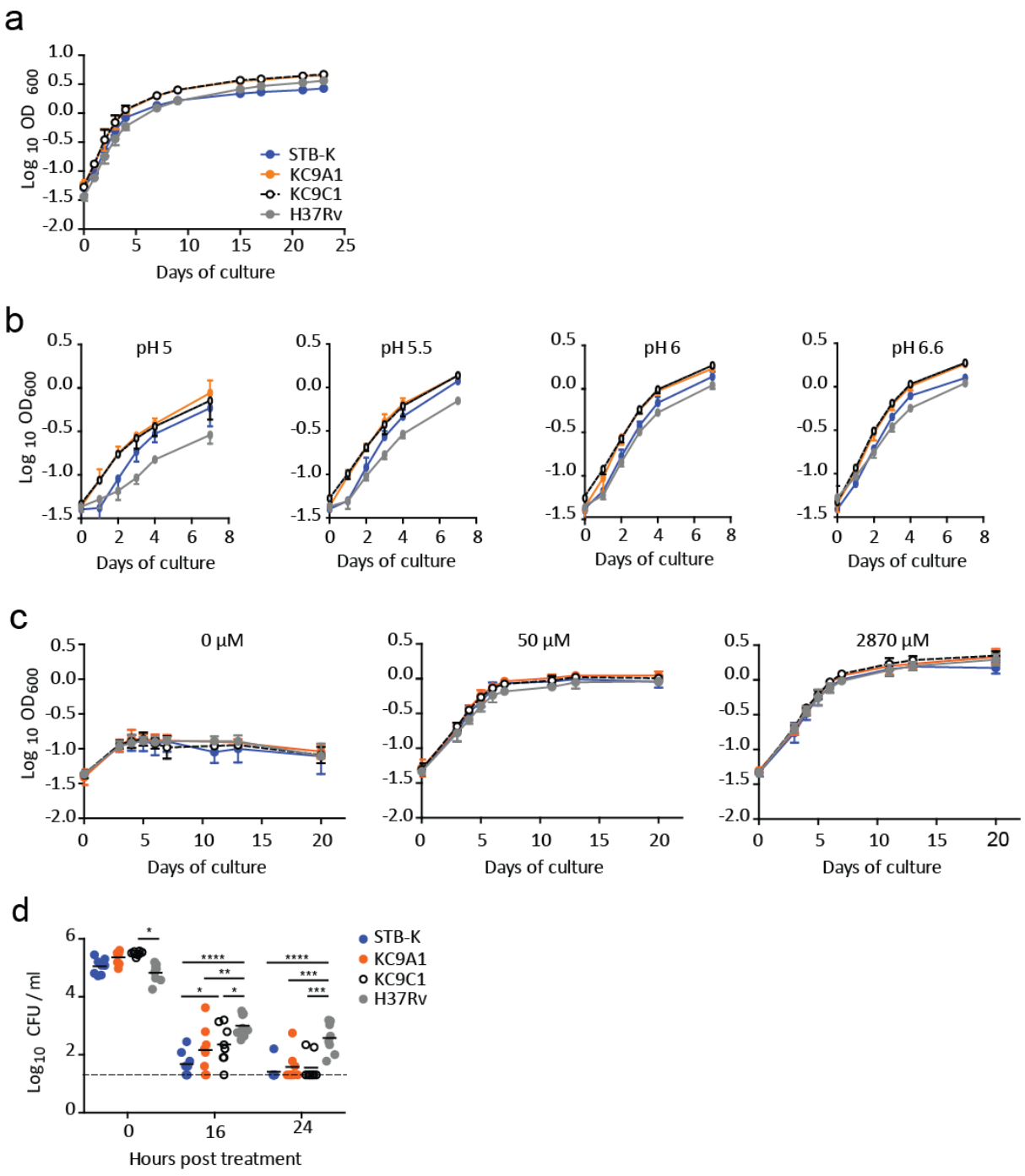


781

782 **Supplementary Figure 9: Replacement of the *rv1339* orthologue in STB-D with the *rv1339***
783 **orthologue of STB-K or KC9A1.**

784 **a**) A *km* resistance cassette was inserted in the *rv1339* orthologue of STB-D (*BN44_11505*) by allelic
785 exchange to create DΔrv1339 (PMM305). DΔrv1339 was then complemented with plasmids
786 containing either the *rv1339* orthologue from STB-K (::rv1339K) or KC9A1 (::rv1339A1). Only the
787 *rv1339* orthologue from STB-K is represented in this figure. The black arrows indicate the primers used
788 to amplify the corresponding regions.

789 **b**) PCR profiles corresponding to a mutant interrupted for the *rv1339* orthologue (PMM305), or the
790 wild-type STB-D strain (WT), when the indicated primer pairs were used for amplification.



791

792 **Supplementary Figure 10: STB-K and the two persistent clones KC9A1 and KC9C1 display similar**

793 **resistance to H_2O_2 and grow similarly in phosphate-limited medium or at acidic pH.**

794 **a**) Flasks of 7H9/ADC/Tween at pH 6.8 were inoculated with STB-K, KC9A1, KC9C1, and *MTB* H37Rv at

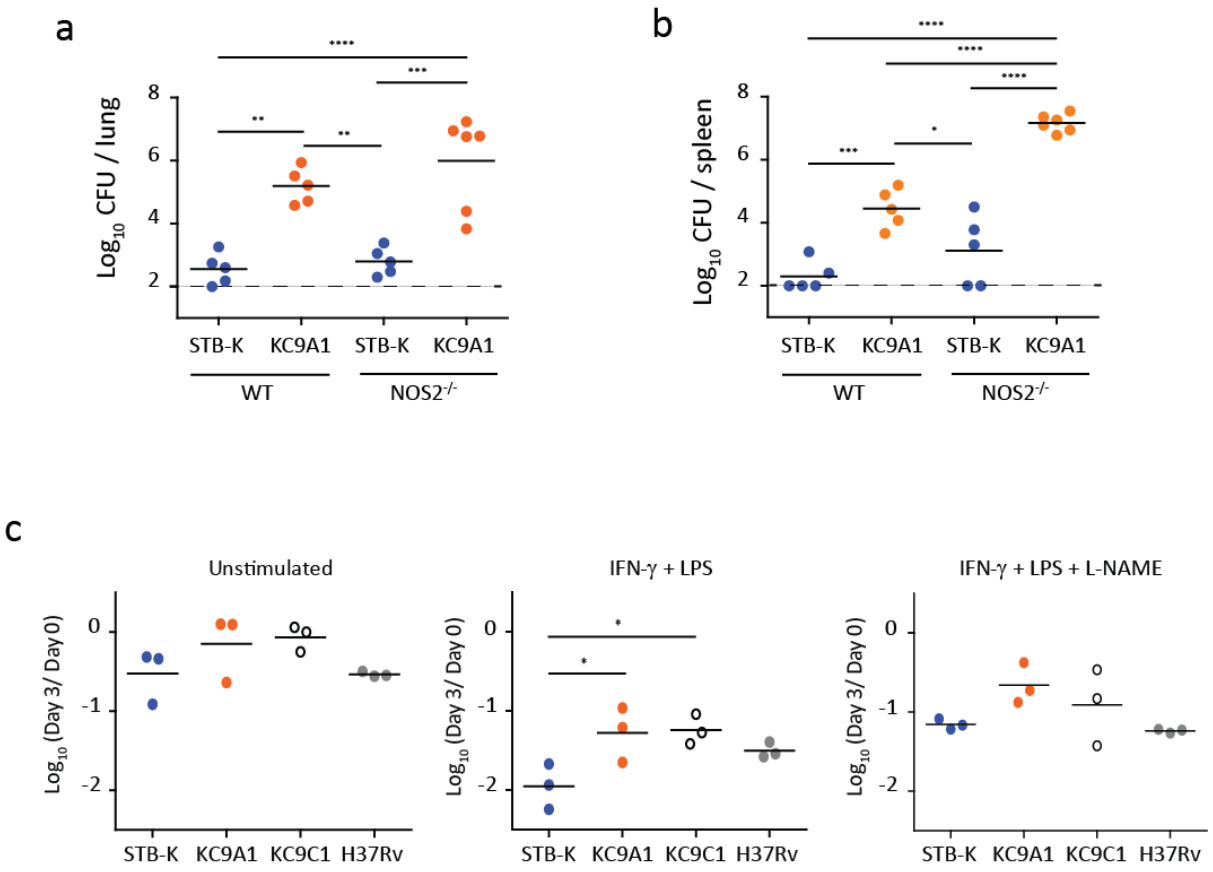
795 a final OD_{600nm} of approximately 0.05 and growth was evaluated by monitoring the OD600. Data are

796 represented as the mean and standard deviation from three independent experiments.

b) STB-K, KC9A1, KC9C1, and H37Rv were inoculated to a final OD_{600nm} of approximately 0.05 into flasks containing 7H9 supplemented with ADC and 0.05% Tyloxapol adjusted to pH 5, 5.5, 6 or 6.6. Strain growth was evaluated by monitoring the OD600. Data are represented as the mean and standard deviation from three independent experiments.

c) STB-K, KC9A1, KC9C1, and H37Rv were inoculated to a final OD_{600nm} of approximately 0.05 into flasks containing Sauton's medium where the final concentration of phosphate had been adjusted to either 0 µM, 50 µM or 2870 µM supplemented with 0.05% Tween 80. Strain growth was evaluated by monitoring the OD600. Data are represented as the mean and standard deviation from three independent experiments.

d) Flasks of 7H9/ADC/Tween + 10 mM H_2O_2 were inoculated with 10^5 CFU of STB-K, KC9A1, KC9C1 and H37Rv. Bacterial survival was monitored by evaluating CFU numbers at 0, 16 and 24 hours post treatment with H_2O_2. Results are shown as $\log_{10}$ CFU/ml for individual data points from eight independent experiments and the mean. When bacteria were undetectable, the CFU value was set at the limit of detection ($\log_{10}$ 1.3) for the purposes of statistical analysis. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by two-way ANOVA with the Bonferroni post hoc test.



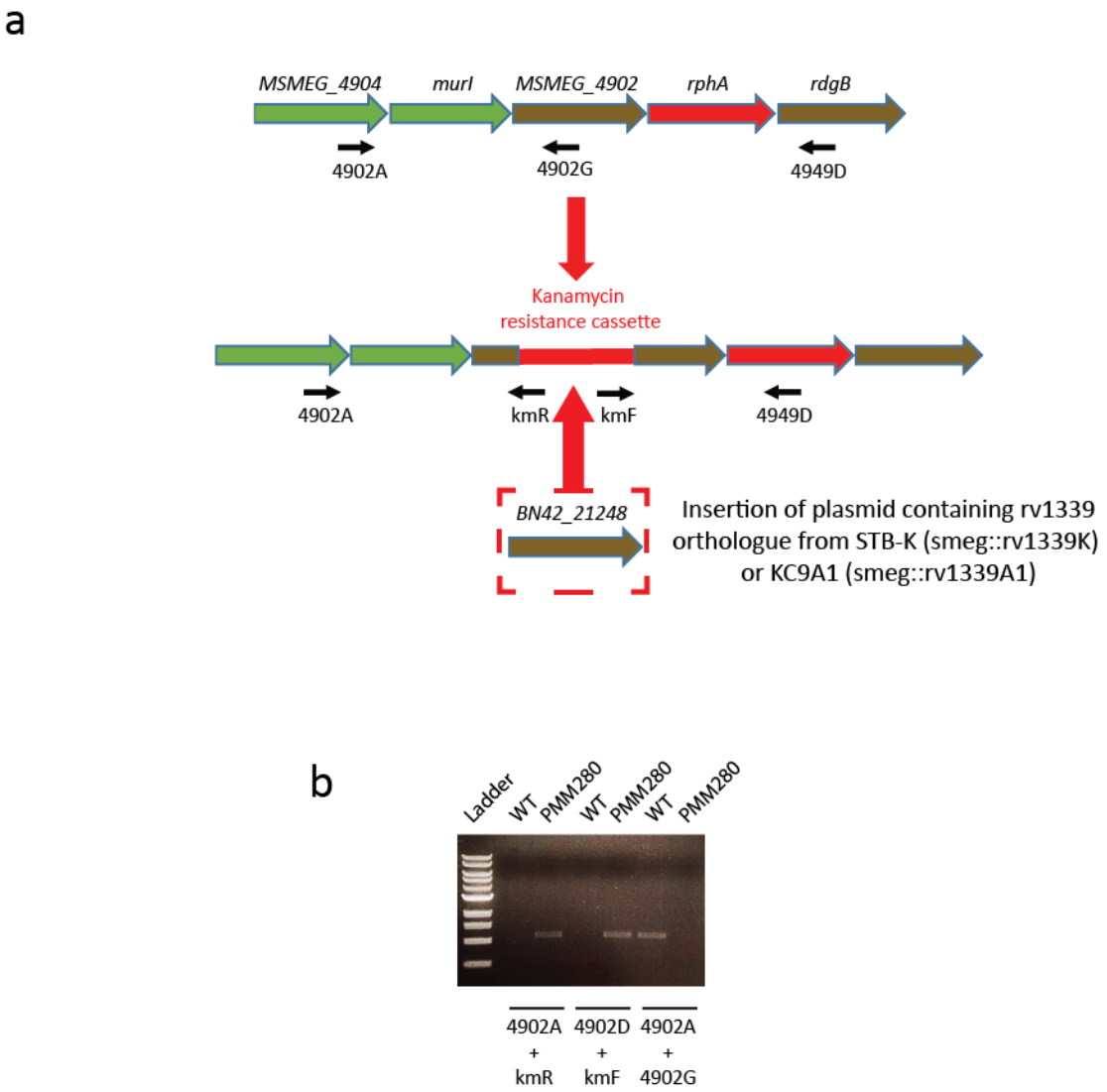
Supplementary Figure 11: The persistence phenotype of the *rv1339* STB-K mutant (KC9A1) is due in part to enhanced resistance to NO.

NOS2$^{-/-}$ mice or C57BL/6 wild-type controls (WT) were infected via aerosol with approximately 100 CFU of STB-K or KC9A1 and bacterial burden in the **a**) lung and **b**) spleen was quantified at 56 days post-infection by CFU plating on solid medium. Results are shown as the $\log_{10}$ CFU for individual mice and the mean from an experiment performed with 5-6 mice per group. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001 by one-way ANOVA with the Bonferroni post hoc test.

Using this route of infection and the C57BL/6 mouse strain, the bacterial burden in the lungs and spleen at day 56 post-infection was consistent with the enhanced persistence phenotype of KC9A1 when compared to STB-K. In NOS2$^{-/-}$ mice, the persistence of STB-K in the spleen and KC9A1 in both organs were increased showing that NO is an important factor reducing STB-K and KC9A1 persistence. However, the difference between STB-K and KC9A1 was maintained in NOS2$^{-/-}$ mice, and even

826 increased in comparison to C57BL/6 mice, indicating that the *rv1339* mutation also confers strong
827 resistance to NOS2$^{-/-}$ independent mechanisms.

828 **c)** BMDMs from BALB/c mice were either unstimulated, or stimulated with IFN-γ (100 U/ml) + LPS (100
829 ng/ml), or stimulated with IFN-γ + LPS and treated with the inducible nitric oxide synthase (iNOS)
830 inhibitor L-NAME (4 mM). BMDMs (2.5 x 10^5 cells/well) were infected at a multiplicity of infection
831 (MOI) of 0.5 bacteria/cell with STB-K, KC9A1, KC9C1 or *MTB* H37Rv and the number of surviving
832 bacteria were evaluated by plating on 7H11 solid medium. The indicated value are the Log_{10} (ratio of
833 bacterial load at day 3 over bacterial load at day 0 in the same experiment). *P < 0.05 by one-way
834 repeated measure ANOVA with the Bonferroni post hoc test.
835 In these experiments, stimulation of BMDMs with IFN-γ and LPS reduced the survival of the 4 strains
836 and this effect is partially suppressed by inhibiting iNOS.
837



838

839 **Supplementary Figure 12: Replacement of the *rv1339* orthologue in *M. smegmatis* mc^{2}155 with the**

840 ***rv1339* orthologue of STB-K or KC9A1.**

841 **a**) A *km* resistance cassette was inserted in the *rv1339* orthologue of *M. smegmatis* mc^{2}155

842 (*MSMEG_4902*) by allelic exchange to create smegΔrv1339 (PMM280). smegΔrv1339 was then

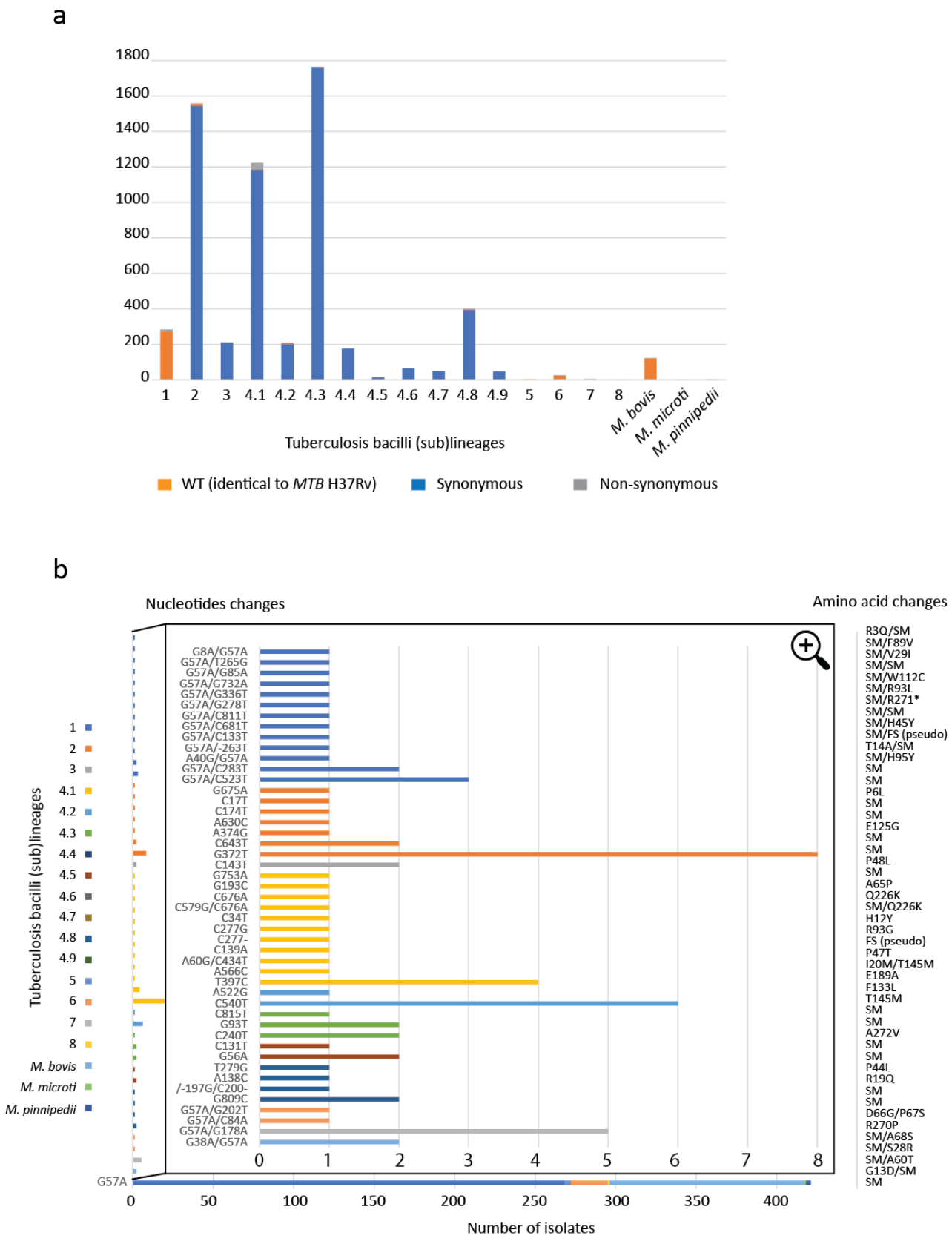
843 complemented with plasmids containing either the *rv1339* orthologue from STB-K (smeg::rv1339K) or

844 KC9A1 (smeg::rv1339A1). The black arrows indicate the primers used to amplify the corresponding

845 regions.

846 **b**) PCR profiles corresponding to a mutant interrupted for the *rv1339* orthologue (PMM280), or the
847 wild-type *M. smegmatis* mc^2155 strain (WT), when the indicated primer pairs were used for
848 amplification.

849



Supplementary Figure 13: Distribution of mutations within the *rv1339* orthologue among *MTB* complex lineages.

a) Wild-type sequences (WT), as well as fixed synonymous and non-synonymous mutations were identified in a collection of genome assemblies from 6167 isolates, which were classified in lineages or main sub-lineages based on phylogenetic SNPs and the conventional nomenclature[66,67].

b) Nucleotide (left) and corresponding amino acid (right) changes or combined changes detected in isolates from different (sub-) lineages are shown. The graph in the center is a zoom-in that represents mutations detected in 8 isolates or less. The G57A mutation detected in >400 isolates (bottom) is a synonymous SNP shared by basal lineages of the *MTB* complex (e.g. lineages 1, 5, 6 and animal lineages). SM, synonymous; FS, frameshift.

These analyses identified about 30 different non-synonymous mutations in *rv1339* in this dataset, which were phylogenetically linked in all cases, i.e. found in one (sub)lineage only. Of note, two frameshift-causing indels and one non-sense mutation were detected, which were however represented only in 2 isolates at most. These results indicate a strong conservation of the *rv1339* gene among *MTB* isolates.

Supplementary Table 1: Mutations identified in STB-K

Cycle 3

Position	Ref	Alt	Mean depth	Ki00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN42_)	Gene (H37Rv)
1582889	C	T	61	0,6					97,8						95,7	93,6	94,9	95,6	94,5	100,0	100,0	100,0	97,4	95,9	21248	*rv1339*
4430235	G	-	88	0,4							95,45														90367	*rv3839*

Cycle 6

Position	Ref	Alt	Mean depth	Ki00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN42_)	Gene (H37Rv)
1582889	C	T	61	0,6					97,8						95,7	93,6	94,9	95,6	94,5	100,0	100,0	100,0	97,4	95,9	21248	*rv1339*
4430235	G	-	88	0,4							95,45														90367	*rv3839*

Cycle 9

Position	Ref	Alt	Mean depth	Ki00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN42_)	Gene (H37Rv)
355663	T	C	19	14,7	90																				20005	
934262	C	T	80	0,4															100,0						20588	*rv0826*
1582889	C	T	86	0,6	100,0	94,8	100,0	96,1	94,5	100,0	92,9	97,8	97,7	97,2	100,0	90,4	100,0	100,0	100,0	100,0	100,0	94,1	95,7	95,9	21248	*rv1339*
1602915	A	G	71	0,4							87,3														intergenic	
2111795	C	G	28	0,9											98,3		100,0	100,0	100,0						30042	*rv1787*
2111828	A	C	20	1,3											97,4		93,8	100,0	100,0						30042	*rv1787*
2113560	-	G	41	2,5						88,68		97,83	90,91	92,73											30044	*rv1789*
2451739	G	A	65	0,6									96,9												30386	*rv2075c*
2451798	T	C	10	2,9									90,0												40077	*rv2162c*
238782	G	C	20	3,8				90,0																	50005	*rv0192*
3952753	G	C	96	0,6																				96,9	50078	*rv3419c*
4202110	C	T	101	0,3							92,1														90152	*rv3649*

Cycle 12

Position	Ref	Alt	Mean depth	Ki00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN42_)	Gene (H37Rv)
355663	T	C	19	14,7	90																				20005	
934262	C	T	80	0,4															100,0						20588	*rv0826*
1582889	C	T	86	0,6	100,0	94,8	100,0	96,1	94,5	100,0	92,9	97,8	97,7	97,2	100,0	90,4	100,0	100,0	100,0	100,0	100,0	94,1	95,7	95,9	21248	*rv1339*
1602915	A	G	71	0,4							87,3														intergenic	
2111795	C	G	28	0,9											98,3		100,0	100,0	100,0						30042	*rv1787*
2111828	A	C	20	1,3											97,4		93,8	100,0	100,0						30042	*rv1787*
2113560	-	G	41	2,5						88,68		97,83	90,91	92,73											30044	*rv1789*
2451739	G	A	65	0,6									96,9												30386	*rv2075c*
2451798	T	C	10	2,9									90,0												40077	*rv2162c*
238782	G	C	20	3,8				90,0																	50005	*rv0192*
3952753	G	C	96	0,6																				96,9	50078	*rv3419c*
4202110	C	T	101	0,3							92,1														90152	*rv3649*

Cycle 15

Position	Ref	Alt	Mean depth	Ki00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN42_)	Gene (H37Rv)
240166	C	G	76	0,4																			100,0		10237	*rv0193c*
382112	C	T	18	0,7										100,0											20026	*rv0297*
386966	C	A	27	1,0			96,3																		20034	*rv0305c*
604649	-	T	113	0,0	92,92																				20231	*rv0501*
1216440	G	A	48	0,0	97,9																				20877	
1511865	G	C	24	1,2														91,7							21173	*rv2885c*
1582889	C	T	108	0,6	100,0	98,8	100,0	100,0	97,4	100,0	99,3	100,0	95,3	99,0	99,0	98,3	99,2	100,0	97,5	99,0	100,0	97,0	98,6	100,0	21248	*rv1339*
2111795	C	G	27	0,9											100,0	100,0	97,2	100,0	100,0						30042	*rv1787*
2111828	A	C	18	1,3											100,0	95,0	100,0	100,0	100,0						30042	*rv1787*
2112495	A	-	117	0,0	98,25	98,05	92,22	100	98,97																30043	*rv1788*
2113560	-	G	52	2,5						96		97,73	91,07	92,86											30044	*rv1789*
2246030	C	T	46	0,0					95,7																30180	
2405288	-	C	99	0,3				95,96																	30346	*rv2042c*
3440220	C	T	99	1,3			100,0																		40962	*rv2962c*
3440958	C	G	82	0,7			100,0																		40963	*rv2963*
3599034	T	G	20	1,8																			95,0		intergenic	
3964805	C	T	66	0,0						100,0	99,0	98,1													intergenic	
3628203	T	C	19	0,0				89,5																	90350	
4407518	T	C	10	0,8				100,0																	90350	

Notes: Position indicates the mutated position in reference to the NCBI GenBank sequence FO203509.1, **Ref** and **Alt** refer to the wild-type and mutated alleles detected respectively, and **Ki00** refers to the frequency of the mutated allele detected in the inoculum. Table cells coloured in grey indicate the deleted region identified in evolution lineage C after cycle 6. The variants identified in this region are presumably false positives due to the recombination of *ppe25* and *ppe27*. Column four mentions mean depth calculated for the 20 sequenced clones from the corresponding cycle at the indicated position.

Supplementary Table 2: Mutations identified in STB-D

Cycle 3

Position	Ref	Alt	Mean depth	Di00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN44_)	Gene (H37Rv)
77436	C	T	267	0,0												100,0			100,0						10083	*rv0066c*
156901	G	A	209	0,0											100,0										10152	*rv0125*
242924	G	A	213	0,0	99,35	100,0																			10234	*rv0200*
303690	G	T	222	0,0						100,0	100,0	100,0	100,0	100,0											10283	*rv0246*
399363	T	G	267	1,7												97,4			98,1						10367	*rv0327c*
443358	C	T	289	29,8	100,0	100,0	100,0	100,0		100,0	99,6	100,0	99,6	100,0	100,0		100,0	100,0		99,6	100,0	100,0	100,0	99,6	10398	*rv0357c*
1914166	G	A	231	0,0																			100,0		20251	*rv1686c*
2234377	C	T	163	0,0																	100,0				40254	
2354094	C	T	204	0,0																				100,0	40367	*rv2075c*
2473696	C	T	236	0,0				99,5																	50119	*rv2184c*
2584587	G	A	228	0,0																	100,0				50225	*rv2281*
2686044	C	T	218	1,3													99,1								intergenic	
2835137	C	A	112	0,8									92,1												50477	
2877937	C	A	177	0,5												99,5			100,0						50519	*rv2524c*
3746850	A	G	166	0,0																			98,5		70122	*rv3109*
4056021	G	A	130	0,0												99,5			100,0						110089	*rv3594*
4156062	A	T	281	2,6												99,6			100,0						120093	*rv3696c*
4156189	G	T	235	28,1	96,6	99,0	98,0	99,0		98,3	98,2	95,7	97,4	100,0	100,0		98,3	98,8		98,1	99,1	98,0	99,1	97,8	120093	*rv3696c*
4374850	C	T	171	3,3												100,0			100,0						120307	*rv3876*

Cycle 6

Position	Ref	Alt	Mean depth	Di00	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	D1	D2	D3	D4	D5	Gene (BN44_)	Gene (H37Rv)
77436	C	T	267	0,0															100,0			100,0	100,0		10083	*rv0066c*
242924	G	A	213	0,0	98,1	99,5	99,5	100,0	98,2																10234	*rv0200*
303690	G	T	222	0,0						100,0	100,0	100,0	100,0	100,0											10283	*rv0246*
399363	T	G	267	1,7															97,2			96,6	96,9		10367	*rv0327c*
443358	C	T	289	29,8	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0						100,0	100,0			100,0	10398	*rv0357c*
966533	C	T	208	0,0																100,0					10955	*rv0873*
1072346	G	C	209	0,0														88,2							11060	*rv0959*
1078392	C	T	154	0,7																98,5					11070	*rv0969*
1364332	C	G	230	0,0																	99,0				11369	*rv1225c*
1655576	G	A	214	0,0												99,5									20018	*rv1464*
2392213	G	A	162	0,0																				99,1	50026	*rv2101*
2877937	C	A	177	0,5															100,0			100,0	100,0		50519	*rv2524c*
3371167	A	T	187	0,0											100,0	100,0	100,0	100,0							60474	*rv2990c*
4100847	G	C	225	0,0																99,1					120034	*rv3645*
4156062	A	T	281	2,6															100,0			100,0	100,0		120093	*rv3696c*
4156189	G	T	235	28,1	98,9	98,2	99,5	100,0	100,0	100,0	98,4	97,9	99,4	98,2						99,6	98,9			99,4	120093	*rv3696c*
4338549	A	G	242	9,0											98,3	98,3	96,8	98,1							120265	
4374850	C	T	171	3,3															100,0			100,0	100,0		120307	*rv3876*

Notes: Position indicates the mutated position in reference to the NCBI GenBank sequence NC_019950.1, **Ref** and **Alt** refer to the wild-type and mutated alleles detected respectively, and **Di00** refers to the frequency of the mutated allele detected in the inoculum. The light orange and green boxes indicate SNPs segregated together in different evolution lineages. Most of these SNPs were detected above background in the inoculum supporting the conclusion that the STB-D inoculum, used to start the experimental evolution, contains at least three genetic profiles.

Supplementary Table 3: *M. canettii, MTB,* and *M. smegmatis* strains used in this study

Strain	Description
STB-K	*M. canettii* CIPT 140010070
STB-D	*M. canettii* CIPT 140060008
STB-A	*M. canettii* CIPT 140060001
STB-J	*M. canettii* CIPT 140070017
STB-L	*M. canettii* CIPT 140070008
KC6A1	Clone isolated from lineage A following 6 cycles of experimental evolution with STB-K
KC6A2	Clone isolated from lineage A following 6 cycles of experimental evolution with STB-K
KC6B1	Clone isolated from lineage B following 6 cycles of experimental evolution with STB-K
KC6B2	Clone isolated from lineage B following 6 cycles of experimental evolution with STB-K
KC6C1	Clone isolated from lineage C following 6 cycles of experimental evolution with STB-K
KC6C2	Clone isolated from lineage C following 6 cycles of experimental evolution with STB-K
KC6D1	Clone isolated from lineage D following 6 cycles of experimental evolution with STB-K
KC6D2	Clone isolated from lineage D following 6 cycles of experimental evolution with STB-K
KC9A1	Clone isolated from lineage A following 9 cycles of experimental evolution with STB-K
KC9B1	Clone isolated from lineage B following 9 cycles of experimental evolution with STB-K
KC9C1	Clone isolated from lineage C following 9 cycles of experimental evolution with STB-K
KC9D1	Clone isolated from lineage D following 9 cycles of experimental evolution with STB-K
KC12A1	Clone isolated from lineage A following 12 cycles of experimental evolution with STB-K
KC12B1	Clone isolated from lineage B following 12 cycles of experimental evolution with STB-K
KC12C1	Clone isolated from lineage C following 12 cycles of experimental evolution with STB-K
KC12D1	Clone isolated from lineage D following 12 cycles of experimental evolution with STB-K
KC15A1	Clone isolated from lineage A following 15 cycles of experimental evolution with STB-K
KC15B1	Clone isolated from lineage B following 15 cycles of experimental evolution with STB-K
KC15C1	Clone isolated from lineage C following 15 cycles of experimental evolution with STB-K
KC15D1	Clone isolated from lineage D following 15 cycles of experimental evolution with STB-K
DC6A1	Clone isolated from lineage A following 6 cycles of experimental evolution with STB-D
DC6B1	Clone isolated from lineage B following 6 cycles of experimental evolution with STB-D
DC6C1	Clone isolated from lineage C following 6 cycles of experimental evolution with STB-D
DC6D1	Clone isolated from lineage D following 6 cycles of experimental evolution with STB-D
KC9A1::EV	KC9A1 complemented with a derivative of the integrative plasmid pMV361
KC9C1::EV	KC9C1 complemented with a derivative of the integrative plasmid pMV361
KC9C1::K	KC9C1 complemented with integrative plasmid pWM361 containing *ppe25-ppe27* of STB-K
D::EV	STB-D complemented with a derivative of the integrative plasmid pMV361
DΔppe25-ppe27 (PMM304)	STB-D deleted for the *ppe25-ppe27* region
DΔppe25-ppe27::EV	PMM304 complemented with a derivative of the integrative plasmid pMV361
DΔppe25-ppe27::D	PMM304 complemented with plasmid pWM385 containing *ppe25-ppe27* of STB-D
DΔppe25-ppe27::K	PMM304 complemented with plasmid pWM361 containing *ppe25-ppe27* of STB-K
DΔppe25-ppe27::C1	PMM304 complemented with plasmid pWM386 containing *ppe25-ppe27* of KC9C1
DΔrv1339 (PMM305)	STB-D interrupted for the gene *BN44_11505* (*rv1339* orthologue)
::rv1339A1 (PMM305::pAH01H)	PMM305 complemented with plasmid pAH01H containing *BN42_21248* of KC9A1 (mutated *rv1339)*
::rv1339K (PMM305::pAH02H)	PMM305 complemented with plasmid pAH02H containing *BN42_21248* of STB-K (wild-type *rv1339)*
H37Rv	*MTB* Lineage 4
Erdman	*MTB* Lineage 4
IO367	*MTB* Lineage 1
HN878	*MTB* Lineage 2
mc^2155	*M. smegmatis* mc^2155
smegΔrv1339 (PMM280)	mc^2155 interrupted for the gene *MSMEG_4902* (*rv1339* orthologue)
smeg::rv1339K	PMM280 complemented with plasmid pAH02H containing *BN42_21248* of STB-K (wild-type *rv1339*)
smeg::rv1339A1	PMM280 complemented with plasmid pAH01H containing *BN42_21248* of KC9A1 (mutated *rv1339*)

906 **Supplementary Table 4: Plasmids used in this study**

907

Name	Description
pWM355	*BN44_11505* gene from STB-D cloned into pJET
pWM359	*BN44_11505* gene disrupted by a *km* cassette carried by the pJET vector
pWM361	*ppe25-ppe27* fragment from STB-K cloned into a derivative of pMV361 containing a *hyg* resistance gene
pWM386	*ppe25-ppe27* fragment from KC9C1 cloned into a derivative of pMV361 containing a *hyg* resistance gene
pWM385	*ppe25-ppe27* fragment from STB-D cloned into a derivative of pMV361 containing a *hyg* resistance gene
pAH01H	*BN42_21248* gene from KC9A1 cloned into a derivative of pMV361 containing a *hyg* resistance gene
pAH02H	*BN42_21248* gene from STB-K cloned into a derivative of pMV361 containing a *hyg* resistance gene
pAH06	DNA fragments overlapping the N-terminal and C-terminal parts of *MSMEG_4902* cloned into pJET
pAH10	*MSMEG_4902* gene disrupted by a *km* cassette carried by the pJQ200 vector

908

Supplementary Table 5: Primers used in this study

Name	Sequence
39a	GCGTTGGTGATCGCCTGC
39b	CGGTGTCGGTAGCGACGA
39c	GCGACGATCTGGTCGGC
39d	CCCTGAATCTCTACCAGGG
39e	CCGTGCTCGGATGCTCC
39f	CGTCAGCAGCAGCTCACG
km1	GTCTGACGCTCAGTGGAAC
km2	TTCAGGTGGCACTTTTCGG
k1	TTTCACGAAGTGTCATGAACAATAAAACTGTCTGC
k2	TTTCACTCTGTGTCATGAATTAATTCTTAGAAAAACT
esx1	TTTCACAAAGTGGCCAAACGATCGTTTACGAG
esx2	TTTCACTTCGTGCCGAAGGTCAGCTGCTGT
esx5	TTTCACAGAGTGCTGCAGATAGCGATCGAC
esx6	TTTCACCTTGTGCAGCATGTTCACGATG
esx7	ATCCTCGAGGAGGACGAC
esx8	CGTAGTTGTTGGCGTCGC
26a	ATCATATGGATTTTGGGGCGTTGC
26b	ATACTAGTCTATCCGGCGAAGGGTGG
4902A	GGGTGTCGTGGCAGGGA
4902B	P-ATCGCACACTGCTAGCT
4902C	P-ATCGGCTGACCTGCGCGTCG
4902D	GCCACGCACGCCAGCC
4902E	CCGAACGCAAGGCACCGG
4902F	GCGTTCTCCTCGAAAGTCG
4902G	CCGAAGTCGATGACCATCGG
kmF	GCCATCCTATGGAACTGCC
kmR	GCCTAGAGCAAGACGTTTCC
1339G	ATCATATGCGTCGATGTATTCCGC
1339H	ATACTAGTCTAGCCGGCTCGC
esxA	ATTCTAGAGGGTTCTCATGCGCTCAAG
esxB	ATAAGCTTGGGACGATCAATTGGGCAT
esxC	ATTCTAGACGCTCAAGTCTTTCGTGTCT
esxD	ATAAGCTTCGGTCGATCGCTATCTGC

Acknowledgements

The authors thank T. Seemann for the initial help with the network phylogeny analysis. The authors are grateful to Erica Russo and Philippe Bousso (Institut Pasteur, France) for the gift of $NOS2^{-/-}$ mice. The authors acknowledge support from the GenoToul (Toulouse, France) Bioinformatics and GeT platforms for sequencing, computing and storage resources, and the Anexplo platform for animal experimentation. The authors thank Samantha Milia for technical assistance at the Experimental Histopathology Facility of INSERM / UPS US006 CREFRE, Toulouse Purpan, France. The authors also thank Life Science Editors for their professional editing assistance during the preparation of the manuscript.

This project has received funding from the European Respiratory Society and the European Union's H2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 713406 awarded to A.C.A. This work was also supported by grants from the French National Research Agency (TBemerg contract N°ANR-16-CE35-0009 awarded to R.B., P.S. and C.G.; , Equipex ANINIMIP ANR-11-EQUIPEX-0003 awarded to C.G.; Labex ANR-10-LABX-62-IBEID awarded to R.B.), by a grant from the European Union (PathoNgenTrace contract N°FP7-278864 awarded to P.S.), and by grants from the Fondation pour la Recherche Médicale (Equipe FRM 2016 DEQ20160334879), the Fondation Bettencourt Schueller and the Fondation MSD Avenir awarded to C.G.

Author contributions

C.G., R.B. and P.S. planned the project. A.C.A., C.A.-D. and C.G. designed the experiments. A.C.A., W. M., C.G., F.M., A.H., C.A.-D., A.P., R.A., A.P., W.F., C.B. and C.G. performed the experiments. A.C.A., W.M., C.G., A.V., A.H., C.A.-D., A.P., R.A., A.P., R.B., P.S. and C.G. analysed the data. A.C.A., C.A.-D., R.B., P.S. and C.G. contributed to the manuscript.

Conflict of interest

The authors declare no competing financial interests.

938

939 **References**

940 1 Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Revisiting the timetable of tuberculosis. *BMJ*
941 **362**, k2738 (2018).

942 2 WHO. Global tuberculosis report. (2019).

943 3 Chiner-Oms, A. *et al.* Genomic determinants of speciation and spread of the *Mycobacterium*
944 *tuberculosis* complex. *Sci. Adv.* **5**, eaaw3307 (2019).

945 4 Blouin, Y. *et al.* Progenitor "*Mycobacterium canettii*" clone responsible for lymph node
946 tuberculosis epidemic, Djibouti. *Emerg. Inf. Dis.* **20**, 21-28 (2014).

947 5 Koeck, J.-L. *et al.* Clinical characteristics of the smooth tubercle bacilli '*Mycobacterium canettii*'
948 infection suggest the existence of an environmental reservoir. *Clin. Microbiol. Infect.* **17**, 1013-
949 1019 (2011).

950 6 Fabre, M. *et al.* Molecular characteristics of "*Mycobacterium canettii*" the smooth
951 *Mycobacterium tuberculosis* bacilli. *Infect. Genet. Evol.* **10**, 1165-1173 (2010).

952 7 Lopez, B. *et al.* A marked difference in pathogenesis and immune response induced by
953 different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* **133**, 30-37 (2003).

954 8 Dormans, J. *et al.* Correlation of virulence, lung pathology, bacterial load and delayed type
955 hypersensitivity responses after infection with different *Mycobacterium tuberculosis*
956 genotypes in BALB/c mouse model. *Clin. Exp. Immunol.* **137**, 460-468 (2004).

957 9 Supply, P. *et al.* Genomic analysis of smooth tubercle bacilli provides insights into ancestry
958 and pathoadaptation of *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 172-179 (2013).

959 10 Boritsch, E. C. *et al.* A glimpse into the past and predictions for the future: the molecular
960 evolution of the tuberculosis agent. *Mol. Microbiol.* **93**, 835-852 (2014).

961 11 Fabre, M. *et al.* High genetic diversity revealed by variable-number tandem repeat genotyping
962 and analysis of *hsp65* gene polymorphism in a large collection of "*Mycobacterium canettii*"

963 strains indicates that the *M. tuberculosis* complex is a recently emerged clone of "*M. canettii*".
964 *J. Clin. Microbiol.* **42**, 3248-3255 (2004).
965 12 Gutierrez, M. C. *et al.* Ancient orgin and gene mosaicism of the progenitor of *Mycobacterium*
966 *tuberculosis*. *PLOS Pathog.* **1**, e5 (2005).
967 13 Thomson, M. *et al.* Modulation of the cAMP levels with a conserved actinobacteria
968 phosphodiesterase enzyme reduces antimicrobial tolerance in mycobacteria. *bioRxiv*,
969 08.26.267864 (2020).
970 14 Kramnik, I., Demant, P. & Bloom, B. B. Susceptibility to tuberculosis as a complex genetic trait:
971 analysis using recombinant congenic strains of mice. *Novartis Foundation symposium* **217**,
972 120-131; discussion 132-127 (1998).
973 15 Kramnik, I., Dietrich, W. F., Demant, P. & Bloom, B. R. Genetic control of resistance to
974 experimental infection with virulent *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*
975 **97**, 8560-8565 (2000).
976 16 Lanoix, J. P., Lenaerts, A. J. & Nuermberger, E. L. Heterogeneous disease progression and
977 treatment response in a C3HeB/FeJ mouse model of tuberculosis. *Dis. Mod. Mech.* **8**, 603-610
978 (2015).
979 17 Harper, J. *et al.* Mouse model of necrotic tuberculosis granulomas develops hypoxic lesions. *J.*
980 *Inf. Dis.* **205**, 595-602 (2012).
981 18 Singh, A. *et al. Mycobacterium tuberculosis* WhiB3 maintains redox homeostasis by regulating
982 virulence lipid anabolism to modulate macrophage response. *PLOS Pathog.* **5**, e1000545
983 (2009).
984 19 Vandal, O. H., Pierini, L. M., Schnappinger, D., Nathan, C. F. & Ehrt, S. A membrane protein
985 preserves intrabacterial pH in intraphagosomal *Mycobacterium tuberculosis*. *Nat. Med.* **14**,
986 849-854 (2008).

20 Darwin, K. H., Ehrt, S., Gutierrez-Ramos, J. C., Weich, N. & Nathan, C. F. The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science* **302**, 1963-1966 (2003).

21 Albrethsen, J. *et al.* Proteomic profiling of *Mycobacterium tuberculosis* identifies nutrient-starvation-responsive toxin-antitoxin systems. *Mol. Cell. Prot.* **12**, 1180-1191 (2013).

22 Schaible, U. E., Sturgill-Koszycki, S., Schlesinger, P. H. & Russell, D. G. Cytokine activation leads to acidification and increases maturation of *Mycobacterium avium*-containing phagosomes in murine macrophages. *J Immunol* **160**, 1290-1296 (1998).

23 MacMicking, J. D., Taylor, G. A. & McKinney, J. D. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. *Science* **302**, 654-659 (2003).

24 Neyrolles, O., Wolschendorf, F., Mitra, A. & Niederweis, M. Mycobacteria, metals, and the macrophage. *Immunol. Rev.* **264**, 249-263 (2015).

25 Rifat, D., Bishai, W. R. & Karakousis, P. C. Phosphate depletion: a novel trigger for Mycobacterium tuberculosis persistence. *J. Inf. Dis.* **200**, 1126-1135 (2009).

26 Rengarajan, J., Bloom, B. R. & Rubin, E. J. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA* **102**, 8327-8332 (2005).

27 Wolschendorf, F. *et al.* Copper resistance is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **108**, 1621-1626 (2011).

28 Botella, H. *et al.* Mycobacterial P1-Type ATPases mediate resistance to zinc poisoning in human macrophages. *Cell Host Microbe* **10**, 248-259 (2011).

29 MacMicking, J. D. *et al.* Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc. Natl. Acad. Sci. USA* **94**, 5243-5248 (1997).

30 Nicholson, S. *et al.* Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. *J. Exp. Med.* **183**, 2293-2302 (1996).

31 Rich, E. A. *et al. Mycobacterium tuberculosis* (MTB)-stimulated production of nitric oxide by human alveolar macrophages and relationship of nitric oxide production to growth inhibition of MTB. *Tubercle Lung Dis.* **78**, 247-255 (1997).

32 Martineau, A. R. *et al.* Neutrophil-mediated innate immune resistance to mycobacteria. *J. Clin. Invest.* **117**, 1988-1994 (2007).

33 Martineau, A. R. *et al.* IFN-gamma- and TNF-independent vitamin D-inducible human suppression of mycobacteria: the role of cathelicidin LL-37. *J Immunol* **178**, 7190-7198, (2007).

34 Liu, P. T., Stenger, S., Tang, D. H. & Modlin, R. L. Cutting edge: vitamin D-mediated human antimicrobial activity against *Mycobacterium tuberculosis* is dependent on the induction of cathelicidin. *J Immunol* **179**, 2060-2063 (2007).

35 Stenger, S. *et al.* An antimicrobial activity of cytolytic T cells mediated by granulysin. *Science* **282**, 121-125 (1998).

36 Buchmeier, N. *et al.* A parallel intraphagosomal survival strategy shared by *Mycobacterium tuberculosis* and *Salmonella enterica*. *Mol. Microbiol.* **35**, 1375-1382, (2000).

37 Ng, V. H., Cox, J. S., Sousa, A. O., MacMicking, J. D. & McKinney, J. D. Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. *Mol. Microbiol.* **52**, 1291-1302 (2004).

38 Primm, T. P. *et al.* The stringent response of *Mycobacterium tuberculosis* is required for long-term survival. *J. Bacteriol.* **182**, 4889-4898 (2000).

39 Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275-1280 (2011).

40 Lenski, R. E. Chance and necessity in the evolution of a bacterial pathogen. *Nat. Genet.* **43**, 1174-1176 (2011).

41 Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751-764 (2013).

42 Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W. & Wilson, D. J. Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* **14**, 150-162 (2016).

1038 43 Wirth, T. *et al.* Niche specialization and spread of *Staphylococcus capitis* involved in neonatal
1039 sepsis. *Nat. Microbiol.* **5**, 735-745 (2020).
1040 44 Chan, J., Xing, Y., Magliozzo, R. S. & Bloom, B. R. Killing of virulent *Mycobacterium tuberculosis*
1041 by reactive nitrogen intermediates produced by activated murine macrophages. *J. Exp. Med.*
1042 **175**, 1111-1122 (1992).
1043 45 Flesch, I. E. & Kaufmann, S. H. Mechanisms involved in mycobacterial growth inhibition by
1044 gamma interferon-activated bone marrow macrophages: role of reactive nitrogen
1045 intermediates. *Infect. Immun.* **59**, 3213-3218 (1991).
1046 46 Long, R., Light, B. & Talbot, J. A. Mycobacteriocidal action of exogenous nitric oxide.
1047 *Antimicrob. Agents Chemother.* **43**, 403-405 (1999).
1048 47 Nozaki, Y., Hasegawa, Y., Ichiyama, S., Nakashima, I. & Shimokata, K. Mechanism of nitric
1049 oxide-dependent killing of *Mycobacterium bovis* BCG in human alveolar macrophages. *Infect.*
1050 *Immun.* **65**, 3644-3647 (1997).
1051 48 Mishra, B. B. *et al.* Nitric oxide prevents a pathogen-permissive granulocytic inflammation
1052 during tuberculosis. *Nat. Microbiol.* **2**, 17072 (2017).
1053 49 Mishra, B. B. *et al.* Nitric oxide controls the immunopathology of tuberculosis by inhibiting
1054 NLRP3 inflammasome-dependent processing of IL-1beta. *Nat. Immunol.* **14**, 52-60 (2013).
1055 50 Chandrasekera, N. S. *et al.* Improved phenoxyalkylbenzimidazoles with activity against
1056 *Mycobacterium tuberculosis* appear to target QcrB. *ACS Infect. Dis.* **3**, 898-916 (2017).
1057 51 O'Malley, T. *et al.* Imidazopyridine Compounds Inhibit Mycobacterial Growth by Depleting ATP
1058 Levels. *Antimicrob. Agents Chemother.* **62**, e02439-17 (2018).
1059 52 Nazarova, E. V. *et al.* The genetic requirements of fatty acid import by *Mycobacterium*
1060 *tuberculosis* within macrophages. *eLife* **8** (2019).
1061 53 Karboul, A. *et al.* Frequent homologous recombination events in *Mycobacterium tuberculosis*
1062 PE/PPE multigene families: potential role in antigenic variability. *J. Bacteriol.* **190**, 7838-7846
1063 (2008).

1064 54 Ates, L. S. *et al.* Mutations in *ppe38* block PE_PGRS secretion and increase virulence of
1065 *Mycobacterium tuberculosis. Nat. Microbiol.* **3**, 181-188 (2018).
1066 55 Wang, Q. *et al.* PE/PPE proteins mediate nutrient transport across the outer membrane of
1067 *Mycobacterium tuberculosis. Science* **367**, 1147-1151 (2020).
1068 56 Helsen, J. *et al.* Gene loss predictably drives evolutionary adaptation. *Mol. Biol. Evol.* **37**, 2989-
1069 3002 (2020).
1070 57 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
1071 357-359 (2012).
1072 58 van Kessel, J. C. & Hatfull, G. F. Mycobacterial recombineering. *Methods Mol. Biol.* **435**, 203-
1073 215 (2008).
1074 59 Pelicic, V., Reyrat, J.-M. & gicquel, B. Generation of unmarked directed mutations in
1075 mycobacteria, using sucrose counterselectable suicide vectors. *Mol. Microbiol.* **20**, 919-925
1076 (1996).
1077 60 Stover, C. K. *et al.* New use of BCG for recombinant vaccines. *Nature* **351**, 456-460 (1991).
1078 61 Le Dantec, C., Winter, N., Gicquel, B., Vincent, V. & Picardeau, M. Genomic sequence and
1079 transcriptional analysis of a 23-kb mycobacterial linear plasmid: evidence for horizontal
1080 transfer and identification of plasmid maintenance systems. *J. Bacteriol.* **183**, 2157-2164
1081 (2001).
1082 62 Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci.*
1083 *Rep.* **7**, 16878 (2017).
1084 63 Rodrigues, L., Viveiros, M. & Ainsa, J. A. Measuring efflux and permeability in mycobacteria.
1085 *Methods Mol. Biol.* **1285**, 227-239 (2015).
1086 64 Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol.*
1087 *Biol. Evol.* **23**, 254-267 (2006).
1088 65 Comas, I. *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily
1089 hyperconserved. *Nat. Genet.* **42**, 498-503 (2010).

66 Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4812 (2014).

67 Merker, M. *et al.* Phylogenetically informative mutations in genes implicated in antibiotic resistance in *Mycobacterium tuberculosis* complex. *Genome Med.* **12**, 27 (2020).