Ecology and evolution of *Mycobacterium tuberculosis*

Sebastien Gagneux[1,2]

Abstract | Tuberculosis (TB) is the number one cause of human death due to an infectious disease. The causative agents of TB are a group of closely related bacteria known as the *Mycobacterium tuberculosis* complex (MTBC). As the MTBC exhibits a clonal population structure with low DNA sequence diversity, methods (such as multilocus sequence typing) that are applied to more genetically diverse bacteria are uninformative, and much of the ecology and evolution of the MTBC has therefore remained unknown. Owing to recent advances in whole-genome sequencing and analyses of large collections of MTBC clinical isolates from around the world, many new insights have been gained, including a better understanding of the origin of the MTBC as an obligate pathogen and its molecular evolution and population genetic characteristics both within and between hosts, as well as many aspects related to antibiotic resistance. The purpose of this Review is to summarize these recent discoveries and discuss their relevance for developing better tools and strategies to control TB.

Acid-fast bacilli
Mycobacteria that have a thick, lipid-rich cell wall that retains staining despite acid treatment; hence 'acid-fast'.

Multilocus sequence typing
A standard genotyping method based on sequence data from approximately seven housekeeping genes, which together define strain-specific sequence types.

Professional pathogen
A pathogen with no environmental reservoir that has to cause disease to transmit from host to host.

Fast-growers
Mycobacteria that form colonies in less than 7 days.

[1]*Swiss Tropical and Public Health Institute, Basel, Switzerland.*
[2]*University of Basel, Basel, Switzerland.*
sebastien.gagneux@swisstph.ch

doi:10.1038/nrmicro.2018.8
Published online 19 Feb 2018

More than 1 billion people have died of tuberculosis (TB) during the past 200 years[1]. Today, TB remains the number one cause of human death due to a single infectious agent, causing an estimated 10.4 million new cases and 1.7 million deaths per year[2]. Furthermore, one-quarter of the global human population is latently infected, providing a large reservoir for future cases of active TB[3]. Global TB control is hampered by the fact that it continues to largely rely on a 120-year-old microscopy-based diagnostic technique with low sensitivity, a 100-year-old vaccine (Bacille Calmette–Guérin; BCG) with limited effectiveness and drugs that are between 40 and 60 years old[4]. As a result, multiple-drug resistance (MDR) and extensive-drug resistance (XDR) are a growing problem and, together with HIV co-infection and other comorbidities such as type 2 diabetes, are fuelling the TB epidemics[4]. In addition to outdated control tools, progress in the field has been held back by various dogmas. One of these is the simplistic notion that drug-resistant bacteria are less fit than drug-susceptible strains[5]. Another views TB bacilli as a clone with no relevant genotypic and phenotypic variation across strains[6]. Human TB is mainly caused by members of the *Mycobacterium tuberculosis* complex (MTBC), a group of closely related (>99% nucleotide sequence identity) acid-fast bacilli. As the MTBC has little sequence variation compared with other bacteria[7], multilocus sequence typing, one of the main techniques that is used to study bacterial genetic variation, is of limited use, and much of the ecology and evolution of the MTBC has therefore remained unknown. The recent advances in next-generation sequencing overcome these limitations by indexing MTBC diversity across the whole genome, thereby enabling detailed analyses of the evolutionary forces driving this diversity[8].

In this Review, I discuss the latest insights into how the MTBC developed from an environmental organism into a professional pathogen, when and where this transition occurred and how the evolution of human-adapted MTBC compares with that of the MTBC that infects animals. Moreover, I summarize our current understanding of how the various phylogenetic lineages of the MTBC are distributed around the world, how this relates to their epidemiology and what this tells us about the ecological niche of these microorganisms. Finally, I outline the consequences of strict clonality for the evolution of the MTBC within and between individuals, as well as for the development of antibiotic resistance.

Mycobacteria phylogeny

The genus *Mycobacterium* comprises more than 170 species, most of which are environmental organisms[9]. Traditionally, *Mycobacterium* species have been divided into fast-growers and slow-growers, with the three major mycobacterial pathogens of humans (that is, the MTBC, *Mycobacterium leprae* and *Mycobacterium ulcerans*) belonging to the slow-growers group[10]. In addition, several so-called non-tuberculous mycobacteria (NTMs) can cause disease in immune-compromised individuals. These include *Mycobacterium abscessus*

(a fast-grower) and the slow-growers *Mycobacterium avium*, *Mycobacterium marinum*, *Mycobacterium xenopi*, *Mycobacterium gordonae* and *Mycobacterium kansasii*. However, the human-adapted MTBC is unique as it is an obligate pathogen of humans without any environmental or animal reservoir. The MTBC also has to cause disease to transmit between individuals[11], which is unlike many other pathogens where virulence is not directly linked to their transmission[12]. How did the MTBC evolve to become such a successful pathogen?

From an environmental organism to a professional pathogen. The current view is that the MTBC emerged as a professional pathogen from an environmental *Mycobacterium* through step-wise adaptation to an intracellular milieu. It has been hypothesized that one key step in this process was the ability to survive in free-living protozoa (for example, amoebae), which feed on environmental bacteria[13]. This ability might then have helped the ancestor of the MTBC to infect and multiply within mammalian macrophages, an important characteristic of MTBC virulence[14]. The final crucial step was developing the capacity to directly spread from host to host, which in the case of human TB occurs exclusively through airborne droplets that are generated when infected individuals cough[4]. A recent theoretical study proposed that controlled fire use by early humans may have contributed to the evolution of TB transmission by simultaneously promoting social interaction (for example, around camp fires) and smoke-induced lung damage[15].

What were the genetic changes that enabled the transition of the MTBC ancestor to a professional pathogen? Studies that compared the MTBC genome with those of the closely related NTM species *M. marinum* and *M. kansasii* suggest genome downsizing through deletion of genes that are dispensable for a pathogenic lifestyle, combined with the acquisition of new genes through horizontal gene transfer (HGT)[16]. With a length of 4.4 Mb, the genome of the MTBC is smaller than the genomes of *M. marinum* and *M. kansasii* (6.6 Mb and 6.4 Mb, respectively)[17,18]. At the same time, the MTBC ancestor acquired several hundred genes through HGT since its divergence from *M. marinum* and *M. kansasii*, including genes encoding transferases and genes related to adaptation to anaerobic conditions[19–22]. Importantly, however, these comparative studies also revealed that many of the known virulence factors in the MTBC, including the PhoPR two-component system, the DosR/S/T regulon, the mce-associated genes and components of the ESAT6 secretion (ESX; also known as type VII secretion) system, are present in non-pathogenic NTMs[18]. Similarly, while the proline-glutamic acid (PE) and proline-proline-glutamic acid (PPE) protein families (protein families of mainly unknown function that are characterized by conserved N-terminal domains and highly variable C-terminal domains)[23] are over-represented in pathogenic mycobacteria, they also occur in non-pathogenic NTMs, albeit in fewer numbers[24]. Among the PE and PPE protein families, the subfamily of PE_polymorphic guanine-cytosine-rich sequence (PE_PGRS) proteins is restricted to pathogenic mycobacterial species, including *M. leprae*, *M. marinum*, *M. ulcerans* and the MTBC[24].

Taken together, these observations indicate that no single genomic feature accounts for the full virulence and transmission potential of the MTBC in humans. Instead, epistatic interactions between many genomic loci and their coordinated transcriptional regulation seem to be responsible. Interestingly, the MTBC stands out among most other bacteria by harbouring a disproportionally high number of toxin–antitoxin system genes[18,25], some of which have been shown to be differentially regulated in clinical strains[26]. Toxin–antitoxin systems have been linked to many functions in bacteria. These include the regulation of gene expression, growth rate and persistence, supporting the notion that these systems may have an important role in regulating the pathogenic life cycle of the MTBC[25,27].

***Insights from comparison with* Mycobacterium canettii.** Insights into the emergence of the MTBC as a professional pathogen have also been gained through comparisons with *M. canettii* and other smooth tubercle bacilli (STB), which are phylogenetically the closest relatives of the MTBC[28] (FIG. 1a). Less than 100 STB strains have been isolated so far, all from individuals with TB that have links to the Horn of Africa, particularly Djibouti. No human-to-human transmission of STB has been documented to date, and little is known about the epidemiology of STB, except that a large proportion of strains were isolated from European expatriates living in Djibouti, among whom children seem to be at increased risk of active TB caused by STB[29]. Taken together, these epidemiological findings are consistent with an opportunistic pathogen living in the environment[30]. The genomes of STB are 10–115 kb larger and much more variable in nucleotide sequence diversity than the MTBC genome[31]. For example, MTBC strains differ at most by ~2,000 SNPs from each other[32], whereas STB strains can differ by up to ~65,000 SNPs[31]. Importantly, some of the variation in STB is linked to ongoing HGT[28,31]. Hence, similar to other mycobacteria[33] but in contrast to the MTBC (see below), HGT has an important role in the ongoing evolution of the STB. In particular, distributive conjugal transfer, a recently reported mechanism of HGT resulting in transconjugants with a mosaic genomic structure similar to the outcome of eukaryotic meiosis[34], was recently demonstrated in STB[35,36]. Further genomic comparisons between the MTBC, STB and NTMs revealed that *cobF*, a gene predicted to be involved in cobalamin (vitamin B_{12}) synthesis, is absent from the MTBC but present in STB, *M. marinum* and *M. kansasii*[22], indicating that the MTBC might not be able to synthesize vitamin B_{12} as many other mycobacteria do, but instead has to scavenge it from its host[37]. Conversely, *pe_pgrs33*, a gene encoding an exported protein with strong immunomodulatory effects[38] is present in the MTBC but is lacking in STB, *M. marinum* and *M. kansasii,* thus representing one example of a genomic locus that is associated with the pathogenic lifestyle of the MTBC[22,31]. A recent study suggests that

Slow-growers
Mycobacteria that form colonies in more than 7 days.

PhoPR two-component system
Mycobacterial transcription factors involved in *Mycobacterium tuberculosis* complex virulence.

DosR/S/T regulon
A set of mycobacterial genes involved in latent tuberculosis infection.

mce-associated genes
Mycobacterial genes originally identified as being involved in macrophage entry.

ESAT6 secretion
(ESX). A protein secretion apparatus that, in the case of the *Mycobacterium tuberculosis* complex, exports many virulence determinants.

Toxin–antitoxin system genes
Regulatory systems comprised of two linked genes, one encoding the toxin and the other encoding the neutralizing antitoxin.

Smooth tubercle bacilli
(STB). Organisms that produce smooth colonies on agar plates, which is in contrast to the *Mycobacterium tuberculosis* complex, which produces rough colonies.

Distributive conjugal transfer
A phage-dependent mechanism of DNA transfer between bacteria.

Transconjugants
Bacterial variants that have incorporated DNA from other bacteria through conjugation.

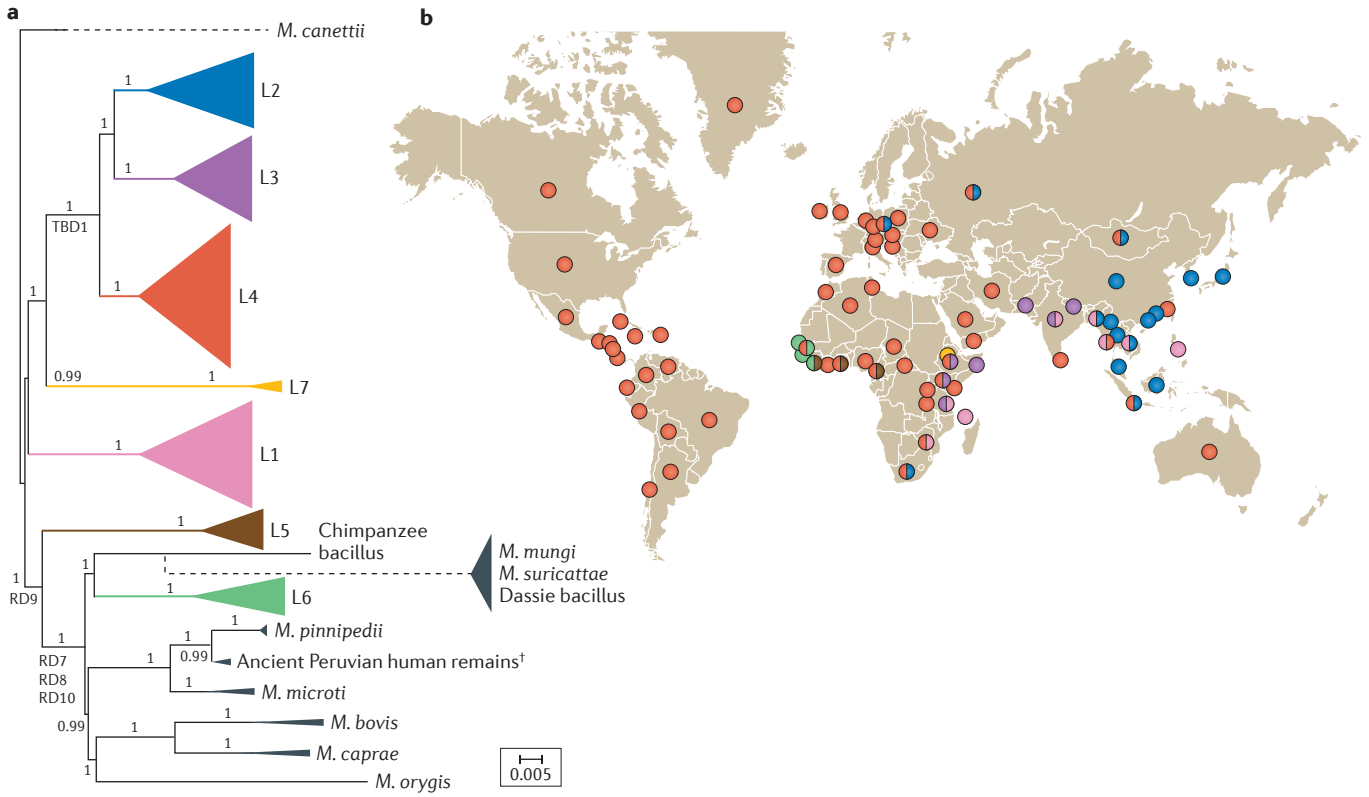


Figure 1 | **Global phylogeography of the human-adapted MTBC. a** | Genome-based phylogeny of the *Mycobacterium tuberculosis* complex (MTBC) rooted with *Mycobacterium canettii* based on previously published data[65]. The MTBC comprises seven human-adapted lineages (in colour) and several lineages adapted to various wild and domestic animals (in grey). Branches of the main lineages are collapsed to improve clarity (indicated by triangles). *M. tuberculosis*-specific deletion 1 (TBD1) indicates that all lineage 2 (L2), L3 and L4 strains share this genomic deletion[57]. Similarly, the deletion of the region of difference 7 (RD7), RD8, RD9 and RD10 is indicated under the respective branches. The grey dotted line leading to *Mycobacterium mungi*, *Mycobacterium suricattae* and the dassie bacillus was added based on previously published genotypic data[48,59] and indicates the most likely phylogenetic relationship of these animal-adapted ecotypes with the other members of the MTBC. The phylogenetic position of these animal-adapted MTBC ecotypes has recently been confirmed on the basis of whole genomes (unpublished observations, S.G.). The dagger indicates genomes generated from ~1,000-year-old MTBC DNA that was recovered from archaeological human remains in Peru[65]. Bootstrap confidence intervals are indicated. Scale bar represents number of nucleotide substitutions per site. **b** | The global distribution of the seven main human-adapted MTBC lineages. Part **a** adapted with permission from REF. 137, Springer.

recombination between two copies of the polyketide synthase *pks5* gene led to a remodelling of the mycobacterial cell surface that is associated with an increase in virulence in the common ancestor of the MTBC[39]. Together, these comparative studies show that in contrast to many other bacterial pathogens in which pathogenicity can be linked to a few chromosomal islands or virulence plasmids, in the MTBC, the determinants of virulence and transmission are more complex and a composite of many interacting factors. Moreover, these bacterial factors also interact with many host factors[27], resulting in variable infection and disease outcomes[40], as well as diverse host preferences among the different members of the MTBC.

Host tropism and adaptation

Following the establishment of the MTBC ancestor as a professional pathogen, the MTBC evolved further into several human-adapted lineages and lineages that are adapted to various wild and domestic mammals (FIG. 1a). Adaptation here refers to the capacity of the pathogen to successfully infect, cause disease and transmit within the primary (or maintenance) host species[11]. This is in contrast to occasional spillover events into other host species in which the full life cycle, in particular the transmission to secondary hosts, is not maintained. This distinction in maintenance versus spillover hosts is important when discussing the host tropism of the different MTBC members. Human TB is mainly caused by *M. tuberculosis sensu stricto* and *Mycobacterium africanum*[41]. These human-adapted members of the MTBC can be further divided into seven phylogenetic lineages: L1, L2, L3, L4 and L7 belong to *M. tuberculosis sensu stricto*, whereas L5 and L6 are traditionally known as *M. africanum* West Africa 1 and 2, respectively (FIG. 1a). The human-adapted MTBC exhibits a strong phylogeographical population structure, with some lineages occurring globally and others showing a strong geographical restriction[32] (FIG. 1b). L2 and L4 are the most widespread globally, with L2 dominating in East Asia. L1 and L3 occur mainly in

Spillover events
The occasional transfer of a particular *Mycobacterium tuberculosis* complex variant from its primary host species into another host species.

regions around the Indian Ocean. L5 and L6 are highly restricted to West Africa, whereas L7 is almost exclusively found in Ethiopia. Various factors are thought to have contributed to the current phylogeographical distribution of human-adapted MTBC members. These are discussed below. Among the animal-adapted members of the MTBC, some have been known to exist for a long time, whereas others have been discovered only recently. As a result, our understanding of the ecology, host preference and evolution of the animal-adapted members of the MTBC, as well as their relationship with the human-adapted MTBC, continues to change.

The animal-adapted members of the MTBC. In 1896, 14 years after the discovery of the human tubercle bacillus by Robert Koch, Theobald Smith showed that the causative agent of TB in animals was distinct from the human bacillus, leading to the first description of *Mycobacterium bovis*[42]. Although primarily a pathogen of cattle, *M. bovis* has been isolated from many mammalian species, including humans. Yet, how many of these species represent spillover hosts is unclear. For example, even though *M. bovis* and *Mycobacterium caprae* (an MTBC member that is adapted to goats and sheep) together cause 1–3% of human TB depending on the geographical region, they rarely transmit from human to human[43]. Wild badgers, on the other hand, are highly susceptible to *M. bovis* and thus represent an important reservoir that complicates the control of bovine TB in countries such as the UK and Ireland[44]. Other animal-adapted members of the MTBC include *Mycobacterium microti* (voles), *Mycobacterium pinnipedii* (seal and sea lions), *Mycobacterium orygis* (antelopes), *Mycobacterium mungi* (banded mongooses), *Mycobacterium suricattae* (meerkats), the dassie bacillus (hyrax) and the chimpanzee bacillus. Some of these microorganisms were given species names according to the first animal they were isolated from. However, given the close phylogenetic relationship between these lineages (FIG. 1a), they might be better considered as ecotypes within the MTBC[45]. Moreover, the actual host ranges of some of these ecotypes remain poorly defined, partially because of the low numbers of isolates that have been studied to date. Similarly, little is known of the molecular mechanisms that determine host range in the MTBC. One striking feature in this respect is the deletion of the genomic region of difference 1 (RD1) that independently occurred in the genomes of *M. microti*, *M. mungi* and the dassie bacillus[46–48]. RD1 encodes the ESX1 secretion system and is required for full virulence in *M. tuberculosis sensu stricto*[46]. The deletion of RD1 was first described in BCG[49], where it is known to have contributed to the attenuation of this live vaccine during its decade-long propagation in vitro[46].

It has been hypothesized that the convergent deletion of RD1 and the accompanying loss of virulence might have contributed to adaptation in the case of *M. microti*, *M. mungi* and the dassie bacillus, as these bacteria primarily infect burrowing host species that live in crowded settings with poor ventilation where the potential for transmission is high[50]. Another study explored the phenotypic consequences of a mutation in the two-component regulatory system PhoPR, which is shared by all animal-adapted strains and the human-adapted MTBC L6 (FIG. 1a). The authors found that this mutation leads to a downregulation of the PhoP regulon, resulting in the loss of bioactive lipids that are involved in virulence and a reduction in the secretion of ESAT6 (REF. 51). However, how these phenotypic differences contribute to the low fitness of *M. bovis* in humans remains unclear. Similarly, experimental infections of *M. tuberculosis sensu stricto* in cattle revealed that the human TB bacillus is highly attenuated in this host species compared with *M. bovis*[52], which is consistent with the sporadic nature of reverse zoonotic transmission of *M. tuberculosis sensu stricto* to animals[53], but the molecular basis for this host preference remains unknown.

The origin of the human-adapted MTBC

As *M. bovis* has a broader host range than *M. tuberculosis sensu stricto*, humans were originally thought to have acquired TB from domestic animals during the Neolithic period[54]. Later, the availability of the first full bacterial genome sequences revealed that *M. bovis* had a smaller genome than *M. tuberculosis,* indicating that *M. tuberculosis* was unlikely to have evolved from *M. bovis*[55,56]. Several studies showed that the order of gene loss across the various members of the MTBC supported an evolutionary scenario in which humans transmitted TB to animals[57,58]. Recent findings based on whole-genome sequences support this scenario as the human-adapted members of the MTBC occupy a more ancestral phylogenetic position than the animal-adapted forms (FIG. 1a). By contrast, all animal-adapted ecotypes were initially thought to form a monophyletic group defined by deletions in RD7, RD8 and RD10 (REFS. 22,57,58); however, more recent data indicate that there might be two separate phylogenetic branches leading to the animal-adapted MTBC[48,59] (FIG. 1a). One of these gave rise to the classical animal-adapted ecotypes, including *M. bovis*, *M. caprae*, *M. orygis*, *M. pinnipedii* and *M. microti*. The other branch includes *M. mungi*, *M. suricattae*, the chimpanzee bacillus and the dassie bacillus and shares a common ancestor with MTBC L6. L6 causes up to 50% of human TB in some West African countries[60], but because of the close phylogenetic relationship with the animal-adapted MTBC, an animal reservoir for L6 has been hypothesized[61]; however, no such reservoir has yet been discovered, indicating that L6 might be primarily human adapted. Hence, considering the phylogenetic position of L6, multiple host jumps might have occurred in the evolution of the RD7–RD8–R10-deleted clade, both from humans to animals and back[41] (FIG. 1a).

The origin of the human-adapted MTBC in space and time. Most evidence to date supports an African origin for the human-adapted MTBC. Specifically, the STB, the closest living relatives of the common ancestor of the MTBC, are almost exclusively found in the Horn of Africa[28,31]. Africa is also the only continent that harbours all seven human-adapted MTBC lineages, including L5,

Ecotypes
An alternative classification of bacterial genotypes that incorporates ecological characteristics.

Box 1 | **Immune subversion**

The *Mycobacterium tuberculosis* complex (MTBC) is thought to have evolved from a *Mycobacterium canettii*-like ancestor into an obligate pathogen[31]. Unlike many other pathogens where transmission is not directly linked to virulence[12], the human-adapted MTBC has to cause pulmonary disease to transmit between individuals[11]. Tuberculosis (TB) is mainly a consequence of the host inflammatory immune responses to antigenic load, which lead to destruction of the lung tissue and the formation of cavities, thereby dramatically increasing the transmission potential of the pathogen[27]. Recent studies have shown that, in contrast to many other pathogens that evade host immunity by antigenic variation, the large majority of known human T cell epitopes in the MTBC are hyperconserved[85–88], including those encoded by *pe_pgrs* genes that are otherwise highly variable[130]. This suggests that the MTBC has adopted a strategy of immune subversion and that the human immune responses that are elicited by the MTBC benefit the pathogen by increasing the transmission potential. This notion is supported indirectly by the observation that in HIV–TB co-infected individuals, the frequency of lung cavitation is inversely correlated with the number of circulating $CD4^+$ T cells[131]. In other words, HIV–TB co-infected individuals with low $CD4^+$ T cell counts have on average fewer lung cavities, and these individuals are therefore expected to transmit less than TB patients who are not infected with HIV (BOX 2). These observations have important implications for TB vaccine design, as standard vaccinology tends to target antigens that are highly conserved among strains of a pathogen[132]. Owing to the recent clinical failure of the most advanced new TB vaccine candidate[133], an alternative strategy that focuses on vaccine antigens that are variable (that is, under diversifying selection) among MTBC clinical strains might be more promising, as the immune responses that are directed to these variable antigens might be more deleterious to the pathogen[87]. An alternative explanation for the high conservation of T cell epitopes in the MTBC is that these regions encode proteins that are essential for virulence and thus cannot be mutated without a fitness loss in vivo[31]. However, genes that are essential for mycobacterial growth are not generally over-represented in T cell epitopes[87], and the observation that epitope regions of T cell antigens are more conserved than other regions in the same antigens supports a role for the host immune response in the hyperconservation of T cell epitopes in the human-adapted MTBC[85,87].

L6 and L7, which are restricted to the African continent (FIG. 1b). Thus, Africa harbours the largest diversity of human-adapted MTBC, and this diversity decreases with increasing distance from Africa[62]. Similarly, various phylogeographical analyses of MTBC whole-genome sequences indicate Africa as the most likely region of origin[63], and most animal-adapted MTBC members that infect wild animals are found in Africa (FIG. 1a). Unlike its geographical origin, there is currently no consensus on the age of the most recent common ancestor of the MTBC. Different approaches have been used to date MTBC phylogenies, and these have yielded different results. One study built on the hypothesis of co-divergence between the MTBC and *Homo sapiens* since the first out-of-Africa migrations of modern humans and estimated that the most recent common ancestor of the MTBC existed 70,000 years ago[63]. Two other studies have used ancient MTBC DNA extracted from ~200-year-old mummies from Hungary and ~1,000-year-old human remains from Peru, respectively, to calibrate the MTBC phylogeny and obtained an estimate of less than 6,000 years ago[64,65].

These younger age estimates are incompatible with evidence from ancient MTBC DNA and cell wall lipids associated with TB found in ~11,000-year-old[66] and ~9,000-year-old[67] human remains in Syria and Israel, respectively, as well as from remains of a bison in Wyoming dated as ~17,000 years old[68]. However, the authenticity of some of these previous findings has been debated[65]. More generally, it is unclear whether mutation rates obtained from epidemiological studies of contemporary MTBC samples[69] or relatively recent ancient DNA[64,65] can be extrapolated to substitution rates over longer periods of time[70], particularly because not all MTBC populations show a clear molecular clock-like accumulation of sequence diversity over time in their genomes[71] and because of our limited understanding of the effects of latency in the molecular evolution of these organisms[72]. To be able to explore the evolutionary history of the MTBC over millennia, analyses of more ancient MTBC genomes will be necessary to establish a robust long-term substitution rate for the MTBC.

The ecology of the human-adapted MTBC

The observation that some human-adapted MTBC lineages are geographically restricted (FIG. 1b) has led to the hypothesis that these variants might be adapted to the local human host populations[73]. Local adaptation refers to the phenomenon in which a pathogen that is adapted to one host species has a reduced capacity to spread among other host species[74]. As the human-adapted MTBC is an obligate pathogen that manipulates the human immune system to promote its replication and spread[27] (BOX 1), local adaptation to the immunological characteristics of its specific host population is expected[74]. Several features in the epidemiology of human TB are consistent with local adaptation, notwithstanding the possible confounding effects of social factors. Specifically, the strong sympatric host–pathogen associations in TB (FIG. 1b) are maintained in metropolitan cities[75–77], and MTBC strains preferentially transmit among their sympatric patient populations in these settings[78,79]. Moreover, these sympatric host–pathogen associations are perturbed by HIV co-infection, as HIV co-infected TB patients are more likely to be infected with allopatric MTBC strains, and the likelihood of allopatric TB increases with increasing immune deficiency[78,79] (BOX 2). Evidence indicating local adaptation has also been reported in Ghana; two separate studies found that MTBC L5 independently associated with a particular patient ethnicity[80,81].

Generalists and specialists. From an ecological point of view, locally restricted MTBC genotypes represent specialists with a narrow niche that corresponds to their specific host population[82]. Conversely, broadly distributed MTBC lineages (FIG. 1) are generalists capable of occupying many different ecological niches. MTBC L4 is the most widespread cause of human TB globally and thus can be viewed as a generalist (FIG. 1b). However, two recent whole-genome studies have shown that L4 can be further divided into multiple sublineages[83,84] (FIG. 2a). Subsequent genotyping of 3,366 L4 clinical isolates from 100 countries revealed that some of these sublineages occurred globally, indicating generalists, whereas others were geographically restricted specialists[84] (FIG. 2b). Intriguingly, even though most of the known human T cell epitopes of the MTBC are conserved[85–88] (BOX 1), the L4 generalist sublineages were found to have a larger proportion of variable epitopes than the specialists,

Sympatric
Host and pathogen variants that co-occur in a given geographical setting.

Allopatric
Host and pathogen variants that usually occur in geographically separate settings.

T cell epitopes
Parts of the *Mycobacterium tuberculosis* complex proteome (that is, peptides) that are recognized by T lymphocytes.

Founder effects
The random introduction of a particular bacterial variant into a given setting.

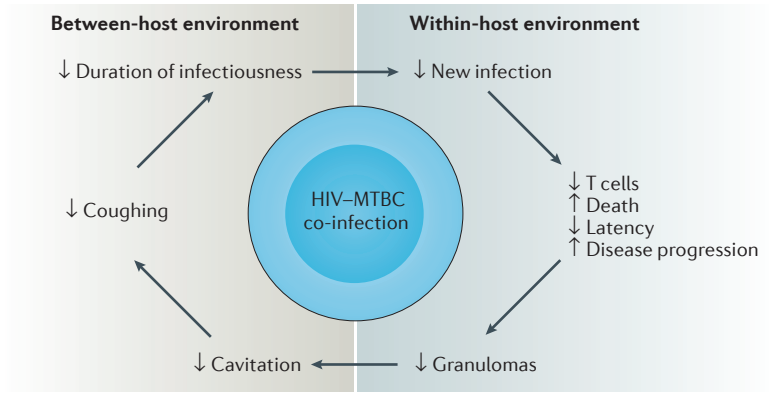
Homoplasies
Characters acquired independently by two or more bacterial variants that do not share an immediate common ancestor.

possibly indicating interaction with a diverse host population, immune escape mechanisms or a combination of both[84]. An alternative but not mutually exclusive explanation for the geographical restriction of specific MTBC genotypes might be founder effects combined with limited population mobility since the initial introduction of a bacterial variant to a given population. Similarly, specialist MTBC genotypes might have emerged more recently than generalists and might thus have had insufficient time to spread globally. Functional studies will be necessary to differentiate between the biological and sociological factors that determine the spread of the various human-adapted MTBC lineages and sublineages.

Virulence evolution in the human-adapted MTBC. An important feature of human TB is latency, during which the TB bacilli are controlled by the host immune system until it is weakened by factors such as old age, malnutrition or co-infection with HIV (BOX 2; FIG. 3), leading to disease reactivation and onward transmission[40]. As latency can last for decades, it was hypothesized to reflect MTBC adaptation to small hunter–gatherer host populations, allowing the pathogen to jump generations and access new birth cohorts of susceptible hosts[89]. In such a scenario, TB would have exerted little selective pressure on humans if active TB occurred mainly later in life, that is, after reproduction. Furthermore, ecological theory predicts that a growing host population density (for example, that linked to urbanization) selects for increased pathogen virulence and shorter latency in obligate pathogens such as the MTBC[12]. This model is supported by recent mathematical modelling[90] and the reported correlation between the duration of urbanization in a given setting and the frequency of human genetic variants linked to resistance to intracellular pathogens[91]. The various human-adapted MTBC variants are known to differ in virulence, progression to disease and transmission potential[32,92], and one could therefore hypothesize that these phenotypic differences reflect the distinct demographic histories of their respective human host populations[93]. Alternatively, given the large number of latently infected individuals and the comparably small number of TB cases, latent TB infection might be seen as a symbiotic relationship that partially benefits the host, for example, by providing immune stimuli that protect the host from other diseases or by contributing important micronutrients[94].

Box 2 | **Impact of HIV on the evolution of the MTBC**

HIV co-infection is a strong and well-known risk factor for the development of active tuberculosis (TB)[131]. However, if and how HIV co-infection influences the evolution of the *Mycobacterium tuberculosis* complex (MTBC) within and between patients remains largely unexplored[11]. To date, only three studies have been published on the subject. One study from Argentina found that HIV co-infection had no effect on the mutation rate of the MTBC and was also not associated with faster de novo development of drug resistance in individuals[134]. Moreover, HIV co-infection did not affect the transmission potential of these individuals. However, most of these individuals were on antiretroviral treatment, and because $CD4^+$ T cell counts were not available, the degree of immune deficiency could not be controlled for in this analysis[131]. A second study from South Africa found that HIV co-infection had no effect on the overall population structure of the MTBC circulating in the area, but HIV exerted weak selection on human MTBC T cell epitopes[135] (BOX 1). In a third study, an *M. tuberculosis* strain was serially passaged in immune-competent and T cell-deficient mice, with the latter mimicking the environment within severely immune-supressed patients[136]. The authors found overall strong evidence for purifying selection of the MTBC, supporting the notion that the MTBC is highly adapted to live inside mammalian hosts. Moreover, the bacterial populations accumulated slightly more mutations in the presence of T cells, suggesting a role of T cell immunity in driving MTBC sequence diversity. HIV co-infection affects many aspects of the life cycle of the MTBC both within and between individuals. These effects are also a function of the level of immune suppression, as measured by the number of circulating $CD4^+$ T cells in individuals with HIV–TB co-infection (see the figure). Some of the most important effects of HIV co-infection are an accelerated progression to disease (that is, reduced latency); a shorter period of infectiousness, as HIV–TB co-infected individuals tend to have a lower life expectancy than individuals who have TB but are not infected with HIV, particularly in the absence of antiretroviral treatment; a higher proportion of extrapulmonary disease; and reduced formation of lung cavities[131], all of which lead to a reduced potential for TB transmission.



Consequences of strict clonality

Ongoing HGT is negligible in the MTBC. There has been a debate as to whether HGT is still occurring between clinical strains of the MTBC or whether this phenomenon is limited to STB and other mycobacteria. In addition to being of academic interest, this issue has important implications for our understanding of antibiotic resistance. Most of the evidence to date supports the view that ongoing HGT does not occur at detectable levels in the MTBC. Specifically, an analysis of minisatellites distributed around the MTBC genome revealed strong linkage disequilibrium (that is, loci carrying particular minisatellite variants in a given bacterial strain are linked and will not be disrupted by recombination)[95]. Moreover, genomic deletions that were used to classify MTBC genotypes show no homoplasies[57,58,75,78]; phylogenies that were generated using different molecular markers are congruent[63,93,96,97], and homoplasies at single nucleotide sites are extremely rare, except in the case of convergent evolution of mutations related to antibiotic resistance[97–99]. Yet, one study has challenged this notion and suggested that strains of the MTBC frequently exchange small DNA fragments, but because of the limited sequence variation, these events would remain unnoticed[100]. Most recently, a co-culturing approach provided strong experimental evidence for the capacity of STB to undergo HGT, which was, however, not observed in the MTBC[36]. In summary, the available data strongly support a clonal population structure for the MTBC with little ongoing HGT between strains[101].

The balance between natural selection and genetic drift. Strict clonality, combined with the obligatory pathogenic lifestyle, has important consequences for the molecular evolution of the MTBC, in particular by influencing the balance between natural selection and selectively neutral mechanisms such as random genetic drift[61,102]. In the MTBC, selective sweeps and background selection are mechanisms linked to the action of natural selection, which together lead to a reduction in genetic diversity. In a selective sweep, a positively selected locus (for example, a drug-resistance-conferring mutation) reaches fixation, and because of strong linkage, the entire chromosome achieves fixation by association, leading to a reduction in diversity[103]. A striking example of this phenomenon was recently reported in an individual who developed XDR-TB during treatment[104]. Background selection refers to the converse phenomenon, where a deleterious mutation is selected against, leading to the removal of all chromosomal variation linked to this locus[105]. One of the features of MTBC population genetics is the high proportion of low-frequency genetic variants, particularly singletons (that is, mutations that occur in only one strain)[86,93,106,107]. This phenomenon has been proposed to reflect the effect of background selection[86,106]. Alternatively, a high proportion of low-frequency mutations is also compatible with recent population expansion, similar to what has been observed in human populations[108]. Changes in human demography are expected to be reflected in the population structure of the human-adapted MTBC[41], and studies have detected signals of population expansion in the MTBC that coincide with human population increases during the Neolithic age and the Industrial Revolution[63,86,109,110]. Importantly, however, the bias towards low-frequency bacterial variants is also observed within individuals with TB[107], suggesting that purifying and background selection might be common features of MTBC evolution both within and between patients. Further examples of purifying selection acting within individuals have recently been reported in a study of MTBC strains that were isolated from individuals who died prior to the initiation of treatment[111]. Another study followed the population dynamics of the MTBC during treatment and found a stronger signal of purifying selection in individuals receiving effective drugs than in patients receiving suboptimal antibiotic treatment[107]. In summary, purifying selection seems to be an important feature of MTBC evolution, and most new mutations arising in these bacteria are likely to be deleterious, likely reflecting the fact that the human-adapted MTBC is already well adapted to live in and spread between humans[41] (FIG. 3a).

Population bottlenecks, structured populations and the effect of genetic drift. In addition to purifying selection, population bottlenecks can also lead to a reduction in diversity, as most bacterial variation emerging de novo within patients is lost each time MTBC transmits to a new host (FIG. 3b). There are experimental data showing that a new TB infection can be established by a single bacterial colony-forming unit[112], although the actual infectious dose in human TB remains unknown. Moreover, MTBC populations are divided into subpopulations: within individuals[111], between patients and geographically (FIG. 1b). Transmission bottlenecks, population substructuring and clonality all lead to a reduction in the effective population size[113]. An important consequence of a small population size is that natural selection is less efficient and the effect of random genetic drift is comparably increased. The notion that genetic drift has an evolutionary role in the MTBC is supported by the observation that, in contrast to STB and NTMs, about two-thirds of SNPs are non-synonymous, including those fixed in different MTBC lineages, and a large proportion of these are predicted to affect gene function[26,93].

In summary, the available data indicate that strict clonality in the MTBC leads to a reduction in diversity through the combined effects of selective sweeps, purifying selection and background selection as well as transmission bottlenecks. While the effect of purifying selection is more readily detected (BOX 1), the role of genetic drift in the molecular evolution of the MTBC needs to be explored further. An improved understanding of the role of these evolutionary forces is required to better predict the impact of future interventions into the evolutionary trajectory of the MTBC.

Drug resistance

An illustrative example of positive selection acting on the MTBC is the emergence of MDR-TB and XDR-TB[4]. The mutation rate of the MTBC in clinical settings has been estimated at 0.3–0.5 substitutions per genome per year, which is orders of magnitude lower than in most other clinically relevant bacteria, likely reflecting the long generation time of TB bacilli in vivo[114]. Yet, despite this low mutation rate and the lack of ongoing HGT and resistance plasmids, antibiotic resistance in the MTBC can arise quickly[104,115] (FIG. 3d). This is because the frequency of resistance-conferring mutations is a function of the inherent mutation rate in addition to the mutational target size for resistance, the fitness effects of these resistance mutations and the bacterial population size within individuals[116]. The bacterial population size can be very large, particularly in individuals who are failing to respond to treatment. Mutational target sizes are related to the mechanism of resistance and thus

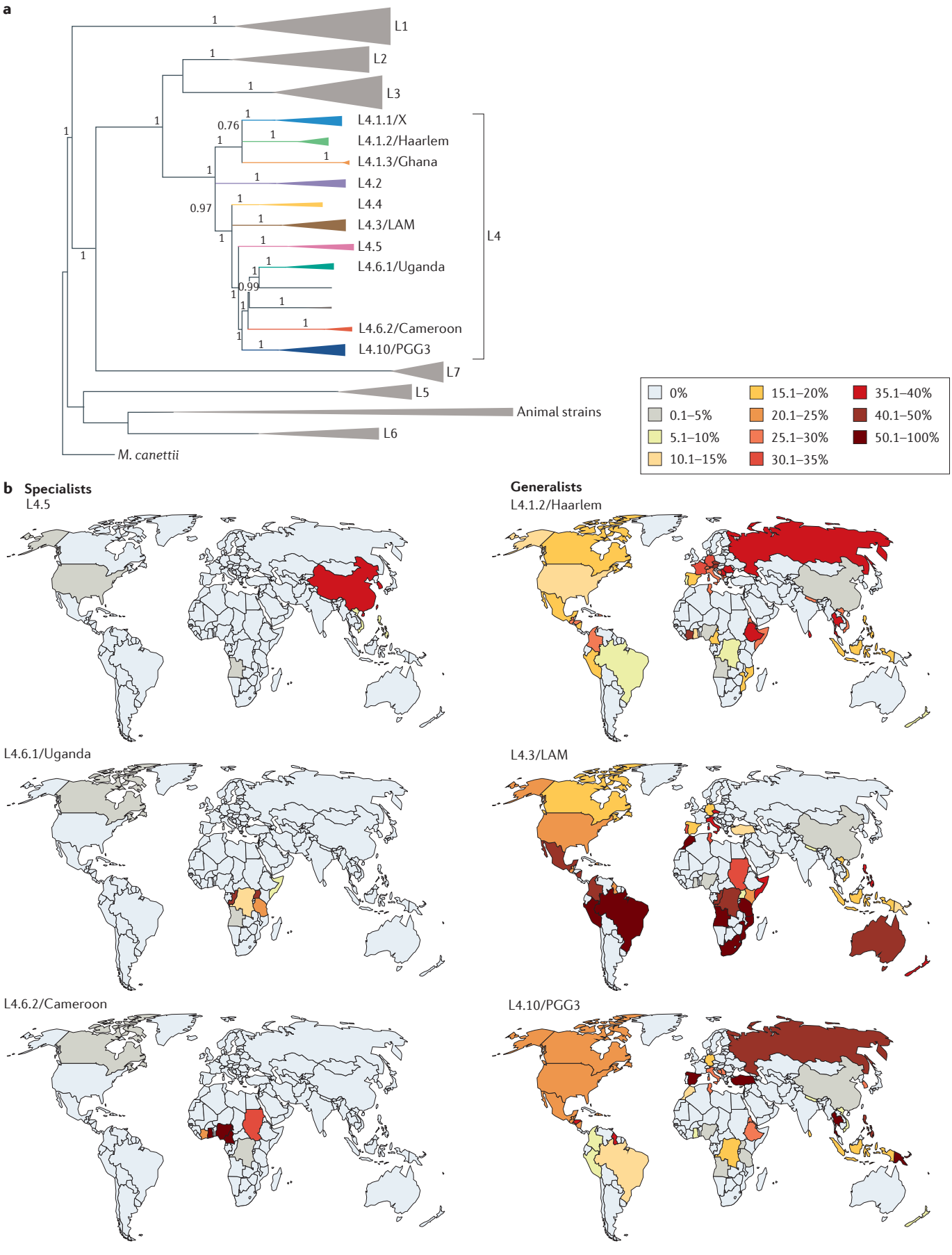
Selective sweeps
Positive selection that leads to the fixation of a new beneficial mutation.

Background selection
Selection against a deleterious mutation that leads to the elimination of any mutation linked to the target of selection.

Purifying selection
Selection against detrimental mutations.

Transmission bottlenecks
A type of population bottleneck in which only a subset of the bacterial diversity present in one host is transmitted to the next.

Figure 2 | **MTBC L4 can be separated into specialists and generalists.** *Mycobacterium tuberculosis* complex (MTBC) lineage 4 (L4) is geographically the most widespread cause of human tuberculosis (FIG. 1b). **a** | Genome-based phylogeny showing that L4 can be further subdivided into at least ten sublineages. Branches of the main lineages and sublineages are collapsed to improve clarity (indicated by triangles). Bootstrap confidence intervals are indicated. **b** | Genotypic screening of 3,366 L4 isolates from 100 countries revealed that some of these L4 sublineages are geographically restricted, corresponding to ecological specialists, and others are globally distributed (generalists). Percentages correspond to the proportion of the respective L4 sublineage in a given country. LAM, Latin American-Mediterranean. PGG3, principle genotypic group 3. Figure adapted from REF. 84, Macmillan Publishers Limited.



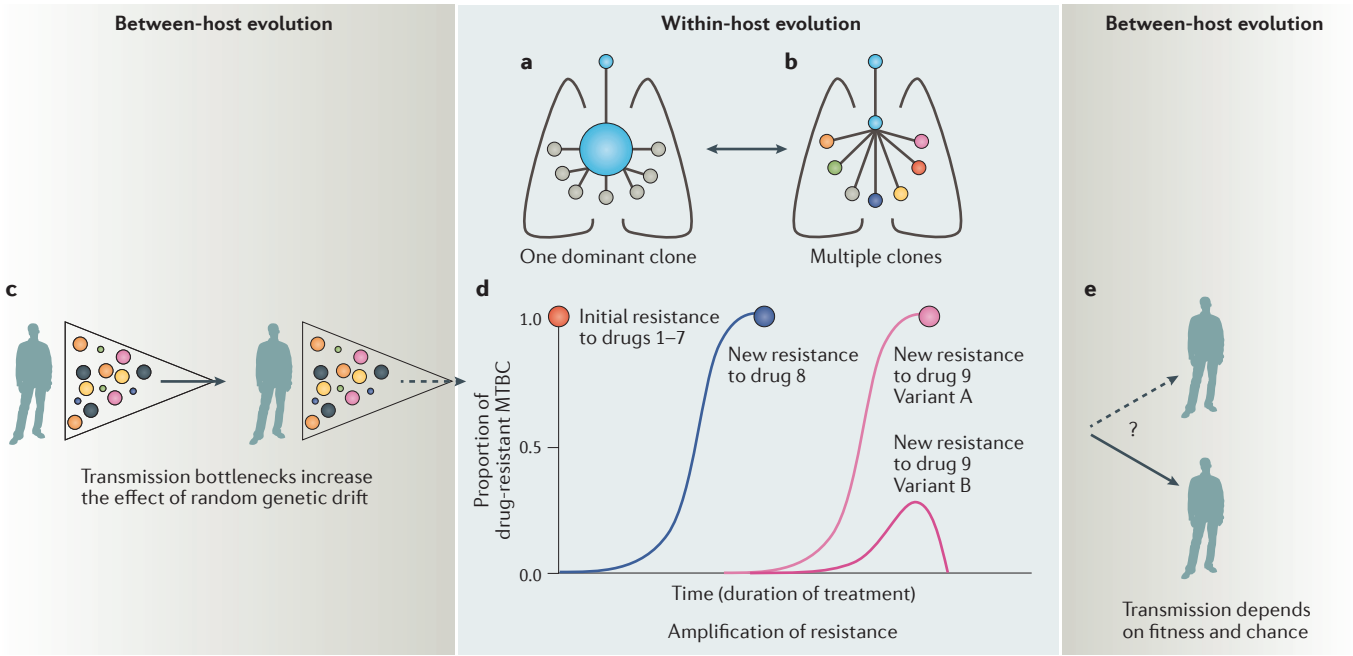


Figure 3 | **The role of natural selection and genetic drift in the evolution of the MTBC and the emergence of drug resistance. a** | The within-host evolution of the *Mycobacterium tuberculosis* complex (MTBC) is characterized by purifying selection as most mutations that emerge de novo in individual patients are lost[107]. This is illustrated here by the large blue dot that represents the dominant bacterial clone, from which new variants emerge that are being removed by purifying selection (small grey dots). This reflects the fact that the human-adapted MTBC is already well adapted to its human host, and most new mutations are thus likely to be deleterious. **b** | This is in contrast to a scenario in which opportunistic pathogens such as *Pseudomonas aeruginosa* or *Burkholderia dolosa* in individuals with cystic fibrosis develop many de novo host adaptations in bacterial subpopulations that evolve separately in individual patients over years (many small different coloured dots). **c** | As MTBC transmits between individuals, it undergoes a population bottleneck at each transmission event. This leads to a reduction in bacterial diversity, a reduction in the effective population size and an increase in the effect of random genetic drift. This is illustrated here with differently coloured dots that represent bacterial variants emerging through micro-evolution within patients. The difference in the size of these dots illustrates differences in fitness, with smaller dots having a reduced fitness compared with that of larger dots. Increased genetic drift during transmission could lead to chance events in which a low-fitness variant gets transmitted (the solid arrow in the first transmission event). Depending on the environmental conditions, for example, in the presence of antibiotics, a fitter variant can be transmitted (the dashed arrow in the second transmission event). **d** | This graph illustrates the gain of additional resistance determinants during patient treatment, which is also referred to as amplification of drug resistance. When individuals infected with MTBC are under ineffective treatment, they are more likely to acquire additional resistance mutations[107]. In this example, an individual with TB was originally infected with an MTBC strain that was already resistant to seven drugs. The red dot indicates that 100% sequencing reads detected resistance alleles for these seven drugs at the time of diagnosis[115]. Thus, highly drug-resistant strains can also be transmitted, either by chance or because of increased fitness as discussed in part **c**. During treatment of this single patient, which lasted a total of 5 years, the bacteria acquired resistance to several additional drugs, including the two new anti-TB drug candidates bedaquiline and delamanid (illustrated as drug 8 in blue and drug 9 in pink). Some of these new drug-resistant variants reached fixation, corresponding to a selective sweep[102]. Multiple independent drug-resistant subpopulations can emerge during the process, with some taking over and others becoming lost again over time (resistance to drug 9 variant A versus variant B). **e** | The likelihood of successful transmission of the bacterial variants with amplified resistance will depend on the inherent fitness of these variants (FIG. 4) and on chance, that is, genetic drift (part **c**). This is illustrated here with a dotted arrow (high fitness) and a solid arrow (low fitness) in line with part **c**. In summary, the emergence, amplification and transmission of drug resistance in TB is governed by multiple selective and neutral evolutionary forces acting concomitantly on the MTBC both within and between patients. However, little is known on the relative importance of these forces in different epidemiological settings (indicated by the question mark).

vary depending on the antibiotic that is administered[116]. These mechanisms of resistance and the various health system factors that drive TB drug resistance in clinical settings are generally well understood (for comprehensive reviews see REFS 116,117). By contrast, the ecological and evolutionary aspects of drug resistance are only gradually being elucidated. Studies in the 1950s showed that some drug-resistant MTBC clinical strains were attenuated in animal models[118]. Hence, in the past, when a universal fitness cost was assumed, the problem of drug-resistant TB was primarily attributed to the de novo development of resistance linked to patient non-adherence and other problems inherent to poor TB control[5]. However, it has now become clear that the global epidemics of MDR-TB and XDR-TB are driven by a combination of de novo development and direct

transmission of drug-resistant strains[119] (FIG. 3c–e). Even though drug-resistance-conferring mutations in the MTBC are often associated with a fitness cost in the absence of the drug, some mutations show little or no cost, and these tend to be preferentially selected for in clinical settings[120,121]. Moreover, compensatory evolution can overcome initial fitness deficits that are linked to particular resistance mutations[98,122] (FIG. 4). For example, secondary mutations in various subunits of the mycobacterial RNA polymerase (RNAP) restore transcriptional activity in strains carrying rifampicin-resistance-conferring mutations in the RNAP genes[123] and are associated with increased transmission and acquisition of additional resistance[98,124]. Moreover, epistatic interactions between mutations causing resistance to different drugs can lead to a reduction in the fitness cost that is associated with each individual mutation[125]. Finally, the genetic background of a strain can also influence the pathway to resistance (FIG. 4), for example, by modulating the level of resistance that is caused by a particular mutation[126]. The Beijing family of strains, which is part of MTBC L2 (FIG. 1a), has repeatedly been associated with multidrug resistance for reasons that remain unclear[127]. A higher inherent mutation rate has been suggested, but experimental studies addressing this hypothesis have produced conflicting results[128,129]. A recent study revealed that the ongoing MDR-TB epidemics in many former Soviet republics[2] were caused by just a few highly resistant clones of the Beijing family[110]. The success of these clones might be due to random genetic drift; that is, being in the right place at the right time[69] (FIG. 3b). Alternatively, as some Beijing family strains have been associated with increased virulence and increased transmission[32,92], they might have more opportunities to develop resistance and/or better tolerate the burden of carrying multiple resistance determinants[127]. In summary, epistatic interactions between drug-resistance-conferring mutations, compensatory mutations and the genetic background of the strain affect the biology and epidemiology of drug-resistant TB (FIG. 4). Improving our understanding of these interactions will be crucial for improved surveillance of antibiotic resistance when introducing novel treatment regimens for TB[4].

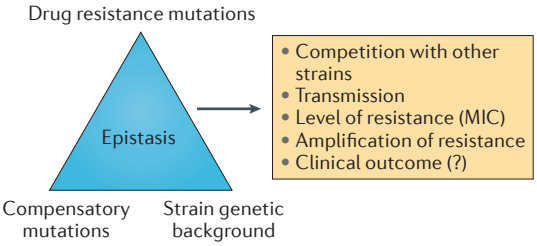


Figure 4 | **The role of epistasis in the evolution of multidrug-resistant tuberculosis.** The epistatic interactions between different drug resistance mutations, fitness compensatory mutations and the genetic background of the strain influence various aspects of drug resistance evolution in tuberculosis. For example, mycobacteria carrying a mutation in DNA-directed RNA polymerase subunit-β (RpoB), which causes resistance to rifampicin, and a fluoroquinolone-resistance-conferring mutation in DNA gyrase subunit A (GyrA) can have a higher competitive fitness than strains carrying only one of these mutations[125]. Strains carrying an RpoB mutation and a compensatory mutation in DNA-directed RNA polymerase subunit-β′ (RpoC) have a higher fitness than strains carrying only the RpoB mutation[98,123,124,138] and are also more likely to acquire additional resistance mutations (that is, amplification of resistance)[124]. The minimum inhibitory concentration (MIC) associated with mutations conferring resistance to isoniazid can differ depending on the specific strain background these mutations arise in[126]. Whether similar epistatic interactions also impact clinical outcome remains to be determined. Figure adapted from Gygli, S. M., Borrell, S., Trauner, A. & Gagneux, S. Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS Microbiol. Rev.* **41**, 354–373 (2017), by permission of Oxford University Press (REF. 116).

Conclusions

Many new insights have been gained during recent years into the origin, ecology and evolution of human TB, but much work remains to be done[8]. We know that the MTBC developed into a professional pathogen from an environmental organism through the acquisition and loss of genes. However, no single genomic feature in the MTBC can account for the obligate pathogenic lifestyle of these organisms or the different host preferences of the various MTBC ecotypes. Comparing host tropisms across these ecotypes might help us to understand the bacterial and host determinants of pathogenicity and transmission, potentially leading to the development of novel antibiotics and vaccines for the control of TB. We also know that the transition to a professional pathogen most likely occurred in Africa, but the age of the MTBC remains unknown. More studies of ancient DNA will help to date the origin of MTBC. The human-adapted MTBC exhibits a specific phylogeographical population structure, which is consistent with an ecological separation into specialists and generalists, but whether these patterns reflect adaptive strategies as opposed to random founder effects or mere social factors remains unclear. Functional studies should be carried out to answer these questions, given their relevance for the design of universally effective vaccines (BOX 1). Current evidence supports a clonal population structure for the MTBC with negligible ongoing HGT. Strict clonality, combined with serial transmission bottlenecks and population subdivisions, affect the balance between natural selection and random genetic drift, but more work is needed to better understand the various evolutionary forces acting within and between patients and how these forces differ from those that act on other pathogens. Finally, epistatic interactions between drug-resistance-conferring mutations, compensatory mutations and the strain genetic background influence the ecology and evolution of antibiotic-resistant TB. These phenomena need to be studied further, as a broader understanding of the role of epistasis in antibiotic resistance will help preserve current and future treatment options against TB and other bacterial infections.

1. Paulson, T. Epidemiology: a mortal foe. *Nature* **502**, S2–S3 (2013).
2. World Health Organization. *Global tuberculosis control — surveillance, planning, financing*. (WHO, Geneva, Switzerland, 2017).
3. Houben, R. M. & Dodd, P. J. The global burden of latent tuberculosis infection: a re-estimation using mathematical modelling. *PLOS Med.* **13**, e1002152 (2016).
4. Dheda, K., Barry, C. E. III & Maartens, G. Tuberculosis. *Lancet* **387**, 1211–1226 (2016).
5. Dye, C., Williams, B. G., Espinal, M. A. & Raviglione, M. C. Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science* **295**, 2042–2046 (2002).
6. Comas, I. & Gagneux, S. The past and future of tuberculosis research. *PLOS Pathog.* **5**, e1000600 (2009).
7. Achtman, M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**, 53–70 (2008).
8. Gagneux, S. *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control* (Springer, Heidelberg, 2017).
9. Fedrizzi, T. et al. Genomic characterization of Nontuberculous Mycobacteria. *Sci. Rep.* **7**, 45258 (2017).
10. Rogall, T., Wolters, J., Flohr, T. & Bottger, E. C. Towards a phylogeny and definition of species at the molecular level within the genus Mycobacterium. *Int. J. Syst. Bacteriol.* **40**, 323–330 (1990).
11. Brites, D. & Gagneux, S. Old and new selective pressures on *Mycobacterium tuberculosis*. *Infect. Genet. Evol.* **12**, 678–685 (2012).
12. Ebert, D. & Bull, J. J. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol.* **11**, 15–20 (2003).
13. Jang, J., Becq, J., Gicquel, B., Deschavanne, P. & Neyrolles, O. Horizontally acquired genomic islands in the tubercle bacilli. *Trends Microbiol.* **16**, 303–308 (2008).
14. VanderVen, B. C., Huang, L., Rohde, K. H. & Russell, D. G. The minimal unit of infection: *Mycobacterium tuberculosis* in the macrophage. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.TBTB2-0025-2016 (2016).
15. Chisholm, R. H., Trauer, J. M., Curnoe, D. & Tanaka, M. M. Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. *Proc. Natl Acad. Sci. USA* **113**, 9051–9056 (2016).
16. Veyrier, F. J., Dufort, A. & Behr, M. A. The rise and fall of the *Mycobacterium tuberculosis* genome. *Trends Microbiol.* **19**, 156–161 (2011).
17. Stinear, T. P. et al. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* **18**, 729–741 (2008).
18. Wang, J. et al. Insights on the emergence of *Mycobacterium tuberculosis* from the analysis of *Mycobacterium kansasii*. *Genome Biol. Evol.* **7**, 856–870 (2015).
19. Becq, J. et al. Contribution of horizontally acquired genomic islands to the evolution of the Tubercle Bacilli. *Mol. Biol. Evol.* **24**, 1861–1871 (2007).
20. Veyrier, F., Pletzer, D., Turenne, C. & Behr, M. A. Phylogenetic detection of horizontal gene transfer during the step-wise genesis of *Mycobacterium tuberculosis*. *BMC Evol. Biol.* **9**, 196 (2009).
21. Reva, O., Korotetskiy, I. & Ilin, A. Role of the horizontal gene exchange in evolution of pathogenic Mycobacteria. *BMC Evol. Biol.* **15** (Suppl. 1), S2 (2015).
22. Boritsch, E. C. et al. A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. *Mol. Microbiol.* **93**, 835–852 (2014).
23. Brennan, M. J. & Delogu, G. The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends Microbiol.* **10**, 246–249 (2002).
24. Gey van Pittius, N. C. et al. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evol. Biol.* **6**, 95 (2006).
25. Sala, A., Bordes, P. & Genevaux, P. Multiple toxin-antitoxin systems in *Mycobacterium tuberculosis*. *Toxins (Basel)* **6**, 1002–1020 (2014).
26. Rose, G. et al. Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. *Genome Biol. Evol.* **5**, 1849–1862 (2013).
27. Ernst, J. D. The immunological life cycle of tuberculosis. *Nat. Rev. Immunol.* **12**, 581–591 (2012).
28. Gutierrez, C. et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLOS Pathog.* **1**, 1–7 (2005).
29. Blouin, Y. et al. Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg. Infect. Dis.* **20**, 21–28 (2014).
30. Koeck, J. L. et al. Clinical characteristics of the smooth tubercle bacilli "*Mycobacterium canettii*" infection suggest the existence of an environmental reservoir. *Clin. Microbiol. Infect.* **17**, 1013–1019 (2010).
31. Supply, P. et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 172–179 (2013).
This is the most detailed genomic analysis of *M. canettii* and other STB to date and provides interesting insights into the evolution of the MTBC.
32. Coscolla, M. & Gagneux, S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* **26**, 431–444 (2014).
33. Smith, S. E. et al. Comparative genomic and phylogenetic approaches to characterize the role of genetic recombination in mycobacterial evolution. *PLOS One* **7**, e50070 (2012).
34. Gray, T. A., Krywy, J. A., Harold, J., Palumbo, M. J. & Derbyshire, K. M. Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. *PLOS Biol.* **11**, e1001602 (2013).
This paper is the first description of the novel mechanism of horizontal gene exchange known as distributive conjugal transfer.
35. Mortimer, T. D. & Pepperell, C. S. Genomic signatures of distributive conjugal transfer among mycobacteria. *Genome Biol. Evol.* **6**, 2489–2500 (2014).
36. Boritsch, E. C. et al. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *Proc. Natl Acad. Sci. USA* **113**, 9876–9881 (2016).
This paper provides experimental evidence for ongoing HGT in *M. canettii* and other STB. By contrast, and despite multiple attempts, no evidence of HGT was detected in the MTBC.
37. Young, D. B., Comas, I. & de Carvalho, L. P. Phylogenetic analysis of vitamin B_{12}-related metabolism in *Mycobacterium tuberculosis*. *Front. Mol. Biosci.* **2**, 6 (2015).
38. Zumbo, A. et al. Functional dissection of protein domains involved in the immunomodulatory properties of PE_PGRS33 of *Mycobacterium tuberculosis*. *Pathog. Dis.* **69**, 232–239 (2013).
39. Boritsch, E. C. et al. pks5-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. *Nat. Microbiol.* **1**, 15019 (2016).
40. Cadena, A. M., Fortune, S. M. & Flynn, J. L. Heterogeneity in tuberculosis. *Nat. Rev. Immunol.* **17**, 691–702 (2017).
41. Brites, D. & Gagneux, S. Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunol. Rev.* **264**, 6–24 (2015).
42. Smith, T. A comparative study of bovine tubercle bacilli and of human bacilli from sputum. *J. Exp. Med.* **3**, 451–511 (1898).
43. Muller, B. et al. Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. *Emerg. Infect. Dis.* **19**, 899–908 (2013).
44. Gormley, E. & Corner, L. A. Control strategies for wildlife tuberculosis in Ireland. *Transbound Emerg. Dis.* **60** (Suppl. 1), 128–135 (2013).
45. Smith, N. H. et al. Ecotypes of the *Mycobacterium tuberculosis* complex. *J. Theor. Biol.* **239**, 220–225 (2005).
46. Pym, A. S., Brodin, P., Brosch, R., Huerre, M. & Cole, S. T. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol. Microbiol.* **46**, 709–717 (2002).
47. Mostowy, S., Cousins, D. & Behr, M. A. Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. *J. Bacteriol.* **186**, 104–109 (2004).
48. Alexander, K. A. et al. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg. Infect. Dis.* **16**, 1296–1299 (2010).
49. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* **178**, 1274–1282 (1996).
50. Behr, M. A. Comparative genomics of mycobacteria: some answers, yet more new questions. *Cold Spring Harb. Perspect. Med.* **5**, a021204 (2015).
51. Gonzalo-Asensio, J. et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc. Natl Acad. Sci. USA* **111**, 11491–11496 (2014).
This is an elegant study exploring the impact of mutations in the PhoPR two-component system on the virulence and potential host tropism in the MTBC.
52. Whelan, A. O. et al. Revisiting host preference in the *Mycobacterium tuberculosis* complex: experimental infection shows *M. tuberculosis* H37Rv to be avirulent in cattle. *PLOS One* **5**, e8527 (2010).
53. Ameni, G. et al. Transmission of *Mycobacterium tuberculosis* between farmers and cattle in central Ethiopia. *PLOS One* **8**, e76891 (2013).
54. Manchester, K. Tuberculosis and leprosy in antiquity: an interpretation. *Med. Hist.* **28**, 162–173 (1984).
55. Cole, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
56. Garnier, T. et al. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl Acad. Sci. USA* **100**, 7877–7882 (2003).
57. Brosch, R. et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl Acad. Sci. USA* **99**, 3684–3689 (2002).
58. Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A. & Behr, M. A. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* **186**, 74–80 (2002).
59. Dippenaar, A. et al. Whole genome sequence analysis of *Mycobacterium suricattae*. *Tuberculosis (Edinb.)* **95**, 682–688 (2015).
60. de Jong, B. C., Antonio, M. & Gagneux, S. Mycobacterium africanum — review of an important cause of human tuberculosis in West Africa. *PLOS Negl. Trop. Dis.* **4**, e744 (2010).
61. Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R. & Gordon, S. V. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **7**, 537–544 (2009).
62. Comas, I. et al. Population genomics of *Mycobacterium tuberculosis* in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in sub-Saharan Africa. *Curr. Biol.* **25**, 3260–3266 (2015).
63. Comas, I. et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* **45**, 1176–1182 (2013).
64. Kay, G. L. et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat. Commun.* **6**, 6717 (2015).
65. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
66. Baker, O. et al. Human tuberculosis predates domestication in ancient Syria. *Tuberculosis (Edinb.)* **95** (Suppl. 1), S4–S12 (2015).
67. Hershkovitz, I. et al. Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the Eastern Mediterranean. *PLOS ONE* **3**, e3426 (2008).
68. Lee, O. Y. et al. *Mycobacterium tuberculosis* complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, Bison antiquus. *PLOS One* **7**, e41923 (2012).
69. Eldholm, V. et al. Armed conflict and population displacement as drivers of the evolution and dispersal of *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **113**, 13881–13886 (2016).
70. Ho, S. Y. & Larson, G. Molecular clocks: when times are a-changin'. *Trends Genet.* **22**, 79–83 (2006).
71. Duchene, S. et al. Genome-scale rates of evolutionary change in bacteria. *Microb. Genom.* **2**, e000094 (2016).
72. Comas, I. & Gagneux, S. A role for systems epidemiology in tuberculosis research. *Trends Microbiol.* **19**, 492–500 (2011).
73. Gagneux, S. Host–pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 850–859 (2012).
74. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
75. Hirsh, A. E., Tsolaki, A. G., DeRiemer, K., Feldman, M. W. & Small, P. M. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc. Natl Acad. Sci. USA* **101**, 4871–4876 (2004).

76. Baker, L., Brown, T., Maiden, M. C. & Drobniewski, F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* **10**, 1568–1577 (2004).
77. Reed, M. B. et al. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J. Clin. Microbiol.* **47**, 1119–1128 (2009).
78. Gagneux, S. et al. Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 2869–2873 (2006).
79. Fenner, L. et al. HIV infection disrupts the sympatric host–pathogen relationship in human tuberculosis. *PLOS Genet.* **9**, e1003318 (2013).
80. Asante-Poku, A. et al. Mycobacterium africanum is associated with patient ethnicity in Ghana. *PLOS Negl. Trop. Dis.* **9**, e3370 (2015).
81. Asante-Poku, A. et al. Molecular epidemiology of *Mycobacterium africanum* in Ghana. *BMC Infect. Dis.* **16**, 385 (2016).
82. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Systemat.* **19**, 207–233 (1988).
83. Coll, F. et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4812 (2014).
84. Stucki, D. et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat. Genet.* **48**, 1535–1543 (2016).
In this study, several thousand MTBC L4 clinical isolates are subtyped into sublineages. The geographical distribution of these sublineages supports a classification into ecological specialists and generalists.
85. Comas, I. et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* **42**, 498–503 (2010).
86. Pepperell, C. S. et al. The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLOS Pathog.* **9**, e1003543 (2013).
87. Coscolla, M. et al. *M. tuberculosis* T cell epitope analysis reveals paucity of antigenic variation and identifies rare variable TB antigens. *Cell Host Microbe* **18**, 538–548 (2015).
88. Yruela, I., Contreras-Moreira, B., Magalhaes, C., Osorio, N. S. & Gonzalo-Asensio, J. *Mycobacterium tuberculosis* complex exhibits lineage-specific variations affecting protein ductility and epitope recognition. *Genome Biol. Evol.* **8**, 3751–3764 (2016).
89. Blaser, M. J. & Kirschner, D. The equilibria that allow bacterial persistence in human hosts. *Nature* **449**, 843–849 (2007).
90. Zheng, N., Whalen, C. C. & Handel, A. Modeling the potential impact of host population survival on the evolution of *M. tuberculosis* latency. *PLOS One* **9**, e105721 (2014).
91. Barnes, I., Duda, A., Pybus, O. G. & Thomas, M. G. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution* **65**, 842–848 (2011).
92. Coscolla, M. & Gagneux, S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov. Today Dis. Mech.* **7**, e43–e59 (2010).
93. Hershberg, R. et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLOS Biol.* **6**, e311 (2008).
94. Williams, A. C. & Dunbar, R. I. Big brains, meat, tuberculosis and the nicotinamide switches: co-evolutionary relationships with modern repercussions on longevity and disease? *Med. Hypotheses* **83**, 79–87 (2014).
95. Supply, P. et al. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol. Microbiol.* **47**, 529–538 (2003).
96. Gagneux, S. & Small, P. M. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* **7**, 328–337 (2007).
97. Comas, I., Homolka, S., Niemann, S. & Gagneux, S. Genotyping of genetically monomorphic bacteria: DNA sequencing in mycobacterium tuberculosis highlights the limitations of current methodologies. *PLOS ONE* **4**, e7815 (2009).
98. Comas, I. et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2012).
In this study, the authors use a combination of experimental evolution and molecular epidemiological data to identify fitness compensatory mutations in the RNA polymerase of rifampicin-resistant MTBC.
99. Farhat, M. R. et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 1183–1189 (2013).
100. Namouchi, A., Didelot, X., Schock, U., Gicquel, B. & Rocha, E. P. After the bottleneck: genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. *Genome Res.* **22**, 721–734 (2012).
101. Liu, X., Gutacker, M. M., Musser, J. M. & Fu, Y. X. Evidence for recombination in *Mycobacterium tuberculosis*. *J. Bacteriol.* **188**, 8169–8177 (2006).
102. Smith, N. H., Gordon, S. V., de la Rua-Domenech, R., Clifton-Hadley, R. S. & Hewinson, R. G. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* **4**, 670–681 (2006).
103. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
104. Eldholm, V. et al. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol.* **15**, 490 (2014).
105. Charlesworth, B. Background selection 20 years on: the Wilhelmine, E. Key 2012 invitational lecture. *J. Hered.* **104**, 161–171 (2013).
106. Pepperell, C. et al. Bacterial genetic signatures of human social phenomena among *M. tuberculosis* from an Aboriginal Canadian population. *Mol. Biol. Evol.* **27**, 427–440 (2010).
107. Trauner, A. et al. The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol.* **18**, 71 (2017).
108. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740–743 (2012).
109. Luo, T. et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc. Natl Acad. Sci. USA* **112**, 8136–8141 (2015).
110. Merker, M. et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **47**, 242–249 (2015).
111. Lieberman, T. D. et al. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nat. Med.* **22**, 1470–1474 (2016).
This paper provides a detailed view into the within-host diversity and evolution of the MTBC, using thousands of genome sequences obtained from individuals co-infected with HIV and TB who died before treatment initiation.
112. Dean, G. S. et al. Minimum infective dose of *Mycobacterium bovis* in cattle. *Infect. Immun.* **73**, 6467–6471 (2005).
113. Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
114. Eldholm, V. & Balloux, F. Antimicrobial resistance in *Mycobacterium tuberculosis*: the odd one out. *Trends Microbiol.* **24**, 637–648 (2016).
115. Bloemberg, G. V. et al. Acquired resistance to bedaquiline and delamanid in therapy for tuberculosis. *N. Engl. J. Med.* **373**, 1986–1988 (2015).
This is the first report of the acquisition of resistance to the two new tuberculosis drugs bedaquiline and delamanid during the treatment of a single patient.
116. Gygli, S. M., Borrell, S., Trauner, A. & Gagneux, S. Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. *FEMS Microbiol. Rev.* **41**, 354–373 (2017).
117. Muller, B., Borrell, S., Rose, G. & Gagneux, S. The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Trends Genet.* **29**, 160–169 (2013).
118. Middlebrook, G. & Cohn, M. L. Some observations on the pathogenicity of isoniazid-resistant variants of tubercle bacilli. *Science* **118**, 297–299 (1953).
119. Manson, A. L. et al. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. *Nat. Genet.* **49**, 395–402 (2017).
This study is the largest whole-genome-based analysis of MTBC drug resistance to date.
120. Sander, P. et al. Fitness cost of chromosomal drug resistance-conferring mutations. *Antimicrob. Agents Chemother.* **46**, 1204–1211 (2002).
121. Gagneux, S. et al. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**, 1944–1946 (2006).
122. Sherman, D. R. et al. Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* **272**, 1641–1643 (1996).
123. Song, T. et al. Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the β' subunit of RNA polymerase. *Mol. Microbiol.* **91**, 1106–1119 (2014).
124. de Vos, M. et al. Putative compensatory mutations in the rpoC gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. *Antimicrob. Agents Chemother.* **57**, 827–832 (2013).
125. Borrell, S. et al. Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. *Evol. Med. Public Health* **2013**, 65–74 (2013).
126. Fenner, L. et al. Effect of mutation and genetic background on drug resistance in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **56**, 3047–3053 (2012).
127. Borrell, S. & Gagneux, S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis*. *Int. J. Tuberc Lung Dis.* **13**, 1456–1466 (2009).
128. Werngren, J. & Hoffner, S. E. Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J. Clin. Microbiol.* **41**, 1520–1524 (2003).
129. Ford, C. B. et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45**, 784–790 (2013).
130. Copin, R. et al. Sequence diversity in the pe_pgrs genes of *Mycobacterium tuberculosis* is independent of human T cell recognition. *mBio* **5**, e00960–e00913 (2014).
131. Kwan, C. K. & Ernst, J. D. HIV and tuberculosis: a deadly human syndemic. *Clin. Microbiol. Rev.* **24**, 351–376 (2011).
132. Rappuoli, R., Bottomley, M. J., D'Oro, U., Finco, O. & De Gregorio E. Reverse vaccinology 2.0: human immunology instructs vaccine antigen design. *J. Exp. Med.* **213**, 469–481 (2016).
133. Tameris, M. D. et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* **381**, 1021–1028 (2013).
134. Eldholm, V. et al. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *eLife* **213**, 469–481 (2016).
135. Koch, A. S. et al. The influence of HIV on the evolution of *Mycobacterium tuberculosis*. *Mol. Biol. Evol.* **34**, 1654–1668 (2017).
136. Copin, R. et al. Within host evolution selects for a dominant genotype of *Mycobacterium tuberculosis* while T cells increase pathogen genetic diversity. *PLOS Pathog.* **12**, e1006111 (2016).
137. Brites, D. & Gagneux, S. The nature and evolution of genomic diversity in the *Mycobacterium tuberculosis* complex. *Adv. Exp. Med. Biol.* **1019**, 1–26 (2017).
138. Casali, N. et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286 (2014).

Acknowledgements
The author thanks all the members of his group for the stimulating discussions over the years. Work in the author's laboratory is supported by the Swiss National Science Foundation (grants 310030_166687, IZRJZ3_164171 and IZLSZ3_170834), the European Research Council (309540-EVODRTB) and SystemsX.ch.

Competing interests
The author declares no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Microbiology thanks R. Brosch, H. Donoghue and P. Supply for their contribution to the peer review of this work.