Daniela Brites
Sebastien Gagneux

Co-evolution of Mycobacterium tuberculosis and Homo sapiens

Authors' address

Daniela Brites[1], Sebastien Gagneux[1]
[1]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute and University of Basel, Basel, Switzerland.

Correspondence to:

Sebastien Gagneux
Swiss Tropical and Public Health Institute
Socinstrasse 57
4002 Basel, Switzerland
Tel.: +41 61 284 8369
e-mail: sebastien.gagneux@unibas.ch

This article is part of a series of reviews covering Tuberculosis appearing in Volume 264 of *Immunological Reviews*.

Video podcast available
Go to www.immunologicalreviews.com to watch an interview with Guest Editor Carl Nathan.

Summary: The causative agent of human tuberculosis (TB), *Mycobacterium tuberculosis*, is an obligate pathogen that evolved to exclusively persist in human populations. For *M. tuberculosis* to transmit from person to person, it has to cause pulmonary disease. Therefore, *M. tuberculosis* virulence has likely been a significant determinant of the association between *M. tuberculosis* and humans. Indeed, the evolutionary success of some *M. tuberculosis* genotypes seems at least partially attributable to their increased virulence. The latter possibly evolved as a consequence of human demographic expansions. If co-evolution occurred, humans would have counteracted to minimize the deleterious effects of *M. tuberculosis* virulence. The fact that human resistance to infection has a strong genetic basis is a likely consequence of such a counter-response. The genetic architecture underlying human resistance to *M. tuberculosis* remains largely elusive. However, interactions between human genetic polymorphisms and *M. tuberculosis* genotypes have been reported. Such interactions are consistent with local adaptation and allow for a better understanding of protective immunity in TB. Future 'genome-to-genome' studies, in which locally associated human and *M. tuberculosis* genotypes are interrogated in conjunction, will help identify new protective antigens for the development of better TB vaccines.

Keywords: *host, pathogen, adaptation, virulence, selection*

Introduction

Co-evolution between a host and a pathogen occurs when evolutionary changes in the pathogen that increase infectivity are counteracted by evolutionary changes in the host that increase resistance to infection. Such reciprocal changes have been demonstrated in experimental systems such as bacteria and their bacteriophages (1, 2). However, in host–pathogen associations with higher levels of biological complexity, a rigorous demonstration of these phenomena is not straightforward (3), particularly when considering infectious diseases of humans (4). Yet, in many instances, despite the lack of a formal proof of host–parasite-induced reciprocal changes, co-evolution still remains a likely explanation for many of the patterns observed in host–pathogen interactions.

For co-evolution to occur between a host and a pathogen: (i) pathogen infectivity and host resistance have to have reciprocal antagonistic effects on the fitness of each; (ii) the variation in pathogen infectivity and in host resistance have

to be at least partially genetically determined; and (iii) the outcome of the interactions between the partners has to be dependent on the respective genotypes (5–8). Here, we review the association of *Mycobacterium tuberculosis* complex (MTBC) and *Homo sapiens* in light of these aspects.

Tuberculosis (TB) remains a leading cause of morbidity and mortality in the world (9). Despite years of research, no vaccine currently exists that protects reliably against pulmonary TB in adults, which is the most transmissible form of the disease (10). Indeed, although many components of the host immune response against MTBC are known, the specific molecules and mechanisms underlying protective immunity remain elusive (11, 12). Remarkably, some hallmarks of TB infection and disease, such as latency, remain poorly understood (13, 14). Evidence is emerging that in addition to host and environmental factors, the genetic variation in MTBC also plays a role in the clinical phenotypes of TB (15, 16). However, little is known about the interaction between human and MTBC genetic diversity, and it has been argued that new paradigms and new conceptual frameworks are required to better understand and ultimately better control TB globally (12, 14, 17, 18).

The purpose of this review was to discuss new ideas and concepts that together could guide future research. We focus on the bacterial side of the association, but always in the context of its relation to the host. We start by introducing MTBC, the different phylogenetic lineages that make up this group of organisms, as well as their biogeographical distribution and primary host range. Next, we discuss how MTBC virulence might have evolved in response to different evolutionary forces, in particular human demography. Local adaptation is a phenomenon expected from a co-evolutionary association, and the evidence for it in human TB is presented. Co-evolution is expected to lead to interactions between host and pathogen loci; hence, we summarize the existing evidence with respect to such interactions in TB. Recent research on the age of MTBC is summarized in light of both genetic and archeological data. In the final section of this review, we discuss how genetic drift and natural selection might influence the ability of MTBC to adapt to its host and end by highlighting how understanding more about the co-evolutionary history of MTBC and humans can guide us in our quest for a better TB vaccine.

The *Mycobacterium tuberculosis* complex

MTBC comprises various bacterial species and sub-species sharing 99.9% DNA sequence identity but differing in their primary host range. We describe below the different members of MTBC, dividing them according to their primary host: human-adapted MTBC lineages, the lineages adapted to wild and domestic mammalian hosts, and *Mycobacterium canettii*. For simplicity, throughout this review, we use the term MTBC to refer to all MTBC lineages except *M. canettii*.

The human-adapted MTBC lineages

The human-adapted forms of MTBC include *Mycobacterium tuberculosis sensu stricto* and *M. africanum*. Humans are the only known host where infection and transmission by *M. tuberculosis* and *M. africanum* occur efficiently and where infection cycles are known to be sustainably maintained (19). Unlike postulated until a decade ago, it is now clear that humans did most likely not acquire the tubercle bacilli from cattle during animal domestication, as certain human MTBC lineages are phylogenetically more basal (i.e. 'ancestral') than *M. bovis* (20, 21). Furthermore, *M. bovis* has lost several genes that are still present in all other *M. tuberculosis* lineages and does not contain any gene that does not occur in at least some strains of *M. tuberculosis sensu stricto*. For a better understanding of the terminology used to classify the main lineages of MTBC, we briefly summarize the discoveries underlying that nomenclature. A main step into elucidating the genetic structure of MTBC was the recognition that a group of strains associated with major epidemics like the 'Beijing' and 'Haarlem' strains all shared a deletion (TbD1) that is not present in *M. africanum* strains (20). As large-scale ongoing horizontal gene transfer has not been described in MTBC, it was assumed that the TbD1-deleted strains were younger, in evolutionary terms, than strains without this deletion. Accordingly, TbD1-deleted strains have been referred to as evolutionary modern, whereas *M. africanum* and several other MTBC lineages were considered ancient (20, 22). It was also recognized that TbD1 was not deleted in *M. bovis* and that *M. africanum*, *M. bovis*, and other animal strains shared another deletion (RD9) not present in the modern strains or any other strains belonging to *M. tuberculosis sensu stricto* (20). During the subsequent decade, the phylogenetic relationships among the different MTBC groups were further deciphered using different genetic markers, culminating in large-scale comparative whole-genome sequencing (23–29). The most recent phylogenetic inferences reveal that the human-adapted lineages of MTBC have a strong phylogeographical structure comprising seven major phylogenetic lineages that diverged from a common ancestor and diversified in different regions of the world

(23–28). The TbD1-deleted strains known as modern are now recognized as three separate, albeit related lineages: Lineage 4 (also known as Euro-American) that has a broad distribution in Europe and America but also in Africa and the Middle-East, Lineage 2 (East-Asian lineage, includes the Beijing family of strains) that is widely distributed in East Asian countries, and Lineage 3 that has a relative narrow distribution occurring in East Africa and in Central and South Asia (24, 26) (Fig. 1). In addition to the animal-adapted lineages, the ancient lineages without the deletion

Fig. 1. Whole-genome phylogeny of 220 strains of MTBC. Bootstrap support for the main branches after inference with Neighbor-joining (left) and Maximum-likelihood (right) analyses are shown. The grey circle indicates the 'modern' strains (see main text). Scale bars indicate substitutions per site. The occurrence of the TbD1 deletion is indicated with an asterisk. Adapted from Comas et al. (29).

in TbD1 comprise Lineage 1 (also known as Indo-Oceanic), which occurs around the Indian Ocean and the Philippines, Lineage 5 (also known as *M. africanum* West African 1), and Lineage 6 (*M. africanum* West African 2), which are geographically restricted to Western African countries (30), and the recently described Lineage 7 (Ethiopian lineage), which is restricted to Ethiopia (24, 26, 28). Rooting (Table 1) the phylogenetic relationships between the different MTBC lineages with *M. canettii* reveals that the common ancestor of *M. africanum* Lineages 5 and 6, and the animal-adapted lineages split early from the common ancestor of MTBC. Another early split led to the ancestor from which all lineages belonging to *M. tuberculosis sensu stricto* derived, with Lineage 1 being the most basal within this group of strains, followed by Lineage 7. A later divergence event originated the ancestor of the modern Lineages 2, 3, and 4 (Fig. 1). Note that the modern Lineages 2, 3, and 4 are monophyletic (Table 1) and therefore comprise a more closely related group of organisms, compared to the various ancient lineages, which are paraphyletic (Table 1). The genetic properties of these lineages and the phenotypical and epidemiological consequences of that variation have been reviewed recently elsewhere (15, 16). Here, we focus on the aspects we consider most relevant for understanding the evolution of MTBC and its association with different human population. Members of the modern Lineages 2 and 4 are responsible for most TB cases in the world, and certain strains belonging to these lineages have been implicated in major outbreaks of both drug-sensitive and drug-resistant TB (15, 16, 31). Furthermore, these lineages have been
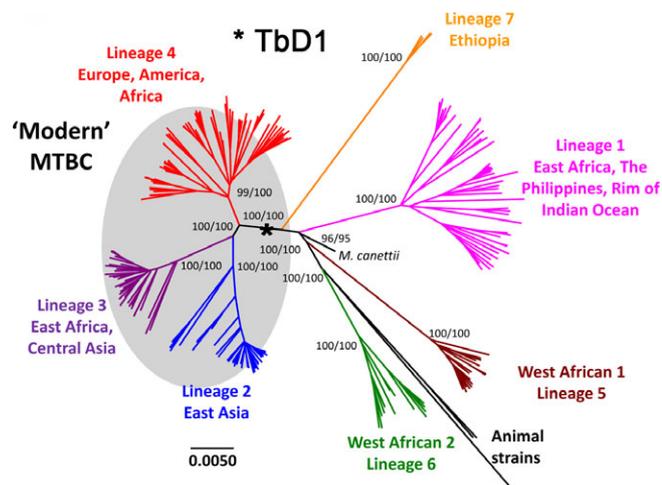
Table 1. Definition of terms used from phylogenetics and population genetics

Term	Definition
Rooting	Providing the evolutionary order of phylogenetic branching events, by defining the position of the common ancestor to all samples on a phylogenetic tree
Monophyletic group	A group that contains all the descendants of a common ancestor
Paraphyletic group	A group that has a common ancestor but that does not included all descendants of that common ancestor
Molecular clock	Amount of differences between the DNA molecules of two species is a function of the time since their evolutionary separation
Calibration point	Attributing one or more dates to tips (e.g. taxa) or nodes of the phylogenetic tree. Dates are usually based on fossils, but they can also be based on known historical events, or in the case of fast evolving microorganism, on isolation times
Coalescent time of a set of strains	The time that has passed since the most recent common ancestor of all those strains existed
Polymorphism	A genetic variant that segregates within a population
Substitution	A genetic variant fixed between two species
Population bottleneck	Reduction on the size of a population, at least for one generation, due to external effects
Founder effect	Establishing a new population by a small number of individuals not genetically representative of the original source population. Leads to loss of genetic diversity by genetic drift
Genetic drift	Random fluctuation of alleles in populations, e.g. after a population bottleneck, randomly, certain alleles are removed, whereas others are kept
Background selection	The effect that selection on deleterious mutations has on other linked non-deleterious mutations, leading to loss of genetic diversity
Positive selection	Selection favoring an advantageous mutation
Negative selection	Selection against a deleterious mutation

shown to cause faster progression to active disease and to be more virulent in animal models than ancient lineages (reviewed in 15, 16). Their monophyletic origin and their expansion through different parts of the world suggest that they could have accompanied the first migrations of modern humans 'Out-of-Africa' and diversified together with different human populations (26, 32). Following this scenario, the prediction would be that particular MTBC lineages would be better adapted to their local host population; this seems to be at least partially the case, as further discussed below. It is notorious that certain lineages, but also sub-groups within the main MTBC lineages (we might call those 'sub-lineages') are geographically more widespread than others. Whether this geographic population structure reflects local adaptation, human demography, founder effects, or a combination of these remains unclear. Comparing the geographically restricted Lineages 5, 6, and 7 with the geographically much more widespread Lineages 2 and 4 provides good examples. Moreover, some sub-lineages within Lineage 4 also differ in their geographic distribution; for example, the Haarlem and the LAM families occur worldwide, while other sub-lineages are geographically restricted. Examples for the latter include the Cameroon and Uganda families of strains, which are the most prevalent strains those countries, respectively (33, 34). Human migrations might account for some of these patterns. For instance, Haarlem and LAM strains were likely brought to the Americas by Europeans during colonization and the massive European migrations of the 18th and 19th centuries to the New World (35). By contrast, even though large numbers of Africans were brought to the Americas during the slave trade, African MTBC variants like *M. africanum*, or the Cameroon and Uganda sub-lineages of Lineage 4 are almost completely absent from the Americas (30). Hence, the differential distribution of MTBC lineages and sub-lineages could also reflect intrinsic characteristics of the strains. For example, some lineages might be thought of 'generalists', i.e. able to persist in different human populations, and other more 'specialists', i.e. able to persist only in one or a few particular host populations (36).

In apparent contradiction with the idea of host-specific pathogen adaptation in MTBC is the fact that transmission of *M. tuberculosis* to animals and *M. bovis* to humans does occur occasionally (37, 38). This also has occurred in our evolutionary past, as revealed by the detection of ancient *M. bovis* DNA in 2000 years old human bones in Siberia (39) and of DNA from strains related to *M. pinnipedii*, an MTBC member usually found in seals and sea lions, in 1000-year-old

human remains from Peru (40). The latter study provided the first whole-genome sequence from ancient MTBC DNA and has been stimulating the discussion on the origin of TB in the Americas (41). There is evidence for Pre-Columbian TB in coastal regions of South America; extant South American MTBC strains, however, were most likely introduced by Europeans, as most of autochthonous TB in that part of the world today is caused by Lineage 4 (24). However, whether the *M. pinnipedii*-related strains could have been a general source of infection in the Americas 1000 years ago and whether these strains were human-adapted remains unknown. Only sequencing more pre- and post-Columbian strains from different geographical regions across the Americas will allow a more definite conclusion.

Although there is some level of cross-species infection and transmission, sustainable human-to-human transmission of animal-adapted MTBC strains has not been demonstrated (42). As we discuss below, the animal strains all derived from one common ancestor that is shared with *M. africanum* Lineage 6, which is currently considered a human-adapted lineage, and with no other MTBC lineage. Given the proximity of domestic animals and humans throughout human evolutionary history, there were certainly opportunities for jumps of other human MTBC lineages to animals. However, this does not seem to have occurred throughout the evolution of MTBC, supporting the notion that the human-adapted MTBC lineages became specialized to infect and persist in human populations.

The animal-adapted MTBC lineages

In contrast to the human-adapted MTBC lineages, some animal-adapted strains of MTBC are able to establish infections and transmit within other animal species aside from their primary hosts. We briefly discuss the animal-adapted lineages, as their geographic distribution, host range, and virulence characteristics might help us to better understand the association of *M. tuberculosis sensu stricto* and *M. africanum* with humans. The animal-adapted MTBC comprise the following: *M. microti*, found in voles, wood mice, and shrews but recently suggested to have spilt-over to domestic cats in the United Kingdom (43); the so-called dassie bacillus, which is a pathogen of hyraxes (44, 45); *M. pinnipedii* found in seals and sea lions; *M. caprae* found in goats but also described as causing sustainable disease in red deer populations (46, 47); and *M. bovis* found in cattle but also able to infect other domestic and wild animals, in particular badgers (19, 48). Other members of the complex have been described more recently. *M. orygis* has been found in oryxes, gazelles, deer,

antelopes, waterbucks, and buffalos, but also in humans from East and Southeast Asia (49, 50); indeed, at least half of the *M. orygis* isolates reported to date have come from human TB patients (49). More recently, *M. mungi* was described from mongoose populations in Botswana (51) and *M. suricattae* in meerkats from South Africa (52). Finally, a novel MTBC strain was isolated from a wild chimpanzee in Cote d'Ivoire (31). All these wild and domestic animal lineages of MTBC and the human-adapted *M. africanum* Lineage 6 share the deletions RD7, RD8, RD9, and RD10, supporting the notion that *M. tuberculosis* did not derive from a bovine form during the Neolithic Revolution, as traditionally postulated. These findings also suggest that the MTBC diversity in wild animals may be much greater than previously appreciated. The hypothesis that all these animal-adapted lineages descended from a singly human-adapted ancestor might also be too simplistic (31). Perhaps, the common ancestor of Lineage 6 and all the animal-adapted strains was not primarily human-adapted but rather a promiscuous bacillus with a broad host spectrum. Host specialization might then have occurred at a later stage during the evolution of these lineages. Interestingly, just like in the human-adapted MTBC lineages, the African continent also harbors most of the diversity in animal-adapted lineages, further supporting the African origin of MTBC (24, 26, 29, 53). Another interesting observation is that most of the animal-adapted lineages affecting primarily wild animals remained geographically restricted to Africa, whereas the lineages adapted to domestic animals spread to different parts of the world, possibly as a consequence of human migration and trade.

Some new insights into the genetic mechanisms that may explain why sustainable infections with animal-adapted strains only rarely occur in humans and the fact that *M. africanum* exhibits lower virulence than *M. tuberculosis sensu stricto* have recently been gained through some elegant work by Gonzalo-Asensio *et al.* (54). The authors showed that a mutation affecting the *phoPR* genes in the common ancestor of *M. africanum* Lineage 5 and Lineage 6 and of animal-adapted strains of MTBC, when transferred to a *M. tuberculosis sensu stricto* genetic background, lead to a decrease in virulence in both mice and human primary macrophages. PhoPR encode a two-component system regulating the synthesis and export of important *M. tuberculosis* virulence determinants such as LipF, ESAT-6, and lipids of the polyacyltrehalose and sulfolipid families (54, 55). However, the data suggest that the secretion of ESAT-6 in *M. africanum* Lineage 6 and *M. bovis* is independent of PhoPR because of the loss of binding sites

of the regulatory proteins EspR and MprAB that normally mediate ESAT-6 secretion in a PhoPR-dependent manner (54). This loss is caused by the deletion of RD8, suggesting that this deletion acted as a compensatory mechanism, restoring virulence in the common ancestor of *M. africanum* Lineage 6 and the animal-adapted lineages (54) (Fig. 2). Strengthening this hypothesis, both *M. tuberculosis sensu stricto* and *M. canettii*, which represent a derived and the ancestor-like (see below) groups of the MTBC phylogeny, respectively, and which have no PhoPR mutations, regulate the secretion of ESAT-6 through PhoPR (54). Interestingly, the same PhoPR mutation is present in *M. africanum* Lineage 5, which has no deletion in RD8. Further work will tell whether an alternative compensatory change could have occurred in Lineage 5.

Mycobacterium canettii

Within the genus *Mycobacterium*, the most likely bacteria to share the most recent common ancestor with MTBC are the so-called 'smooth tubercle bacilli' (STB), which include *M. canettii* (53, 56, 57). *Mycobacterium canettii* is formally part of MTBC, but unlike the other MTBC members described above, undergoes frequent horizontal gene transfer and inter-strain recombination (53, 57). The genetic diversity between different strains of STB is comparable to that of other recombining bacteria, and much higher than the genetic diversity among the other MTBC lineages (53, 57). Comparative genomics has revealed evidence of recombination between STB and other MTBC strains, although it remains to be demonstrated if these recombination events took place before or after the divergence of the different MTBC lineages (57–60). STB share an average nucleotide sequence similarity of over 95% and a high degree of synteny with the other MTBC members, justifying their inclusion in the complex (57). To date, only about 60 STB isolates have been reported, mainly from the Horn of Africa. STB seems to disproportionally affect expatriate individuals, children, and HIV co-infected TB patients (61). Furthermore, transmission of STB among humans has not been demonstrated; hence, these bacteria are probably acquired from the environment (62). However, no animal or environmental reservoir is currently known to harbor STB. Experiments in mice suggest both a lower virulence and shorter persistence during chronic infection with STB compared to H37Rv (57). Overall, the virulence phenotypes, the epidemiological characteristics, as well as certain genomic features such as the presence of different clustered regularly interspaced short palindromic repeats loci and associated *cas* genes
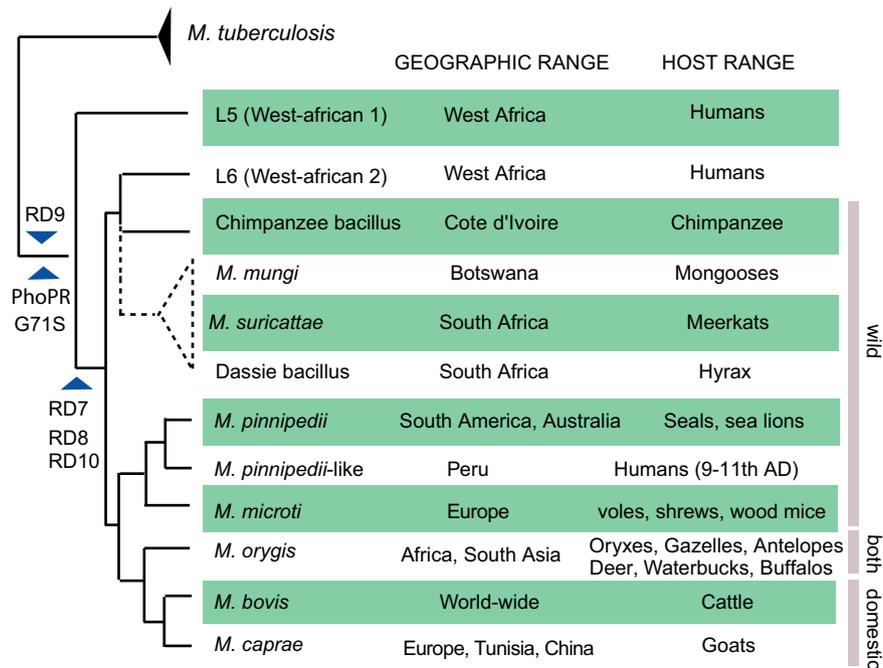
Fig. 2. Simplified schematic illustrating the phylogenetic relationships of animal strains based on different phylogenies published previously (31, 40, 49, 51, 52). The length of the branches does not reflect real phylogenetic distances. Dashed branches represent putative relative relationships. The phylogenetic positions of genomic deletions discussed in the main text and of the mutations in the *phoPR* genes (54) are indicated. The geographical and host ranges of the animal lineages do not include reports from zoos, as those were considered incidental infections.

(CRISPR-Cas), and the presence of CRISPR spacers matching prophages present in other mycobacteria (e.g. *M. marinum*) (61) are consistent with STB being environmental bacteria, only occasionally pathogenic in humans.

This finding is in sharp contrast to the other members of MTBC, which are obligate pathogens. The MTBC possibly evolved from a *M. canettii*-like ancestor, a strain with a loose ability to survive both in the environment and inside a host. The genetic changes associated with the transition to becoming a professional pathogen involved gene loses and gene acquisitions (57–59, 63), and the diversification of gene families such as the pe_pgrs and esx gene families (64, 65). Current examples of facultative intracellular bacteria with environmental reservoirs exist as the opportunistic pathogen *Listeria monocytogenes* illustrates (66). Between the *M. canettii*-like ancestor and the most recent common ancestor of all extant MTBC, the ability to replicate outside a host was lost and the ability to overcome the host defenses and to persist within host cells was acquired. This has come at the cost for the pathogen of harming the host (virulence), because at least some degree of pathology is essential for MTBC transmission to occur. This might have marked the beginning of the co-evolutionary association between *M. tuberculosis* and *Homo sapiens*. Notably, the trajectory toward host special-

ization has allowed this MTBC ancestor to spread all over the world and become one of the most successful human pathogens.

The evolution of virulence in MTBC

In the process of becoming obligatory pathogens, the ancestors of MTBC must have evolved strategies to overcome the immune response of the host, to transmit between hosts, and to avoid becoming extinct in small human populations. The affected humans must have evolved counter strategies, because for transmission of MTBC to occur, human disease is necessary. Specifically, lung cavitation and increased sputum-smear positivity are associated with more efficient transmission of MTBC (67–71). Furthermore, crowding (i.e. increased host density) is known to enhance MTBC transmission (72). These characteristics suggest that MTBC virulence could have evolved under a trade-off: the pathogen has regulated its virulence to be able to persist and spread in human populations exhibiting increased densities and greater resistance or tolerance (5, 73, 74). There are two lines of evidence supporting this suggestion: one is the preferential association of particular MTBC lineages with their local human populations, which is highly suggestive of adaptation to their sympatric host populations, and the

second is that different MTBC genotypes have evolved differences in virulence traits which might reflect the outcome of different strategies to persist in the specific human population they have evolved with. We now further discuss these points below.

Local adaptation in human-adapted MTBC

If hosts are heterogeneous qualitatively (e.g. different genotypes) and quantitatively (e.g. abundance), temporal and spatial dynamics of host–pathogen associations are expected under a co-evolutionary process such that a snapshot in time or space will reveal that certain host genotypes are more susceptible to certain parasite genotypes (75, 76). One example of such a dynamic is local adaptation, whereby a local pathogen population has a higher fitness in a local host (sympatric) than in a non-local host (allopatric) (75, 76). Pathogens are expected to evolve quicker than their hosts due to their comparably reduced generation time, and thus tend to be ahead in the co-evolutionary process. However, the latter does not always hold true, depending on the specific characteristics of the local association (77). There is some evidence for local adaptation of human-adapted MTBC lineages to different host populations. Specifically, the geographic origin of a TB patient is in many cases a good predictor of the infecting MTBC lineage (78). This remains true in cosmopolitan areas such as San Francisco, London, and Montreal, where humans and bacteria intermingle (23, 24, 78, 79). Furthermore, MTBC transmission was shown to be higher in sympatric host–pathogen associations than in allopatric combinations (24). Social factors such as preferential social mixing among ethnic groups are likely to account for some of these observations. However, there are indications that biological factors also play a role. For example, in a nation-wide molecular epidemiology study in Switzerland, it was found that recent transmission of sympatric MTBC strains was more likely than that of allopatric strains among European-born patients. Importantly, in contrast to HIV-negative TB patients, patients that were HIV-infected were more likely to be associated with an allopatric MTBC lineage (80). Moreover, this association became stronger with increased immune suppression as measured by reduced CD4$^+$ T-cell counts (80).

Given that TB disease is necessary for MTBC transmission, what are the implications of local adaptation; are local MTBC strains more or less virulent in their local host population? One study in Texas has shown that different levels of pulmonary impairment could not be explained by

differences in the causative MTBC lineages alone, but by the pathogen–host association; patients infected with sympatric MTBC lineages suffered less lung impairment than patients infected with allopatric lineages (81). Transmission was not measured in this study, so it is not possible to relate it to virulence. Nevertheless, it is likely that the different virulence characteristics of the different MTBC strains and lineages, such as their replication ability, the nature of immune responses elicited, and the duration of latency, are the result of fine-tuned adaptations to the particular local host population.

Variation in virulence in human MTBC

The progression of TB in patients is largely determined by the response of the host immune system and environmental variables (17). In addition, mounting evidence points to an important effect of the MTBC genetic background (reviewed in 15, 16). The most abundant MTBC lineages in the world are representatives of the modern lineages, reflecting their evolutionary success. As already mentioned, these include the more transmissible and virulent strains of MTBC, some of which have been responsible for large outbreaks in different regions of the world, others associated with the emergence and spread of drug resistance. Representatives of Lineage 2 and Lineage 4 have generally a higher replication capacity in human macrophages and animal models of infection. There is also evidence that the modern Lineage 2, 3, and 4 elicit reduced and/or delayed pro-inflammatory immune responses (15, 16). Lineages 2 and 4 have also been associated with higher transmission as measured by the number of secondary cases generated, reflecting a link between increased virulence in animal models and enhanced transmissibility in human populations (15, 16). The high virulence of the modern lineages, in particular Lineages 2 and 4, has been hypothesized to have evolved as a response to the increasing size and density of human populations starting with the Neolithic Demographic Transition and the onset of agriculture (around 10 000 year ago), and later during the Industrial Revolution and the associated urban crowding (300–200 years ago) (26, 29, 32). In contrast, the much more geographically restricted *M. africanum* lineages show lower a prevalence compared to the modern lineages, even within their specific range of distribution in West African (30), with the possible exception of Guinea-Bissau (82). Moreover, there is indication that the prevalence of *M. africanum* lineages is decreasing in some parts of West Africa (82–84), possibly as a result of out-competition by

M. tuberculosis sensu stricto. The *M. africanum* lineages are also known to be less virulent than modern strains in animal models (85), and *M. africanum*-infected patients have lower T-cell responses to ESAT-6, which is an important virulence factor in MTBC (86, 87).

Latency is a hallmark of human TB (13, 14). The majority of people infected with MTBC mount an effective adaptive immune response that controls the replication of the bacteria, leading to a persistent infection with no disease symptoms (88). Only a small fraction of latently infected individuals will re-activate to active disease, often several decades after initial exposure (14). By keeping MTBC replication in check, the host avoids MTBC virulence. However, latency could also have evolved as a strategy for the pathogen to ensure transmission and avoid extinction in small host populations (89). Supporting this view, MTBC evolved several mechanisms to survive during latency (90, 91). Mathematical models suggest that optimal levels of persistence of MTBC in human populations would be achieved with a latency period and the ones currently observed in clinical settings, suggesting that the human immune system keeps the duration of latency and activation rates under their evolutionary optimum (92). Latency seems thus neither uniquely a property of the host or the pathogen, and it has most likely been shaped by the interactions between the two. This is supported by the observation that in the Gambia, patients infected with *M. africanum* Lineage 6 were less likely to progress to active TB than those infected with other MTBC lineages (93). Indeed, the rate of progression to active disease was 5 times higher for Lineage 2 than for *M. africanum* Lineage 6 (93).

Host–pathogen molecular interactions

It is well established that human genetic factors influence clinical TB phenotypes (94, 95). Family aggregation and twin studies have revealed a clear genetic basis for susceptibility to TB infection and disease (reviewed in 95). Furthermore, susceptibility to infection correlates with human ancestry (95–97). However, identifying the human genetic variants underlying the various clinical manifestations of TB has had limited success (reviewed in 94, 95). For example, polymorphisms in the gene *NRAMP1/SLC11A1* have been associated with pulmonary TB (98). Still, despite a strong support for the role of this gene in TB, its effect is highly heterogeneous across different human populations (98–100).

Much of the difficulties in uncovering the human genetic variants underlying responses to MTBC are due to the inability in unambiguously defining and quantifying relevant phenotypes (11, 13, 14). In addition, age, nutritional status, and co-infections can modulate the clinical presentation of TB (101). Finally, co-evolution may also account for some of those difficulties (4, 7). As introduced before, in a co-evolutionary host–pathogen association, the outcome of the interaction between the partners depends on their respective genotypes (5–8). In TB, there is now increasing evidence that the interaction between bacterial and human genetic loci influence clinical phenotypes. For example, one study found that in a Vietnamese population, carrying the T597C allele of the *Toll-like receptor 2 gene (TLR2)* was associated with infection by Lineage 2 (102). Another study from Ghana found that the variant G57E of the *Mannose-binding Lectin (Protein C) 2 gene (MBL2)* was associated with TB caused by *M. africanum* as opposed to *M. tuberculosis sensu stricto* (103). Also in Ghana, the variant 261TT of the *Immunity-related GTPase M gene (IRGM)* was protective against TB caused by Lineage 4, but not against disease caused by other MTBC lineages (104). A recent study reported on the association between different HLA class I types and disease caused by different MTBC strain families in a South African colored population (105). These results suggest that at least to some extent, TB clinical phenotypes can be explained by the interaction between human and MTBC genetic variation. Thus, new insights will be gained when MTBC strain information is accounted for in human genetic studies of TB. Indeed, a recent study in HIV has used a 'genome-to-genome' approach to successfully identify interacting molecular partners in the virus and the human host (106). Co-evolution is thus likely to have important implications for our understanding of the association between humans and MTBC and, more than just stimulating interesting academic discussions, can guide our research toward better tools to control the disease. One particular discussion among academic parties concerns the age of the association between humans and MTBC, which we address in the following section.

The age of the association between MTBC and humans

Particular questions related to the age of this association include how and when the different MTBC lineages emerged and diversified. Moreover, did enough time elapse for differences in virulence phenotypes to evolve, and does the age of MTBC explain its low genetic diversity compared to other

bacteria? Many studies have speculated on the age of the most recent common ancestor (MRCA) of MTBC, but only recently has the subject been addressed in a quantitative way. Four studies have used DNA diversity to infer the age of MTBC (29, 32, 40, 107) (Table 1), while two other studies have inferred the age of particular MTBC variants (23, 108). However, all of these estimates remain controversial. We review here some of these estimates and contrast them with the available archeological evidence. An important point to keep in mind is that using DNA sequences for these types of inferences can only tell us about the MRCA of the DNA sequences used in the analysis. As pointed out by Smith *et al.* (22), if for example the Beijing family had taken over all the remaining lineages in the world, the MRCA of all contemporary MTBC would be a Beijing strain.

What do DNA sequences tell us about the age of MTBC?

Dating the age of bacteria relies on a molecular clock (Table 1), i.e. a way of estimating how much time has elapsed since the corresponding DNA sequences diverged. Dating the MTBC phylogeny has several complications, which the studies we review here tried to circumvent in different ways. It is now widely accepted that there is no universally applicable molecular clock in bacteria, and that the molecular clock of different bacterial species and lineages can differ by orders of magnitude (109, 110). The first study attempting to determine a molecular clock for MTBC was by Wirth *et al.* (32), using microsatellite-like loci with variable numbers of tandem repeats. The common ancestor of MTBC was dated at around 40 000 years ago, and the data supported the divergence of two main lineages around 20 000–30 000 years ago, which co-migrated with modern humans and spread through Africa and Eurasia. The limitation of this study lies on the genetic markers used. Microsatellites are excellent genotyping makers for epidemiological purposes because of their fast rate of evolution, but their molecular evolution patterns cannot be easily extrapolated to the genome as a whole. Therefore, they are of limited use for phylogenetic inferences (111).

More reliable phylogenetic calibration points can be obtained by dating nodes or tips of a phylogenetic tree using ancient bacterial DNA (aDNA), but this was until very recently (see below) not possible for MTBC, because only small fragments of aDNA with little or no phylogenetic information were available. In the absence of calibration points, another possibility is to estimate mutation rates experimentally (112) or from epidemiological outbreaks

(113). Importantly, the relationship between time and substitution rate is poorly understood in MTBC. In other words, the way mutations accumulate during short periods of time, say years to a few decades, is not a good approximation for how mutations accumulate over longer time periods relevant for evolutionary processes (109). Thus, extrapolating from short-term mutation rates to long-term substitution rates is questionable. Moreover, the short-term mutation rates themselves are debatable because of the uncertainty of the *in vivo* generation time and the strong genetic drift associated with population bottlenecks (further discussed below). One alternative way to calibrate phylogenetic splitting events, for which bacterial divergence is strictly a consequence of co-divergence with its host, is to use the host fossil record to date these events. This approach has been particularly powerful in the case of vertical transmitted endosymbiotic bacteria and their insect hosts (114). A similar approach has been used successfully to estimate the age of the association between *Helicobacter pylori* and humans. Even though *H. pylori* is horizontally transmitted, transmission occurs often between close relatives, i.e. from a mother to her newborn child, hence approaching 'vertical' transmission (115). Recently, Comas *et al.* (29) used such an approach to infer the age of the MCRA of MTBC. The starting observation was the striking similarity between the MTBC phylogeny based on 220 full genome sequences representing the global diversity of human-adapted MTBC, and a human phylogeny based on 4955 mitochondrial genomes representing the main human mitochondrial haplogroups (Fig. 3). The authors observed that important branching events coincided in both phylogenies. Specifically, the earliest branching clades had an African origin in both cases, and the branching of Lineage 1 and the Eurasian Lineages 2 to 4 mirrored the main human mitochondrial haplogroups M and N associated with the earliest Out-of-Africa migration of modern humans (Fig. 3). In addition, there were strong quantitative associations between the most frequent MTBC lineages and the most common mitochondrial haplogroups from the same countries. The authors interpreted these observations as evidence for co-divergence, and thus used different events in human evolution to date the MTBC phylogeny (29). Specifically, the coalescent time (Table 1) for the basal MTBC Lineages 5 and 6 was calibrated using three alternative key points in human evolution; the emergence of anatomically modern humans (approximately 185 000 years ago), the emergence with the human mitochondrial DNA haplogroup L3 (approximately 70 000 years ago), and the emergence associated with the beginning of the Neolithic Demographic
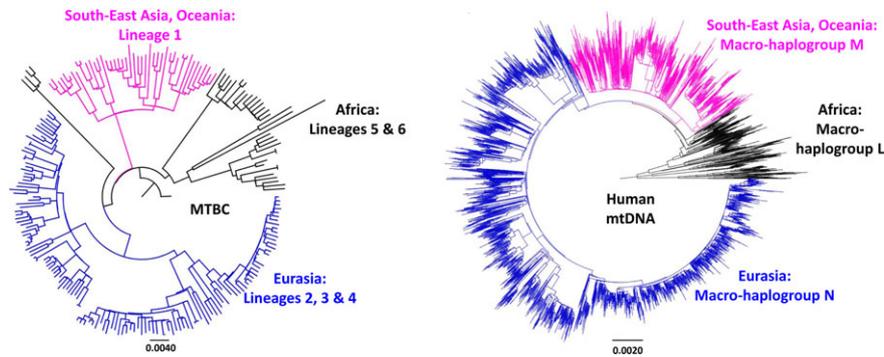
Fig. 3. Comparison of the MTBC phylogeny based on 220 genomes (left) and a phylogeny derived from 4955 mitochondrial genomes representative of the main human mitochondrial haplogroups (right). The color coding highlights the similarities in tree topology, and geographic distribution between MTBC strains and the main human mitochondrial macrohaplogroups (black, African clades: MTBC Lineages 5 and 6, human mitochondrial macrohaplogroups L0–L3; pink, Southeast Asian and Oceania clades: MTBC Lineage 1, human mitochondrial macrohaplogroup M; blue, Eurasian clades: MTBC Lineages 2, 3, and 4, human mitochondrial macrohaplogroup N). MTBC Lineage 7 has only been found in Ethiopia, and its correlation with any of the three main human haplogroups remains unclear. Figure from Comas *et al.* (29).

Transition (approximately 10 000 years ago). In this way, several dates were obtained for the different branching points in the MTBC phylogeny separating the seven known MTBC lineages. Among these three possible scenarios, the one yielding a coalescent time for Lineages 5 and 6 of 70 000 years was considered the most likely, as the MTBC lineage splits could be related to known human migratory events and human archeological evidence. For instance, the first split leading to Lineage 1 was estimated at around 67 000 years ago, coinciding with the first wave of human migrations Out-of-Africa (29). The split leading to Lineages 2 and 4 was estimated at 30 000–46 000 years ago and at 32 000–42 000 years ago, respectively, which correlates well with the first archeological evidence for the presence of modern humans in Europe and East Asia, respectively (116, 117). The age of the MRCA of all members of the complex was thus estimated at 73 000 years (range 50 000–96 000 years ago). Although only based on correlations, the scenario proposed by this study offers plausible explanations for the phylogeographic structure observed in MTBC (Fig. 1). It has one important limitation, which is the *a priori* assumption of co-divergence between the MTBC and humans. Yet, mirrored phylogenies between hosts and their pathogens are not necessarily the result of co-divergence, as switching between closely related hosts and extinctions of certain pathogen lineages can lead to similar phenomena (118, 119). As outlined above, extant human-adapted MTBC are obligate human pathogens not known to sustainably infect any other host species and with no known environmental reservoir, but whether that has always been the case throughout the evolution of MTBC is unknown. Thus, for horizontally transmitted pathogens such as MTBC, ideally,

the coinciding splitting events of both the host and the pathogen lineages should be demonstrated independently. This has been tested in the work by Pepperell *et al.* (107), in which a phylogeny of global human populations based on the Y chromosome (120) and a phylogeny of 48 global MTBC genome sequences were compared. Unlike Comas *et al.* (29), these authors did not find evidence for congruence between the human and the MTBC tree topologies (Fig. 4). It is known that human phylogenies based on Y chromosomes and mitochondrial DNAs can provide evidence for different human evolutionary histories even in the same geographic regions, but broad features of human evolution are consistent between the two markers (120). Accordingly, it appears that in general, the similarity between the topologies of the human phylogeny based on Y chromosomes and MTBC is not as negligible as originally suggested by Pepperell *et al.* (107) (Fig. 4A and C). In Fig. 4A, we added to the internal nodes of the MTBC phylogeny by Pepperell *et al.* (107), the origin of the most recent common ancestor of each main lineage as determined by Comas *et al.* (29) (Fig. 3). This is justified, as patterns consistent with co-divergence are more likely to be found when considering major splits that happened in a more distant evolutionary past rather than considering more recent splitting events such as the ones reflected on the tips of the phylogeny. For example, Lineage 2 strains can be found in South Africa; however, because these strains were most likely introduced by recent (in evolutionary time) human migrations from Asia (121), it is unlikely that a signature of co-divergence will be observed between South African Lineage 2 strains and South African human populations. In addition, based on the original publication (120), we indicated in the human

phylogenetic tree the Y haplogroups relevant for this discussion (Fig. 4C). We highlight here what we consider to be congruent among Figs 3 and 4: (i) a strong concordance in the split of the early branches that are restricted to Africa (Lineages 5 and 6 versus Y chromosome haplogroup A and mitochondrial macrohaplogroup L); (ii) concordance in the split that originated MTBC Lineage 1 and the Y chromosome haplogroup C and the mitochondrial macrohaplogroup M that are found in East Asia, South Asia, and Oceania (120); and (iii) the split of both the Y chromosome haplogroup F and the mitochondrial macrohaplogroup N is in concordance with the split of the modern Eurasian MTBC Lineages 2, 3, and 4.

Important differences between the results of Comas *et al.* (29) and those of Pepperell *et al.* (107) reside in the approach taken to infer the age of the most recent common ancestor of MTBC. In Pepperell *et al.* (107), the rate of evolution of MTBC was determined using two calibration points which are marked in Fig. 4. The first calibration point was applied to eight MTBC strains isolated from Western Canadian Aboriginal TB patients, strains which according to historical records have been introduced from Europe between 1710 and 1870 C.E. through the fur trade (122). The second calibration point was applied to the divergence point between the laboratory strain H37Rv and a closely related Canadian *M. tuberculosis* clinical isolate. The date for the divergence between these two isolates was set at 1905, the date of the original isolation of H37Rv from a patient in New York (123). Using this approach, the authors estimated a substitution rate of 1.3×10^{-7} substitutions/site/year, which is much higher than the one estimated by Comas *et al.* (2.58×10^{-9} substitutions/site/year), and close to the mutation rate (0.2–0.5 SNPs per year) inferred from molecular epidemiology studies (113, 124, 125). This higher substitution rate translates necessarily in a much more recent common origin for all MTBC lineages (Table 2). To test more formally for co-divergence, Pepperell *et al.* (107) correlated divergence time among different MTBC estimates with human estimates of genetic distance between continental population and concluded that this correlation was too
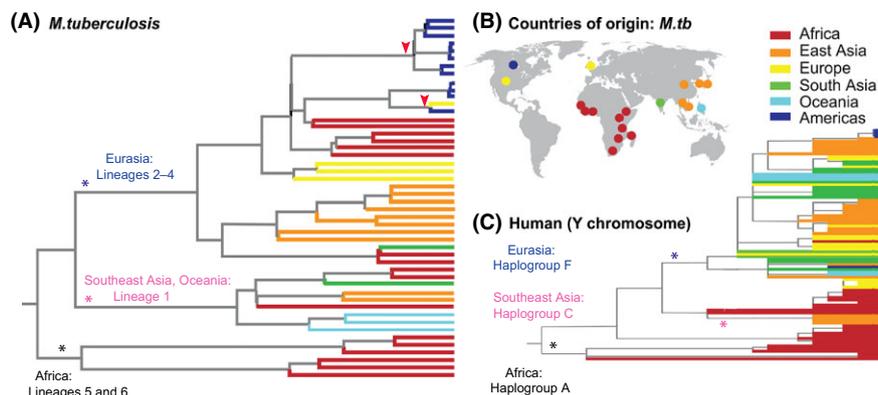

Fig. 4. Geographic and genetic structure of a global sample of MTBC genomes. Adapted from Pepperell *et al.* (107). (A) Maximum clade credibility phylogeny inferred from genome-wide MTBC SNP data. Tips are colored by the geographic origin of the MTBC isolate (see key). Modifications: the lineage nomenclature (indicated with an asterisk) and color codes used by Comas *et al.* (29) was added to the original tree; the red arrows indicate the calibration points used for dating the most recent common ancestor of MTBC. (B) Countries of origin for MTBC isolates used in this study are shown as colored dots on the global map. One dot is shown per country but some countries were represented by >1 MTBC isolate. Colors correspond to global regions (see key). (C) Phylogeny of global human populations based on Y chromosome data (120). Tips are colored according to the same scheme as the MTBC phylogeny (A). Modifications: the names of the relevant human Y haplogroups were added (indicated with an asterisk) and color coded as in Comas *et al.* (29).

Table 2. Comparison of different dating scenarios for the evolution of MTBC

Coalescent times	Comas *et al.* (29)	Pepperell *et al.* (107)	Bos *et al.* (40)
MRCA of MTBC	73 000 (50 000–96 000)	n.r.	(2962–5339)
Lineage 5 and 6	70 000 (48 000–88 000)	2190 (1331–3142)	(2792–5025)
Lineage 1	67 000 (46 000–88 000)	2190 (1331–3142)	(2765–5105)
Lineages 2–4	46 000 (31 000–61 000)	1347 (830–1907)	(1779–3422)

Coalescent times (*Table 1*) are given in thousands of years. The median and/or the 95% HPD (highest probability density) intervals are indicated; n.r., non-reported.

low ($r^2 = 0.46$) to support co-divergence. We agree this is an important analysis, but it should be repeated using more isolates spanning more of the global diversity of MTBC.

In another approach, Bos *et al.* (40) used the *M. pinnipedii*-related aDNA samples from Peru mentioned above (radiocarbon dated to between AD 1028 and AD 1280) to calibrate a global MTBC phylogeny. The substitution rate obtained was 4.6×10^{-8} substitutions/site/year, again much higher than the estimate by Comas *et al.* (29) and closer to the one by Pepperell *et al.* (107) and the various epidemiological studies (113, 124, 125). The MRCA of MTBC was inferred at 4449 years before present, similar to the estimate by Pepperell *et al.* (107) (Table 2). The same work (40) corroborated these recent time estimates using an independent calibration based on a 18th century Hungarian mummy from which MTBC DNA was isolated and genome sequenced (126). The year of the supposed death of the individual carrying the bacillus from which the DNA was isolated (1797 AD) was used (126). The substitution rate was only slightly higher (7.07×10^{-8} substitutions/site/year) than the one obtained with the Peruvian aDNA (40).

The estimates by Pepperell *et al.* (107) and the ones obtained with the Hungarian mummy DNA (40) have some technical limitations. Substitution rates (Table 1) are known to be time-dependent (127, 128). Estimates based on transient mutations (i.e. polymorphisms) such as the ones obtained in epidemiological contexts or by dating the tips of phylogenetic trees built with extant or very recent strains in evolutionary terms, are necessarily higher than estimates based on fixed mutation (i.e. substitutions) because a high proportion of polymorphisms is eliminated by natural selection and genetic drift over time (26, 107). The relative contribution to the evolution of MTBC of natural selection and genetic drift is discussed further below. The important point in the context of dating MTBC phylogenies is that in MTBC, we do not know how far back in time polymorphisms can persist. Using the approximately 1000-year-old DNA to calibrate the MTBC phylogeny should lessen this problem (40). However, the similarity of the substitution rate estimates obtained using this approach and the ones obtained through epidemiological studies is not very reassuring. An additional problem is that the different MTBC lineages most likely evolve at different rates (40), meaning that the molecular clock inferred with the Peruvian aDNA does not necessarily reflect the molecular clock of the different MTBC lineages. To resolve more confidently the age of the most recent common ancestor of MTBC and of its main lineages, other calibration points, preferably based on aDNA from older MTBC strains will be necessary.

The estimates by Bos *et al.* (40) suggest that the ancestors of all current MTBC lineages could have originated as recently as 3000 BCE to 2000 BCE and that the common ancestor of all modern lineages appeared only between 1500 BCE to 200 CE (Table 2, Fig. 5). How then to explain the similarity between the human and the MTBC phylogenies as described in Comas *et al.* (29)? Moreover, how did the modern lineages spread Out-of-Africa to the rest of the world? Did they emerge as one Out-of-Africa lineage, spreading and diversifying through the Indo-European migrations that between 4000 and 1000 BCE connected Asia Minor, Europe, Central Asia, and India (129)? Could the human movements associated with the Silk Road linking different parts of Asia, India, and the Mediterranean region have spread Lineage 2 and Lineage 3 strains (130) (Fig. 5)? These more recent estimates also imply that MTBC has diversified and adapted to a whole range of distinct mammalian species (Fig. 2) during a period of only about 4000 years. Archeology from human remains coupled with molecular biological analyses of mycobacterial aDNA has contributed significantly to our understanding of the association of MTBC and ancient human populations (131, 132). We briefly summarize some of those findings next.

What does the archeological evidence tell us?

In Fig. 5, we summarize the most significant archeological contributions based on bone lesions consistent with TB and molecular evidence for the presence of MTBC in archeological specimens. We briefly discuss here the inconsistencies between the archeological findings and the age estimates based on DNA sequence divergence discussed above. The estimates by Comas *et al.* (29) predate all archeological evidence known to date and cannot thus be refuted. In contrast, there are inconsistencies with the dating estimates obtained by Bos *et al.* (40) and Pepperell *et al.* (107). The oldest known biomolecular evidence for MTBC dates to about 17 500 years before present and was found in a bison bone discovered in Wyoming, USA (133) (Fig. 5). This finding predates the earliest estimates for the age of the MRCA of MTBC by Bos *et al.* (40) (Fig. 5). The molecular evidence indicative of MTBC aDNA was the amplification of IS6110 and presence of spoligotyping patterns, both of which are specific to MTBC today (134). However, the insertion element IS6110 is a mobile genetic element, we can therefore not be sure of the specificity of this element in the past (135). Similarly, the genomic region used for
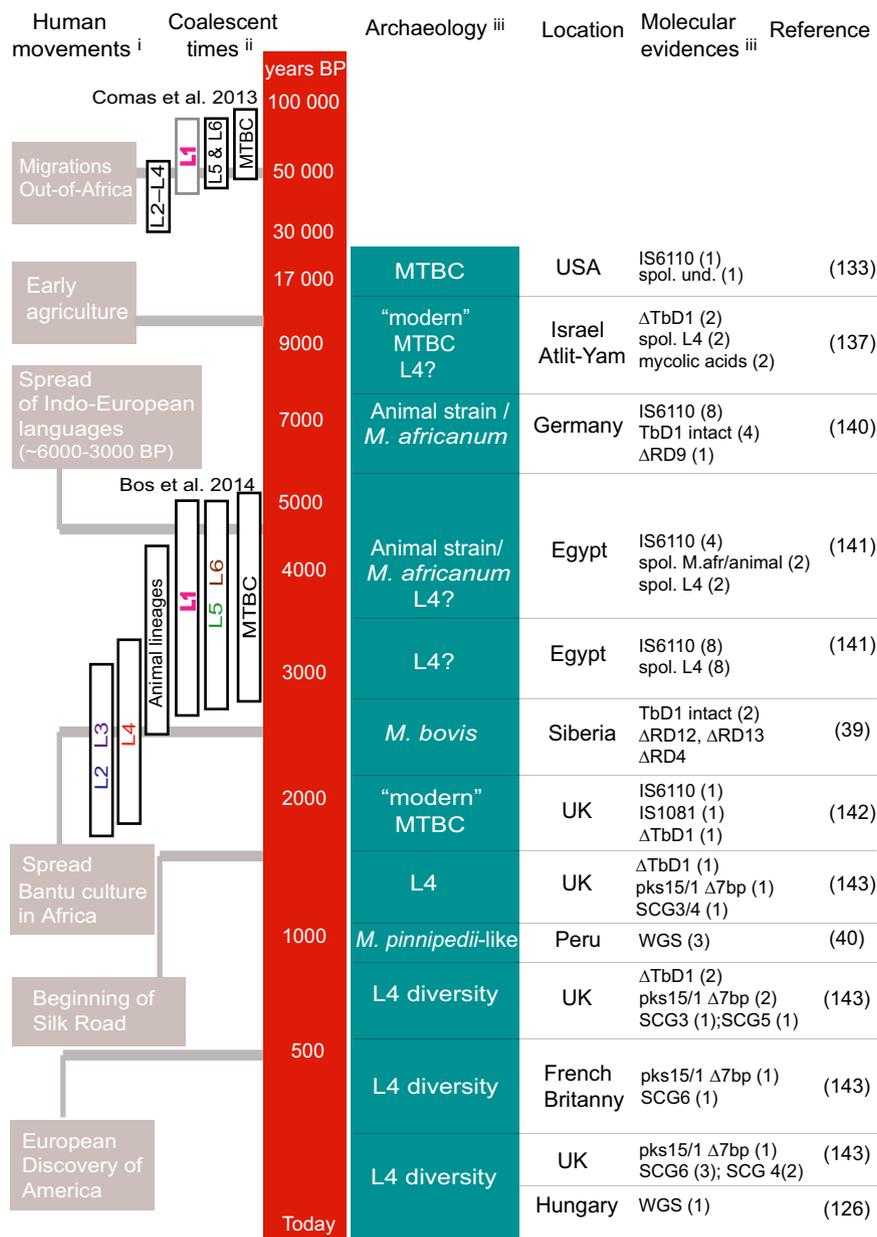
Fig. 5. Summary of the most important archeological findings and molecular evidence indicating the presence of MTBC in human remains. All studies reported MTBC in human skeletons or mummies except in Rothschild *et al.* (133). (i) Human movements possibly involved in the expansion of MTBC; (ii) Intervals for coalescent times (95% HPD) concerning the MRCA of different MTBC lineages obtained from the two most discrepant inferences (29, 40); (iii) Interpretation of both archeological and biomolecular findings. For a better understanding of the molecular typing techniques see Coscolla and Gagneux (15). For each molecular marker, the number of archeological specimens containing DNA with that specific marker is indicated within parentheses. BP, before present; spol., spoligotyping pattern; und., undefined; WGS, whole-genome sequencing.

spoligotyping is known as the CRISPR-Cas region which is prone to horizontal gene transfer in other bacteria (136).

Another archeological finding inconsistent with the age of the MRCA of all modern lineages inferred by Bos *et al.* (40) comes from one report on TB caused by modern strains of MTBC in an approximately 9000 years old human skeletons from Israel (137) (Fig. 5). The strain in question harbored the TbD1 deletion, spoligotyping patterns could be obtained, and the presence of mycolic acids was also

detected. However, the authenticity of some of this material has been the subject of dispute (138, 139). Generally, studies using aDNA are prone to contamination, and many precautions need to be taken for generating convincing evidence (131).

The earliest estimate for the age of the MRCA of all animal-adapted lineages of MTBC was set at around 4000 years by Bos *et al.* (40) (Fig. 5). The oldest archeological evidence for an animal MTBC strain has been reported from a human

skeleton from Germany, which presented evidence of TB caused by a strain with an RD9 deletion dated to around 7000 years ago (140) (Fig. 5). Data from other human skeletons from the same site also suggested that as early as 5000 BCE, MTBC was already infecting humans in Central Europe (140) (Fig. 5), an observation that does not fit with the age of modern MTBC lineages proposed by Bos *et al.* (40). Alternatively, other MTBC strains might have existed in Europe and been later displaced later by modern strains. Modern strains might have been already present 4000 years ago in Egyptian mummies (141), and these strains were also detected as causing TB in the United Kingdom during the Iron Age 2200 years ago (142) (Fig. 5). Another study suggests that *M. bovis* had reached Siberia by 2000 years ago (39). Interestingly, more recent studies have achieved more precise classification of ancient MTBC strains (40, 126, 143), which indicates that, as mentioned before, strains similar to extant *M. pinnipeddi* infected humans in Peru around 1000 years ago (40), and around the same time, different genotypes of Lineage 4 were present in the United Kingdom (143) (Fig. 5). The relative abundance of MTBC DNA in archeological human remains, together with our increased ability to sequence reliably aDNA, will certainly shed further light on the co-evolutionary history of MTBC and humans in the near future. A complementary approach to study the co-evolutionary history of humans and their pathogens is to determine how the genetic variation that characterizes the extant populations of both has been shaped over time (4, 7).

Co-evolutionary genetics: the MTBC perspective

In a co-evolutionary association, molecular interactions between the host and the pathogen are expected (7), such as the ones involving host immune receptors and pathogen surface proteins. For example, if a mutation encodes a certain molecular variant that allows a pathogen to avoid host recognition, the frequency of that mutation will increase. Likewise, mutations in a host leading to better recognition of that pathogen variant will be selected. Over evolutionary time, such dynamics lead to distinctive signatures of selection in the genetic diversity of the affected loci of both the host and pathogen populations (4, 7). The interpretation of these signatures can give us insights into which loci are affected and how they evolve. However, natural host–pathogen associations show us that co-evolving hosts and pathogens have many additional characteristics which complicate our assessment of the relevant loci involved in these interactions, and the relevant time and geographic scales to be

considered (3, 7). For example, humans have long generation times compared to their pathogens and the molecular variation relevant for the recognition of the pathogen may be somatically generated *de novo* (adaptive immune system) and will thus not be encoded in their DNA. Alternatively, pathogens may evolve strategies other than avoiding being recognized by the host. In MTBC, this seems to be the case as suggested by one study which studied variation in 491 experimentally confirmed T-cell epitopes in representatives of several lineages of the complex (27). Surprisingly, that study found that the large majority of these T-cell epitopes had high sequence conservation, indicating that there is a strong selection pressure to keep these T-cell epitopes unchanged (27, 107). One possible reason for this high conservation of T-cell epitopes is that MTBC might benefit from being recognized by T cells because, as referred above, the microbe depends on an efficient host immune response leading to increased lung pathology to transmit to new hosts (27). Support for this view comes from the observation that HIV co-infected TB patients tend to generate fewer secondary cases compared to HIV uninfected patients (144); this is particularly true for HIV/TB patients with low CD4$^+$ T cells (145). Interestingly, sequence variation in T-cell epitopes is also low in *M. canettii* (57) and in BCG strains, which evolved during *in vitro* passage (10), suggesting that recognition by T cells is probably not the only reason underlying epitope conservation in MTBC.

In addition to mutations causing polymorphisms in populations, duplication, and expansion of gene families has evolved in almost all organisms as a way to generate genetic diversity (146). In particular, multi-copy gene families have evolved as a strategy to generate antigenic variation in different pathogens; classical examples include the *var* genes of *Plasmodium falciparum* and *pil* genes of *Neisseria meningitis* (147). In MTBC and other virulent mycobacteria, the *pe_pgrs* genes underwent extensive duplication (148). Moreover, these genes harbor higher diversity than other parts of the MTBC genome (149, 150). However, similar to the hyperconservation of T-cell epitopes encoded by other parts of the MTBC genome (27), the regions of *pe_pgrs* genes that encode T-cell epitopes are mostly confined to the PE domains, which exhibit high sequence conservation when comparing different clinical strains of MTBC (10). Thus, the diversity of *pe_pgrs* genes cannot be explained by evasion of T-cell recognition. Perhaps, the importance of the genetic diversity of *pe_pgrs* genes lies not so much on the diversity generated at the population level but rather on the diversity generated within each bacterial cell, for instance by differential expres-

sion of these genes under different conditions. Another possibility is that the diversity of *pe_pgrs* genes might be driven by host antibody recognition (10). In summary, there is currently no evidence for immune evasion through antigenic variation in the genomes of MTBC.

Is there any other evidence of adaptation in MTBC genomes? The emergence of drug resistance clearly illustrates that MTBC has the potential to adapt. Yet, except in the context of drug resistance, the current evidence of ongoing adaptive changes in MTBC is scarce. One important limitation to adaptive evolution in MTBC might be the lack of recombination and horizontal gene transfer (151). Recombination reduces the extent to which the effect of natural selection on new mutations is reduced by the effect of selection imposed by the genetic background (background selection, Table 1), i.e. recombination can separate advantageous and neutral mutations from harmful mutations (152). In non-recombining chromosomes, advantageous mutations might get 'trapped' with mutations that are slightly deleterious, in which case, selection on the advantageous mutation has to be strong for it to be selected (152).

Nevertheless, evidence for adaptation is emerging as more full genomes of MTBC clinical strains become available for comparative analyses. A recent analysis by Osorio *et al.* (153), reported two genes unrelated to drug resistance showing evidence of positive selection in MTBC. The study by Gonzalo-Asensio *et al.* (54) described above on the molecular basis of virulence in Lineage 6 and animal-adapted lineages presents compelling molecular evidence for adaptation. From a more speculative point of view, it is also plausible that MTBC and humans might be close to an evolutionary 'optimum'. This would imply that the human-adapted MTBC lineages are generally well adapted to their host and *vice versa*. If so, most mutations responsible for host adaptation would already have reached fixation. New mutations will mostly be neutral or slightly deleterious and hence assume the form of transient polymorphisms, which end up disappearing from the population over time. Such a scenario has been suggested for other genetically monomorphic bacterial pathogens (154). Thus, most of the main reciprocal adaptations between MTBC and humans would have happened in the distant evolutionary past. This does not mean that MTBC and humans do not co-evolve anymore, but adaptation now might involve much smaller effects. Some indirect support for this idea comes from two studies of ancient and historical patterns of alleles conferring human resistance to TB. One concluded that the dramatic decline of TB in Europe starting in the mid-nineteenth century, long

before modern interventions such as BCG vaccination and antibiotic treatment were introduced, cannot be the result of selection for higher genetic resistance to TB (155). Contrastingly, and supporting the view that the bulk of co-evolution occurred in the evolutionary past, a study by Barnes *et al.* (156), suggested that the time since ancient urbanization in the Old World (the earliest settlement considered was 6000 BCE) can explain TB genetic resistance in today's urban human populations.

Alternatively, it has been hypothesized that infection by MTBC might have been beneficial to early human populations during times of nutrient shortage (157) and/or by providing some level of immunological protection against other pathogens (29). It is provocative to propose that a pathogen killing 1.3 million people every year might have reached some sort of evolutionary optimum or might even offer a net benefit to infected humans. At the same time, it is clear that only a small proportion of the more than 2 billion individuals estimated latently infected individuals whose immune system is able to control the infection will never die of or exhibit any symptoms of TB.

Genetic drift versus adaptation

It has been proposed that the evolutionary trajectory of new mutations in MTBC might be mostly governed by genetic drift, negative selection, and background selection (26, 60, 107) (Table 1), except in context of drug treatment or HIV co-infection (158). Generally, due to their (on average) negative functional effects, non-synonymous single nucleotide mutations are removed more efficiently by natural selection than synonymous mutations (generally considered to be neutral). In MTBC, a high proportion of non-synonymous mutations are maintained across almost all genes categories relative to the corresponding synonymous mutations (26, 107). Moreover, close to half of these non-synonymous mutations have been predicted to impact protein function, presumably mainly in a negative way (26, 159). These phenomena are consistent with an important role for genetic drift (i.e. 'chance', Table 1) in the evolution of MTBC. This is likely, given the strong sequential population bottlenecks (Table 1) during patient-to-patient transmission, where new infections are initiated by only a single or a few bacterial cells (160). The effects of genetic drift on the evolutionary trajectory of mutations can be very strong, because drift reduces the effective population size of a population (the number of individuals that contribute to the progeny of the following generation), decreasing the efficiency of natural selection in removing deleterious mutations.

Moreover, mutations that are neutral or have only a small advantageous effect on fitness can be lost by chance. The notion that genetic drift plays an important role in MTBC is not to say that selection is completely absent. Indeed, as discussed above, some signatures of positive selection (i.e. adaptation) have been observed.

Concluding remarks and outlook

The observations we reviewed here favor the hypothesis of co-evolution between humans and certain MTBC lineages. Humans differ in their genetic susceptibility to TB, and the genetic diversity in MTBC is also increasingly recognized as a factor influencing the outcome of TB infection and disease. Moreover, recent studies have reported interactions between human genetic diversity and bacterial diversity, reinforcing the view that in addition to environmental variables, TB susceptibility varies as a function of both the human and MTBC genotype. The age of the association between humans and MTBC is subject to intense debate, but irrespective of the exact duration of this association, different phylogenetic lineages of MTBC are associated with different geographic regions, and possibly adapted to specific human populations.

In addition, some lineages seem to behave in a generalist way, while others appear to be specialists restricted to a smaller group of human host populations.

Co-evolution is defined as reciprocal adaptive changes in interacting host and pathogen genetic loci. Conceptually, these interacting loci can be seen as in 'conflict' with each other (161). Identifying these conflicting host and pathogen loci could shed new light onto the host–pathogen interaction by identifying the actual molecular determinants involved in this interaction. Moreover, from a vaccinology point of view, these conflicting loci might be attractive new vaccine targets, because the host immune responses they elicit will most likely be detrimental to MTBC. This is in contrast to the highly conserved T-cell epitopes that elicit immune responses that might offer a net benefit to the bacteria (27). Hence, in addition to being an interesting academic subject of discussion, a better understanding of the co-evolutionary history of MTBC and *H. sapiens* could be exploited in vaccine development. For example, as illustrated in the recent genome-to-genome study of HIV mentioned above (106), a similar approach could be used to identify new protective antigens for TB vaccines.

References

1. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc R Soc Lond B Biol Sci 2002;**269**:931–936.
2. Betts A, Kaltz O, Hochberg ME. Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. Proc Natl Acad Sci USA 2014;**111**:11109–11114.
3. Sorci G, Moller AP, Boulinier T. Genetics of host-parasite interactions. Trends Ecol Evol 1997;**12**:196–200.
4. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet 2014;**15**:379–393.
5. Anderson RM, May RM. Coevolution of hosts and parasites. Parasitology 1982;**85**(Pt 2):411–426.
6. May RM, Anderson RM. Epidemiology and genetics in the coevolution of parasites and hosts. Proc R Soc Lond B Biol Sci 1983;**219**:281–313.
7. Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet 2002;**32**:569–577.
8. Hamilton WD. Sex versus non-sex versus parasite. Oikos 1980;**35**:282–290.
9. WHO. *Global Tuberculosis Report* 2013. Geneva, Switzerland: World Health Organization, 2013.
10. Copin R, Coscolla M, Efstathiadis E, Gagneux S, Ernst JD. Impact of in vitro evolution on antigenic diversity of Mycobacterium bovis bacillus Calmette-Guerin (BCG). Vaccine 2014;**32**:5998–6004.

11. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MP. The immune response in tuberculosis. Annu Rev Immunol 2013;**31**:475–527.
12. Ernst JD. The immunological life cycle of tuberculosis. Nat Rev Immunol 2012;**12**:581–591.
13. Esmail H, Barry CE 3rd, Young DB, Wilkinson RJ. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci 2014;**369**:20130437.
14. Young DB, Gideon HP, Wilkinson RJ. Eliminating latent tuberculosis. Trends Microbiol 2009;**17**:183–188.
15. Coscolla M, Gagneux S. Does M. tuberculosis genomic diversity explain disease diversity? Drug Discov Today 2010;**7**:e43–e59.
16. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. Semin Immunol 2014;**26**:441–444.
17. Comas I, Gagneux S. The past and future of tuberculosis research. PLoS Pathog 2009;**5**: e1000600.
18. Young D, Verreck FA. Creativity in tuberculosis research and discovery. Tuberculosis 2012;**92** (Suppl 1):S14–S16.
19. Smith NH, et al. Ecotypes of the Mycobacterium tuberculosis complex. J Theor Biol 2006;**239**:220–225.
20. Brosch R, et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA 2002;**99**:3684–3689.

21. Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr MA. Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. J Infect Dis 2002;**186**:74–80.
22. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson G. Bottlenecks and broomsticks: the molecular evolution of Mycobacterium bovis. Nature Rev Microbiol 2006;**4**:670–681.
23. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci USA 2004;**101**:4871–4876.
24. Gagneux S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci USA 2006;**103**:2869–2873.
25. Tsolaki AG, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc Natl Acad Sci USA 2004;**101**:4865–4870.
26. Hershberg R, et al. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 2008;**6**: e311.
27. Comas I, et al. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat Genet 2010;**42**:498–503.
28. Firdessa R, et al. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg Infect Dis 2013;**19**:460–463.
29. Comas I, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium

tuberculosis with modern humans. Nat Genet 2013;**45**:1176–1182.

30. de Jong BC, Antonio M, Gagneux S. Mycobacterium africanum–review of an important cause of human tuberculosis in West Africa. PLoS Negl Trop Dis 2010;**4**:e744.

31. Coscolla M, et al. Novel Mycobacterium tuberculosis complex isolate from a wild chimpanzee. Emerg Infect Dis 2013;**19**:969–976.

32. Wirth T, et al. Origin, spread and demography of the Mycobacterium tuberculosis complex. PLoS Pathog 2008;**4**:e1000160.

33. KoroKoro F, et al. Population dynamics of tuberculous Bacilli in Cameroon as assessed by spoligotyping. J Clin Microbiol 2012;**51**:299–302.

34. Wampande EM, et al. Long-term dominance of Mycobacterium tuberculosis Uganda family in peri-urban Kampala-Uganda is not associated with cavitary disease. BMC Infect Dis 2013;**13**:484.

35. Paulsen HJ. Tuberculosis in the native American: indigenous or introduced? Rev Infect Dis 1987;**9**:1180–1186.

36. Kirzinger MW, Stavrinides J. Host specificity determinants as a genetic continuum. Trends Microbiol 2012;**20**:88–93.

37. Rivero A, et al. High rate of tuberculosis reinfection during a nosocomial outbreak of multidrug-resistant tuberculosis caused by Mycobacterium bovis strain B. Clin Infect Dis 2001;**32**:159–161.

38. Ameni G, et al. Mycobacterium tuberculosis infection in grazing cattle in central Ethiopia. Vet J 2011;**188**:359–361.

39. Taylor GM, Murphy E, Hopkins R, Rutland P, Chistov Y. First report of Mycobacterium bovis DNA in human remains from the Iron Age. Microbiology 2007;**153**:1243–1249.

40. Bos KI, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 2014;**514**:494–497.

41. Mackowiak PA, Blos VT, Aguilar M, Buikstra JE. On the origin of American tuberculosis. Clin Infect Dis 2005;**41**:515–518.

42. Berg S, Smith NH. Why doesn't bovine tuberculosis transmit between humans? Trends Microbiol 2014;**22**:552–553.

43. Smith NH, Crawshaw T, Parry J, Birtles RJ. Mycobacterium microti: more diverse than previously thought. J Clin Microbiol 2009;**47**:2551–2559.

44. Wagner JC, Buchanan G, Bokkenheuser V, Leviseur S. An acid-fast bacillus isolated from the lungs of the Cape hyrax, Procavia capensis (Pallas). Nature 1958;**181**:284–285.

45. Mostowy S, Cousins D, Behr MA. Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the Mycobacterium tuberculosis complex. J Bacteriol 2004;**186**:104–109.

46. Schoepf K, et al. A two-years' survey on the prevalence of tuberculosis caused by Mycobacterium caprae in red deer (Cervus elaphus) in the Tyrol, Austria. ISRN Vet Sci 2012;**2012**:245138.

47. Chiari M, et al. Isolation of Mycobacterium caprae (Lechtal genotype) from red deer (Cervus elaphus) in Italy. J Wildl Dis 2014;**50**:330–333.

48. Jenkins HE, Cox DR, Delahay RJ. Direction of association between bite wounds and Mycobacterium bovis infection in badgers: implications for transmission. PLoS ONE 2012;**7**:e45584.

49. van Ingen J, et al. Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. Emerg Infect Dis 2012;**18**:653–655.

50. Gey van Pittius NC, van Helden PD, Warren RM. Characterization of Mycobacterium orygis. Emerg Infect Dis 2012;**18**:1708–1709.

51. Alexander KA, et al. Novel Mycobacterium tuberculosis complex pathogen, M. mungi. Emerg Infect Dis 2010;**16**:1296–1299.

52. Parsons DC, Drewe JA, Gey van Pittius NC, Warren RM, van Helden PD. Novel cause of tuberculosis in Meerkats. South Africa. Emerg Infect Dis 2013;**19**:2004–2007.

53. Gutierrez MC, et al. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 2005;**1**:e5.

54. Gonzalo-Asensio J, et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci USA 2014;**111**:11491–11496.

55. Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I. The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol 2006;**60**:312–330.

56. van Soolingen D, et al. A novel pathogenic taxon of the Mycobacterium tuberculosis complex, Canetti: characterization of an exceptional isolate from Africa. Int J Syst Bacteriol 1997;**47**:1236–1245.

57. Supply P, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 2013;**45**:172–179.

58. Veyrier FJ, Dufort A, Behr MA. The rise and fall of the Mycobacterium tuberculosis genome. Trends Microbiol 2011;**19**:156–161.

59. Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O. Horizontally acquired genomic islands in the tubercle bacilli. Trends Microbiol 2008;**16**:303–308.

60. Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EP. After the bottleneck: genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. Genome Res 2012;**22**:721–734.

61. Blouin Y, et al. Progenitor "Mycobacterium canettii" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg Infect Dis 2014;**20**:21–28.

62. Koeck JL, et al. Clinical characteristics of the smooth tubercle bacilli 'Mycobacterium canettii' infection suggest the existence of an environmental reservoir. Clin Microbiol Infect 2010;**17**:1013–1019.

63. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, Neyrolles O. Horizontal transfer of a virulence operon to the ancestor of Mycobacterium tuberculosis. Mol Biol Evol 2006;**23**:1129–1135.

64. Cole ST, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;**393**:537–544.

65. Uplekar S, Heym B, Friocourt V, Rougemont J, Cole ST. Comparative genomics of Esx genes from clinical isolates of Mycobacterium tuberculosis provides evidence for gene conversion and epitope variation. Infect Immun 2011;**79**:4042–4049.

66. Ramaswamy V, et al. Listeria–review of epidemiology and pathogenesis. J Microbiol Immunol Infect 2007;**40**:4–13.

67. Salinas C, et al. [Longitudinal incidence of tuberculosis in a cohort of contacts: factors associated with the disease]. Arch Bronconeumol 2007;**43**:317–323.

68. Rodrigo T, et al. Characteristics of tuberculosis patients who generate secondary cases. Int J Tuberc Lung Dis 1997;**1**:352–357.

69. Rouillon A, Perdrizet S, Parrot R. Transmission of tubercle bacilli: the effects of chemotherapy. Tubercle 1976;**57**:275–299.

70. Rose CE Jr, Zerbe GO, Lantz SO, Bailey WC. Establishing priority during investigation of tuberculosis contacts. Am Rev Respir Dis 1979;**119**:603–609.

71. Capewell S, Leitch AG. The value of contact procedures for tuberculosis in Edinburgh. Br J Dis Chest 1984;**78**:317–329.

72. Clark M, Riben P, Nowgesic E. The association of housing density, isolation and tuberculosis in Canadian First Nations communities. Int J Epidemiol 2002;**31**:940–945.

73. Anderson RM, May RM. Population biology of infectious diseases: part I. Nature 1979;**280**:361–367.

74. Levin BR. The evolution and maintenance of virulence in microparasites. Emerg Infect Dis 1996;**2**:93–102.

75. Kawecki TJ, Ebert D. Conceptual issues in local adaptation. Ecol Lett 2004;**7**:1225–1241.

76. Blanquart F, Kaltz O, Nuismer SL, Gandon S. A practical guide to measuring local adaptation. Ecol Lett 2013;**16**:1195–1205.

77. Gandon S, Van Zandt PA. Local adaptation and host-parasite interactions. Trends Ecol Evol 1998;**13**:214–216.

78. Reed MB, et al. Major Mycobacterium tuberculosis lineages associate with patient country of origin. J Clin Microbiol 2009;**47**:1119–1128.

79. Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. Emerg Infect Dis 2004;**10**:1568–1577.

80. Fenner L, et al. HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. PLoS Genet 2013;**9**:e1003318.

81. Pasipanodya JG, et al. Allopatric tuberculosis host-pathogen relationships are associated with greater pulmonary impairment. Infect Genet Evol 2013;**16**:433–440.

82. Groenheit R, et al. The Guinea-Bissau family of Mycobacterium tuberculosis complex revisited. PLoS ONE 2011;**6**:e18601.

83. Niobe-Eyangoh SN, et al. Genetic biodiversity of Mycobacterium tuberculosis complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol 2003;**41**:2547–2553.

84. Godreuil S, et al. First molecular epidemiology study of Mycobacterium tuberculosis in Burkina Faso. J Clin Microbiol 2007;**45**:921–927.

85. de Jong BC, et al. Differences between tuberculosis cases infected with Mycobacterium africanum, West African type 2, relative to Euro-American Mycobacterium tuberculosis: an update. FEMS Immunol Med Microbiol 2010;**58**:102–105.

86. Bold TD, Davis DC, Penberthy KK, Cox LM, Ernst JD, de Jong BC. Impaired fitness of Mycobacterium africanum despite secretion of ESAT-6. J Infect Dis 2012;**205**:984–990.

87. de Jong BC, et al. Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J Infect Dis 2006;**193**:1279–1286.

88. Pai M, Zwerling A, Menzies D. Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med 2008;**149**:177–184.

89. Blaser MJ, Kirschner D. The equilibria that allow bacterial persistence in human hosts. Nature 2007;**449**:843–849.

90. Gengenbacher M, Kaufmann SH. Mycobacterium tuberculosis: success through dormancy. FEMS Microbiol Rev 2012;**36**:514–532.

91. Ehrt S, Rhee K. Mycobacterium tuberculosis metabolism and host interaction: mysteries and paradoxes. Curr Top Microbiol Immunol 2012;**374**:163–188.

92. Zheng N, Whalen CC, Handel A. Modeling the potential impact of host population survival on the evolution of M. tuberculosis latency. PLoS ONE 2014;**9**:e105721.

93. de Jong BC, et al. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. J Infect Dis 2008;**198**:1037–1043.

94. Moller M, Hoal EG. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis 2010;**90**:71–83.

95. Abel L, El-Baghdadi J, Bousfiha AA, Casanova JL, Schurr E. Human genetics of tuberculosis: a long and winding road. Philos Trans R Soc Lond B Biol Sci 2014;**369**:20130428.

96. Stead WW, Senner JW, Reddick WT, Lofgren JP. Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med 1990;**322**:422–427.

97. Stead WW. Genetics and resistance to tuberculosis. Could resistance be enhanced by genetic engineering? Ann Intern Med 1992;**116**:937–941.

98. Li HT, Zhang TT, Zhou YQ, Huang QH, Huang J. SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. Int J Tuberc Lung Dis 2006;**10**:3–12.

99. Greenwood CM, et al. Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am J Hum Genet 2000;**67**:405–416.

100. Malik S, et al. Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease. Proc Natl Acad Sci USA 2005;**102**:12183–12188.

101. Young D, Perkins MD, Duncan K, Barry CE 3rd. Confronting the scientific obstacles to global control of Tuberculosis. J Clin Invest 2008;**118**:1255–1265.

102. Caws M, et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog 2008;**4**:e1000034.

103. Thye T, et al. Variant G57E of mannose binding lectin associated with protection against tuberculosis caused by Mycobacterium africanum but not by M. tuberculosis. PLoS ONE 2011;**6**:e20908.

104. Intemann CD, et al. Autophagy gene variant IRGM -261T contributes to protection from tuberculosis caused by Mycobacterium tuberculosis but not by M. africanum strains. PLoS Pathog 2009;**5**:e1000577.

105. Salie M, et al. Associations between human leukocyte antigen class I variants and the Mycobacterium tuberculosis subtypes causing disease. J Infect Dis 2014;**209**:216–223.

106. Bartha I, et al. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. Elife 2013;**2**:e01123.

107. Pepperell CS, et al. The role of selection in shaping diversity of natural M. tuberculosis populations. PLoS Pathog 2013;**9**:e1003543.

108. Hughes AL, Friedman R, Murray M. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of Mycobacterium tuberculosis. Emerg Infect Dis 2002;**8**:1342–1346.

109. Morelli G, et al. Microevolution of Helicobacter pylori during prolonged infection of single hosts and within families. PLoS Genet 2010;**6**:e1001036.

110. Kuo CH, Ochman H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol Direct 2009;**4**:35.

111. 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 2009;**4**:e7815.

112. Ford CB, et al. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet 2011;**43**:482–486.

113. Walker TM, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 2013;**13**:137–146.

114. Moran NA, Munson MA, Baumann P, Ishikawa H. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc R Soc Lond B Biol Sci 1993;**253**:167–171.

115. Moodley Y, et al. The peopling of the Pacific from a bacterial perspective. Science 2009;**323**:527–530.

116. Jin L, Su B. Natives or immigrants: modern human origin in east Asia. Nat Rev Genet 2000;**1**:126–133.

117. Stewart JR, Stringer CB. Human evolution out of Africa: the role of refugia and climate change. Science 2012;**335**:1317–1321.

118. Charleston MA, Robertson DL. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst Biol 2002;**51**:528–535.

119. Paterson AM, Banks J. Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. Int J Parasitol 2001;**31**:1012–1022.

120. Underhill PA, Kivisild T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 2007;**41**:539–564.

121. Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. Lancet Infect Dis 2010;**10**:103–111.

122. Pepperell CS, et al. Dispersal of Mycobacterium tuberculosis via the Canadian fur trade. Proc Natl Acad Sci USA 2011;**108**:6526–6531.

123. Kim TH, Kubica GP. Long-term preservation and storage of mycobacteria. Appl Microbiol 1972;**24**:311–317.

124. Bryant JM, et al. Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. Lancet Respir Med 2013;**1**:786–792.

125. Bryant JM, et al. Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. BMC Infect Dis 2013;**13**:110.

126. Chan JZ, et al. Metagenomic analysis of tuberculosis in a mummy. N Engl J Med 2013;**369**:289–290.

127. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. Syst Biol 2007;**56**:515–522.

128. Rocha EP, et al. Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 2006;**239**:226–235.

129. KNOWLEDGE IFTSOH. Available at: http://www.humanjourney.us.

130. Foundation S. Available at: http://www.silkroadfoundation.org/toc/index.html.

131. Donoghue HD, et al. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. Lancet Infect Dis 2004;**4**:584–592.

132. Djelouadji Z, Raoult D, Drancourt M. Palaeogenomics of Mycobacterium tuberculosis: epidemic bursts with a degrading genome. Lancet Infect Dis 2011;**11**:641–650.

133. Rothschild BM, et al. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clin Infect Dis 2001;**33**:305–311.

134. Thierry D, et al. IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res 1990;**18**:188.

135. Coros A, DeConno E, Derbyshire KM. IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. J Bacteriol 2008;**190**:3408–3410.

136. He L, Fan X, Xie J. Comparative genomic structures of Mycobacterium CRISPR-Cas. J Cell Biochem 2012;**113**:2464–2473.

137. Hershkovitz I, et al. Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. PLoS ONE 2008;**3**:e3426.

138. Donoghue HD, et al. Biomolecular archaeology of ancient tuberculosis: response to "Deficiencies and challenges in the study of ancient tuberculosis DNA" by Wilbur et al. (2009). J Archaeol Sci 2009;**36**:2797–2804.

139. Wilbur AK, et al. Deficiencies and challenges in the study of ancient tuberculosis DNA. J Archaeol Sci 2009;**36**:1990–1997.

140. Nicklisch N, et al. Rib lesions in skeletons from early neolithic sites in Central Germany: on the trail of tuberculosis at the onset of agriculture. Am J Phys Anthropol 2012;**149**:391–404.

141. Zink AR, et al. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol 2003;**41**:359–367.

142. Taylor GM, Young DB, Mays SA. Genotypic analysis of the earliest known prehistoric case of tuberculosis in Britain. J Clin Microbiol 2005;**43**:2236–2240.

143. Muller R, Roberts CA, Brown TA. Genotyping of ancient Mycobacterium tuberculosis strains reveals historic genetic diversity. Proc R Soc Lond B Biol Sci 2014;**281**:20133236.

144. Kwan CK, Ernst JD. HIV and tuberculosis: a deadly human syndemic. Clin Microbiol Rev 2011;**24**:351–376.

145. Huang CC, et al. The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. Clin Infect Dis 2014;**58**:765–774.

146. Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 2004;**38**:615–643.

147. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol 2009;**7**:493–503.

148. Brennan MJ, Delogu G. The PE multigene family: a 'molecular mantra' for mycobacteria. Trends Microbiol 2002;**10**:246–249.

149. McEvoy CR, et al. Comparative analysis of Mycobacterium tuberculosis pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. PLoS ONE 2012;**7**:e30593.

150. Copin R, et al. Sequence diversity in the pe_pgrs genes of Mycobacterium tuberculosis is independent of human T cell recognition. MBio 2014;**5**:e00960–00913.

151. Arjan JA, Visser M, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. Diminishing returns from mutation supply rate in asexual populations. Science 1999;**283**:404–406.

152. Rice WR. Experimental tests of the adaptive significance of sexual recombination. Nat Rev Genet 2002;**3**:241–251.

153. Osorio NS, et al. Evidence for diversifying selection in a set of Mycobacterium tuberculosis genes in response to antibiotic- and nonantibiotic-related pressure. Mol Biol Evol 2013;**30**:1326–1336.

154. Achtman M. Insights from genomic comparisons of genetically monomorphic bacterial pathogens. Philos Trans R Soc Lond B Biol Sci 2012;**367**:860–867.

155. Lipsitch M, Sousa AO. Historical intensity of natural selection for resistance to tuberculosis. Genetics 2002;**161**:1599–1607.

156. Barnes I, Duda A, Pybus OG, Thomas MG. Ancient urbanization predicts genetic resistance to tuberculosis. Evolution 2010;**65**:842–848.

157. Williams A, Dunbar RIM. Big brains, meat, tuberculosis and the nicotinamide switches: co-evolutionary relationships with modern repercussions on longevity and disease? Int J Tryptophan Res 2013;**15**:73–88.

158. Brites D, Gagneux S. Old and new selective pressures on Mycobacterium tuberculosis. Infect Genet Evol 2012;**12**:678–685.

159. Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB. Mapping of genotype-phenotype diversity among clinical isolates of mycobacterium tuberculosis by sequence-based transcriptional profiling. Genome Biol Evol 2013;**5**:1849–1862.

160. Lin PL, et al. Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. Nat Med 2014;**20**:75–79.

161. Kodaman N, Sobota RS, Mera R, Schneider BG, Williams SM. Disrupted human-pathogen co-evolution: a model for disease. Front Genet 2014;**5**:290.