Published in final edited form as:
Semin Immunol. 2014 December ; 26(6): 431–444. doi:10.1016/j.smim.2014.09.012.

Consequences of genomic diversity in *Mycobacterium tuberculosis*

Mireia Coscolla[1,2] and **Sebastien Gagneux**[1,2]

[1]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland [2]University of Basel, Petersplatz 1, Basel 4003, Switzerland

Abstract

The causative agent of human tuberculosis, *Mycobacterium tuberculosis* complex (MTBC), comprises seven phylogenetically distinct lineages associated with different geographical regions. Here we review the latest findings on the nature and amount of genomic diversity within and between MTBC lineages. We then review recent evidence for the effect of this genomic diversity on mycobacterial phenotypes measured experimentally and in clinical settings. We conclude that overall, the most geographically widespread Lineage 2 (includes Beijing) and Lineage 4 (also known as Euro-American) are more virulent than other lineages that are more geographically restricted. This increased virulence is associated with delayed or reduced pro-inflammatory host immune responses, greater severity of disease, and enhanced transmission. Future work should focus on the interaction between MTBC and human genetic diversity, as well as on the environmental factors that modulate these interactions.

Keywords

Mycobacterium tuberculosis; genome; diversity; SNP; lineage; virulence; transmission

1. Introduction

Tuberculosis (TB) in humans is mostly caused by the members of the *Mycobacterium tuberculosis* complex (MTBC) known as *Mycobacterium tuberculosis* sensu stricto and *Mycobacterium africanum.* MTBC are gram-positive acid-fast bacteria transmitted via aerosols generated by patients with pulmonary TB. The outcome of TB infection and disease is highly variable: exposure to MTBC can be followed by rapid clearance through innate

Corresponding author: Sebastien Gagneux, Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Socinstrasse 57, 4051 Basel, Switzerland, Phone: +41-61-284-8369; Sebastien.Gagneux@unibas.ch.

Conflict of interest statement
The authors declare no conflicts of interest.

immunity, direct development of active disease, or latent infection that may or may not re-activate up to several decades following initial exposure. Active TB disease comprises a range of presentations, including classical pulmonary TB, and various forms of extrapulmonary disease such as TB meningitis and miliary TB. Each of these different forms of TB feature a variety of symptoms that are associated with diverse host responses to the pathogen [1]. Traditionally, the different outcomes of TB infection and disease have been attributed to host and environmental variables [2]. Various human genetic determinants are known to influence the susceptibility to TB [3] (see also contributions by Meyer et al. and Bustamante et al. to this special issue). Environmental factors such as overcrowding and poor ventilation increase exposure to infectious particles [4], and implementing improved ventilation has been shown to reduce MTBC transmission [5]. Increasingly however, it is becoming clear that better knowledge of the bacterial determinants of virulence and their interaction with host and environmental factors will improve our understanding of the pathogenesis of TB [6].

Many experimental studies have provided evidence that clinical strains of MTBC differ in virulence (reviewed in [7–9]). However, unlike many other pathogenic bacteria like *Corynebacterium diphtheriae*, *Escherichia coli* O157:H7, *Shigella dysenteriae*, *Vibrio cholerae*, and *Salmonella typhi*, MTBC lacks canonical virulence factors such as, e.g., toxins. Hence in TB, determining the effects of strain-specific variation on infection and disease is challenging. Some of the limitations for discovering general patterns in genotype-phenotype associations has been the lack of phylogenetically robust classification systems for MTBC clinical strains [10]. Furthermore, identifying genetic determinants of virulence requires analytical methods that index genomic diversity in a more comprehensive manner as opposed to mere “genotyping”. During the past five years, advances in DNA sequencing technologies have made available many whole-genome sequences of MTBC clinical strains from around the world [11]. This has led to a better understanding of the global phylogenetic diversity of MTBC. Many genotyping schemes have been developed for MTBC in the past, (reviewed in [12]), but only comparative whole-genome sequencing (WGS) provides the phylogenetically robust framework for strain classification coupled with the sensitivity required to unmask the genetic particularities of different strains in detail. In addition, DNA sequence information can be further exploited using various comparative genomic and population genetic tools to predict the potential phenotypic impact of particular genetic polymorphisms [13,14].

In this review, we start by describing the latest insights from WGS data into the nature of MTBC genomic diversity, and provide a list of new databases and analysis platforms for WGS data relevant to MTBC. We then review recent attempts to predict the impact of MTBC genomic diversity on gene function and host immune recognition. We continue by summarizing the most recent experimental and epidemiological evidence supporting the relevance of MTBC genomic diversity for different virulence phenotypes, and end with an outlook on future research directions.

2. The nature of MTBC genomic diversity

2.1. Genotyping methods to classify MTBC clinical strains

During the last decades, various molecular techniques of DNA fingerprinting have been used to discriminate between clinical strains of MTBC (reviewed in [12]). Restriction fragment length polymorphisms (RFLP) typing is based on differences in copy number and the differential genomic location of the insertion sequence (IS) *6110*, and became the first gold standard method for genotyping MTBC [15]. This technique has been used successfully to define chains of ongoing TB transmission, discriminate relapse from re-infection, and to detect laboratory cross-contaminations [16,17]. IS*6110* is an IS element of 1,361 bp flanked by 28 bp inverted repeats which is differently inserted in the genome across strains. IS*6110*-RFLP analysis relies on digestion of genomic DNA with endonucleases followed by electrophoretic separation, blotting onto nylon filter, and probing of the restriction fragments with a IS*6110*-specific DNA fragment. Some of the limitations of IS*6110*-RFLP are that this technique is difficult to reproduce between laboratories, and that it requires large amounts of good quality DNA.

Because of these limitations, several PCR-based methods have been developed for MTBC strain typing that require little DNA. These include spoligotyping and Mycobacterial Interspersed Repeat Units (MIRUs) typing, which together have been recently defined as the new gold standard for molecular epidemiological investigation of TB [12]. Spoligotyping patterns are defined based on the presence or absence of 43 unique regions intercalated between direct repeats in the Clustered Regularly Interspaced Short Palindromic Repeats region (CRISPRs) of the MTBC genome. MIRUs profiles classify MTBC strains by the number of repeats at different Variable Number of Tandem Repeats (VNTRs) loci. Spoligotypes and MIRU-types can be compared using the SITVITWEB database that includes thousands of spoligotyping patterns [18] and MIRU-VNTRplus [19]. However, the use of spoligotyping and MIRUs is limited for phylogenetics and strain classification because of the propensity of the corresponding molecular markers for convergent evolution; i.e. because these markers change rapidly, the same or similar patterns can emerge by chance in strains that are phylogenetically unrelated [10].

To get around this problem, genomic deletions, often referred to as Regions of Difference (RDs) or Large Sequence Polymorphism (LSPs) have been used as markers to classify groups of MTBC strains into main phylogenetic lineages [20–24], and sub-lineages [25,26]. Although horizontal gene transfer between MTBC and *M. canettii* has been detected [27], the population structure within MTBC is largely clonal [2,20]. Because on-going horizontal gene exchange is rare in MTBC, LSPs are essentially irreversible, making them ideal phylogenetic markers for strain classification.

Following the completion of the first MTBC genomes [28–30], comparative genomics identified sets of phylogenetically informative single nucleotide polymorphisms (SNPs) that were used to establish various strain-typing methodologies [29,31–33]. In addition, several groups have developed SNP-typing schemes using multilocus sequencing analyses [34–36]. More recently, accumulating WGS data have led to the development of novel SNP-typing methods that rely on a broader understanding of the global MTBC phylogenetic diversity

[11]. Some of these methods have been built into highly multiplexed assays [37–40]. In addition to SNPs, other forms of genetic diversity have been incorporated in high-throughput genotyping schemes, including spoligotyping data [41,42], and drug resistance-conferring mutations [43].

2.2. New software and databases to explore genomic MTBC diversity from WGS data

Despite the usefulness of the various genotyping assays discussed above [11], WGS remains the only tool that can classify MTBC strains robustly and simultaneously index genomic diversity at all levels, be it at the level of whole populations [44,45], within large outbreaks [46,47], during household transmission [48], within single patients [49,50], or during in vitro evolution [51–54]. Moreover, DNA sequence data can be used to measure phylogenetic distances, and hence quantify the amount of genetic diversity within and between groups of strains. Finally, WGS allows discovering new mutations in particular strains, or groups of strains, associated with particular phenotypes. While WGS is becoming cheaper and more widely available, the analysis of WGS data remains often limiting. Hence, rapid and user-friendly analysis methods are required. Recently, several new databases have been developed that make WGS data readily available. These platforms also include tools to visualize and analyse MTBC genomic diversity (Table 1). They contain multiple types of information on multiple MTBC genomes, including DNA and protein sequences, maps, assemblies, annotations, and bibliography, as well as gene expression and protein data associated with different genomic regions. Whilst the Tuberculist database [55] offers diverse information on the reference *M. tuberculosis* genome of the laboratory strain H37Rv, these new databases incorporate tools for comparative genomics, and include data from various MTBC clinical strains as well as other species within the genus *Mycobacterium* (Table 1). Most of these databases build on already available genomes. In addition, new tools have been developed recently that help explore unknown genomic diversity based on newly generated WGS data. For example, SpolPred predicts the spoligo-type from short DNA sequencing reads [56]. Similarly, KvarQ provides a robust SNP typing schemes using short DNA sequencing reads from WGS, and offers a robust classification in the main human- and animal-adapted lineages of MTBC. In addition, KvarQ identifies drug resistance-conferring mutations [57]. In summary, recent advances in WGS of MTBC clinical strains are revealing a larger amount of between-strain genomic diversity than generally appreciated (see below). Increasingly, it is also becoming clear that this genomic diversity translates into relevant phenotypic variation.

2.3. Differences between MTBC strains and lineages

The MTBC comprises various closely related bacterial species and sub-species, including *M. tuberculosis* sensu stricto and *M. africanum* which are adapted to humans, as well as several animal-adapted forms [58], i.e. *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii*, *M. origys*, *M. mungi*, *M. suricattae*, the dassie bacillus, and the chimpanzee bacillus [59–62]. In addition to these classical members of MTBC, the Complex also comprises more distantly related bacteria known as *M. canettii* and other so-called “smooth tuberculosis bacilli (STBs)”, which are characterized by a smooth colony morphology (STB) [63–65]. The STBs exhibit several other important features, including strong evidence of horizontal gene transfer, which sets them apart from the other members of MTBC [66]. Only approximately

60 isolates of STB have been described so far, most of which from the Horn of Africa. Increasing epidemiological evidence suggests STBs are environmental organisms that only occasionally infect humans [67].

One of the first evolutionary reconstructions of the genetic population structure of the whole MTBC highlighted a group of strains harbouring a deletion in the genomic region known as TbD1 [68]. TbD1-deleted strains have been referred to as evolutionarily “modern” compared to the strains without this deletion, which collectively have been referred to as evolutionarily “ancestral” or “ancient”. Today, based on WGS analyses, we know that MTBC comprises seven human-adapted lineages (Lineage 1 to Lineage 7 in Figure 1A), where the “modern” clade form a monophyletic group comprising the TbD1-deleted Lineages 2, 3 and 4, as these lineages diversified more recently than the remaining MTBC strains. By contrast, the “ancestral” strains are paraphyletic, meaning they do not comprise a single phylogenetic group. Many studies have shown that the human-adapted MTBC lineages show a strong phylogeographical population structure, with the different lineages associated with distinct geographical regions [20,24,33–35,69,70]. Some of these lineages are also more globally widespread than others (Figure 1). Specifically, the most widely distributed groups are Lineage 2 and Lineage 4 (Figure 1B). Lineage 2 (also known as East-Asian lineage, includes the Beijing family of strains) predominates in East Asia, but is also present in Central Asia, Russia and South-Africa. Lineage 4 (also known as the Euro-American lineage) occurs frequently in populations from Asia, Europe, Africa and America. Lineage 1 and Lineage 3 show a more restricted geographical distribution limited to East Africa, Central-, South- and South-East Asia (Figure 1C). The most geographically restricted lineages are Lineage 5, Lineage 6 and Lineage 7, which are all associated with specific regions of Africa (Figure 1D). Lineage 5 and Lineage 6 are also known as *M. africanum* West Africa 1 and West Africa 2, respectively, and almost exclusively occur in West Africa or in recent immigrants from those regions [71]. Lineage 6 occurs primarily in the Western part of West Africa, whereas Lineage 5 dominates further to the East in regions bordering the Gulf of Guinea [71,72]. Similarly, the recently discovered Lineage 7 is confined to Ethiopia and recent immigrants from that part of the world. The reasons for why these three lineages are limited to specific regions of Africa are unknown [73,74]. Finally, two other lineages within the classical MTBC are adapted to different wild or domestic animal species (Figure 1A). One of these two lineages comprises the classical animal-associated strains *M. bovis* (includes the vaccine strains BCG), *M. caprae*, *M. microti*, *M. pinnipedii* and *M. orygis*. The other animal-adapted lineage includes the chimpanzee bacillus, and, although genome data are not yet available, likely also the dassie bacillus, *M. mungi* and *M. suricatae* [59–62]. Recently, a new group of MTBC has been identified in 1000-year old human remains from Peru [75]. These ancient MTBC strains were distinct from any known human-adapted MTBC, but most closely related to contemporary *M. pinnipedii* which is adapted to seals and sea lions. These findings suggest that marine mammals could have played a role in spreading TB from Africa across the Atlantic Ocean to the New World and transmitting to pre-Columbian human populations [75], Figure 1A). For the remainder of this review, we will focus on the classical human-adapted members of MTBC, i.e. the Lineages 1 to 7 depicted in colour in Figure 1.

Strains of MTBC differ in their content of SNPs, small insertion and deletions (indels), large genomic deletions, large duplications and insertion sequences. Unlike MIRU and spoligotyping patterns, LSP are robust makers for phylogenetic classification, but LSPs are not polymorphic enough to differentiate among closely related strains, e.g. within an outbreak or transmission chain. By contrast, WGS reveals all types of mutations, and provides the best discriminatory power to differentiate between strains. Moreover, WGS allows computing phylogenetic distances and quantifying genomic diversity within and between groups of strains. To obtain an initial estimate of within MTBC diversity, we have used previously published whole genome sequences of 217 globally distributed clinical strains [76] and calculated the number of SNPs between any pair of strains. On average, two human-adapted MTBC strains differed by about 1,200 SNPs, which corresponds to 0.03% of the genome when excluding repetitive sequences (Figure 2). By contrast, the corresponding difference between any classical MTBC strain and *M. canettii* was about 2.7 % [65], which is 90 times larger than the average SNP-distance among human-adapted members of MTBC. Next, we calculated the average SNP distances within and between human-adapted MTBC lineages (Figure 2). The geographical origin of strains included in this analysis was diverse, and except for Lineage 7 that is only found in Ethiopia and Ethiopian immigrants, every lineage included strains from various countries. We found that Lineage 1 harboured the largest genetic diversity with an average of 730 SNPs between any two strains belonging to this lineage. The corresponding average distance was lowest for Lineage 7 with only 230 SNPs. In terms of between lineage diversity, the strains belonging to the “modern” Lineages 2, 3 and 4 differed by 970 SNPs in average. Strains belonging to the “ancestral” Lineages 1, 5, and 6 were more distantly related with an average of 1,500 SNPs between them. The maximum SNP distance of 1,800 SNPs was observed between strains of Lineage 7 and either Lineage 1, 5, or 6 (Figure 2). Although these estimates might change when more genomes are considered, they provide a first indication of the relative genomic distances within and between the different human-adapted lineages of MTBC.

3. The consequences of genomic diversity in MTBC

3.1. Predictions of the impact of MTBC genomic diversity

In addition to exploring and quantifying the genomic diversity in MTBC, WGS data has been used together with computational methods to predict the impact of this variation on the bacterial phenotype. One of the striking observations has been that contrary to many other organisms, in MTBC about two thirds of SNPs in coding regions are non-synonymous (i.e. amino acid changing) [29,35,76,77]. Moreover, using sequence data from 89 genes in 107 MTBC strains, Hershberg et al. [35] observed that 58% of the non-synonymous mutations fell in positions that were highly conserved in other mycobacteria, suggesting that most of these mutations in MTBC might have functional consequences. More recently, Rose et al. used WGS data to identify all SNPs specific to the different human-adapted lineages of MTBC; these are SNPs that are fixed in the corresponding bacterial populations [78] (Figure 3A). The number of fixed SNPs (i.e. synonymous and non-synonymous) ranged from 124 in Lineage 2 to 698 in Lineage 5, but in all cases, at least 44% of the non-synonymous SNPs fixed in one or the other lineage were predicted to impact gene function (Figure 3B). Taken together, these findings illustrate that even though the overall genomic diversity of MTBC is

low compared to other bacteria [79], a large proportion of the mutations that have accumulated in the different phylogenetic lineages of MTBC are likely to lead to phenotypic differences.

An alternative way to predict the functional consequences of mutations is to determine the evolutionary conservation of the corresponding genomic regions and the selection pressures in response to which these mutations evolve. The importance of a protein residue is reflected in its evolutionary conservation [80]. For example, loci coding for important general functions will be under purifying selection and thus highly conserved. Alternatively, functional innovations might be under positive selection, and hence detecting positively selected sites in genomes can point to adaptive processes. One of the strongest on-going selection pressures in MTBC has been drug pressure, and detection of positive selection has been used successfully to identify molecular markers of drug resistance [81–83]. One particular signature of positive selection is the independent multiple appearance of the same mutation; a phenomenon referred to as convergent evolution. Convergent evolution leading to so-called homoplasies are generally uncommon in MTBC, except for mutations involved in drug resistance [52]. Recently, signatures of positive selection have also been detected in membrane proteins of MTBC, suggesting a role of the corresponding mutations in the interaction with the host immune system [83].

WGS data can also be used to explore the putative impact of host recognition in MTBC. Many pathogens evade host immunity by means of antigenic variation [84]. Intriguingly, a comparative WGS analysis of 21 MTBC clinical strains from global sources revealed that the majority of 495 experimentally confirmed human T-cell epitopes in these strains were highly conserved [77]. The conservation of T-cell epitopes in human-adapted MTBC might reflect the fact that the host immune response is largely responsible for the lung damage during TB disease (i.e. cavitation), which increases patient coughing and hence contributes the successful transmission of the pathogen. Indeed, TB patients with cavitary disease are more likely to generate secondary cases [85]. These findings have implications for the development of new TB vaccines and diagnostics, as highly conserved antigens might represent ideal diagnostic targets, but not adequate vaccine components if the immune responses they elicit benefit the bacteria rather than the host.

3.2. The phenotypic consequences of genomic variation

Four years ago, we reviewed the available experimental and epidemiological evidence for strain-specific phenotypic diversity [7]. Since then, additional studies have come out providing additional support for the view that strain variation in MTBC has important phenotypic consequences. Different MTBC strains and lineages differ in their growth rates in liquid culture [86,87], in monocyte derived macrophages [87], and in mice [88,89]. MTBC strains and lineages also differ in their gene expression [78,88] and metabolic profiles [90]. A recent clustering analysis derived from the differential use of various substrates separated strains according to three major lineages [90]. Most recently, Portevin et al. reported significant differences in mycolic acid profiles between different MTBC strains and lineages [91]. Mycolic acids constitute the most abundant cell wall lipid in

MTBC, and play an important role in the host immune response [92]. They also help the bacteria resist oxidative stress [93].

One of the challenges in studying the effect of MTBC strain variation has been linking genotype to phenotype. As mentioned above, a wide range of mutations occur across MTBC genomes including SNPs, deletion, duplications and mobile elements, all of which can have an impact on bacterial phenotypes. In the next few paragraphs, we review recent studies that have managed to link a given mutation in MTBC clinical strains to a particular phenotype.

Single nucleotide polymorphisms—SNPs are natural polymorphism that can be classified as synonymous, non-synonymous, nonsense, or intergenic. Because non-synonymous SNPs change the amino acid content of the corresponding protein, they are the principal contributors to functional mutations. A special category of SNPs which are better referred to as (*de novo*) single nucleotide mutations occur in genes associated with resistance to anti-TB drugs. These mutations decrease the susceptibility to a particular drug by either modifying the drug-target, increasing the expression of the gene product targeted by the drug, or by reducing the drug activation in the case of prodrugs [94,95]. Other non-synonymous SNPs have been shown to impact on the MTBC phenotype irrespective of drug resistance, with the PhoPR two-component system providing a particularly illustrative example. The PhoPR components are essential for MTBC virulence in animal models of TB (reviewed in [96]), and several point mutations in either component of this system have been shown to lead to important changes in bacterial phenotype. For example, one amino acid change at position 219 of PhoP in the laboratory strain H37Ra alters the binding capacity of PhoP to its own promotor [97]. As a result, H37Ra is highly attenuated compared to its virulent counterpart H37Rv. Similarly, one amino acid change at position 71 of PhoR occurs naturally in all MTBC strains belonging to Lineage 5 and 6, and in all animal-adapted MTBC strains. This mutations was shown to affect the PhoPR regulatory system, and consequently, the synthesis of important surface lipids and the secretion of the 6-kDa antigenic target ESAT-6 were reduced in the corresponding strains [98]. Non-synonymous SNPs can also lead to epigenetic changes. Recently, Shell et al. described one non-synonymous SNP in the methiltransferase *mamA* in Lineage 2 and another non-synonymous SNP in the active site of *hsdM* in Lineage 4 that lead to variation in adenine methylation levels in strains belonging to these lineages [99].

Compared to non-synonymous SNPs, intergenic and synonymous SNPs are often referred to as evolutionary neutral because on average, their effect on phenotype, and hence fitness, is less. However, some synonymous SNPs do have important phenotypic effects. For example, a synonymous SNP within the gene immediately upstream of DosR has been shown to generate an alternative internal transcriptional start site (TSS) in Lineage 2/Beijing strains. This TSS has been associated with increased expression of the *dosR* regulon [78]. Similarly, a synonymous SNP upstream of *malQ* in Lineage 1 strains has been found to create a new TSS associated with increased expression of *mal*Q [78]. Finally, a recent study reported a synonymous SNP leading to a new internal TSS in *mabA* which was demonstrated to confer isoniazid resistance in MTBC clinical strains, representing a novel mechanism of resistance to this important first line anti-TB drug [100]. Additionally, SNPs in non-coding regions can have important functional consequences. Non-coding SNPs in promoter regions can modify

the promoters and therefore alter transcriptional regulation. For example, several intergenic SNPs in the promoter of the *inh*A gene, increase the transcription level of *inhA*, thereby conferring resistance to isoniazid and ethionomide [101,102].

Finally, nonsense mutations result in a premature stop codon that produce truncated peptides with a likely impact on the functionality of the product encoded; this process is also referred to as pseudogenization. Unlike *M. leprae* where almost half of the coding DNA sequences are pseudo-genes [103], pseudogenization in MTBC seem to be a more recent process and not as significant as in *M. leprae* [104]. However, some nonsense mutations are found within MTBC, where Lineage 6 and the *M. bovis* clades contain more pseudo-genes than other lineages [104]. Many of these nonsense mutations are in genes that have analogous genes or pathways in the MTBC genome and presumably, these mutations will have a minor phenotypic impact due to genomic redundancy [104].

Gene Duplications—Gene duplications and the expansion of gene families are important sources of genetic diversity with the potential to lead to evolutionary innovations [105]. Duplications of the ESAT-6 gene clusters in an ancestor of MTBC resulted in the expansion of the PE/PPE gene family [106]. PE/PPE genes have been postulated to play a role in antigenic variation [107–109]. Indeed, these genes are highly polymorphic [110–114], but the evolutionary forces driving this diversity remain unknown. Nevertheless, differences in nucleotide variation and gene expression in some PE/PPE genes have been associated with virulence [115,116].

Although there is currently no evidence of large duplications being a major source of genomic diversity within MTBC, several instances have been reported. A 350kb genomic duplication that includes the DosR operon has been observed in some Lineage 2 strains. This duplication has been suggested to be partially responsible for increased expression of the *dosR* regulon in Lineage 2/Beijing strains discussed above [117]. However, this duplication is not seen in all Lineage 2/Beijing strains [78], possibly because the duplicated regions is lost during *in vitro* cultivation [117]. Interestingly, in addition to some Lineage 2 strains, some strains belonging to Lineage 4 have been shown to carry independent duplications spanning the same genomic region [118]. As outlined before, the presence of convergent evolution events in MTBC is rare [52]. Therefore, the convergent duplication of *dosR* regulon might indicate some evolutionary advantage of strains harbouring this duplication.

Large scale duplications might also have played a role during the *in vitro* evolution of BCG strains. Two tandem duplications termed DU1 and DU2, of 29,668 bp and 36,161 bp, respectively, are present in *M.bovis* BCG Pasteur compared to H37Rv [119]. Intriguingly, DU2 showed four alternative forms evolved in the different laboratories where the different BCG strains were passaged, leading to independent duplications of similar genomic regions [120]. If and how these duplications impact BCG phenotypes, including vaccine efficacy, remains unknown [54].

Repetitive and mobile genetic elements—Repetitive elements such as CRISPs, VNTRs and ISs have been used as molecular markers for MTBC strain genotyping

(discussed above and reviewed in [12]). These elements are an important source of genomic variation that could impact bacterial phenotypes.

The CRISPR region in MTBC is known to be polymorphic and shows convergent deletions in phylogenetically unrelated strains [121]. However, the functional role (if any) of polymorphisms affecting to the CRISPR region in MTBC, or the impact of CRISPR polymorphisms on mycobacterial phenotypes have not been determined.

In humans, variation in some VNTR loci is associated with complex diseases such as type-1 diabetes [122]. In bacteria, VNTRs have been implicated in the ON/OFF switching of phase variable genes in *Neisseria* [123,124]. In mycobacteria, intragenic VNTR loci variation has been shown to modify the structure and function of the proteins affected [125]. Intergenic VNTR variation can alter promoter activity in downstream genes [126–128]. For example, an increased number of repeats in the VNTR3239 locus of MTBC leads to a higher expression of the downstream gene *fpg1,* a DNA glycosylase shown to be involved in the protection against oxidative DNA damage [126].

Finally, differential insertion of the IS*6110* sequences in MTBC clinical strains has been shown to lead to important phenotypic effects linked to changes in gene expression. When inserted in the upstream region of a gene, the insertion sequence can modify the transcription of neighbouring genes by acting as a promotor [129–131]. In addition, insertion of IS*6110* can disrupt genes. Indeed, the presence of independent disruptions of the same gene in different clinical strains, i.e. another example of convergent evolution, might be associated with an evolutionary advantage for the corresponding strains [132]. For example, one hyper-virulent *M. bovis* strains that caused an outbreak of human TB in Spain showed an IS*6110* insertion upstream of the *phoP* regulon [98]. The expression of the *phoP* regulon is generally diminished in animal strains and Lineage 6 strains due to the non-synonymous SNP in *phoP* mentioned above. The insertion of IS*6110* upstream of *phoP* regulon in this hyper-virulent *M. bovis* restores *phoP* transcription, compensating for the effect of the non-synonymous SNP [98]. *M. bovis* is generally not an efficient pathogen in humans, among which it rarely transmits. However, this single IS*6110* insertion seemed to have transformed an MTBC strain adapted to cattle into a variant capable of sustaining a transmission cycle in a new host (i.e. humans).

Genomic deletions—Large genomic deletions are a substantial source of genomic diversity within MTBC. Genomic deletions can result from transposition of mobile genetic elements, like in the case of the prophages phiRv1 (linked to RD3) and phiRv2 (linked to RD11), and insertion sequences IS*1532* (e.g. in the case of RD6) or IS*6110* (e.g. in the case of RD5) [68,133]. Deletions can also be mediated by homologous recombination between adjacent IS*6110* resulting in the loss of the intervening DNA segment; examples for this mechanism include RvD2, RvD3, RvD4, and RvD5 [68]. In addition, many genomic deletions occur through unknown mechanisms [23].

Deletion of one or several loci can greatly influence the bacterial phenotype. One of the most relevant MTBC phenotype is drug resistance, and deletions have been implicated in drug resistance phenotypes. The presence of repetitive sequences near the *katG* region of the

M. tuberculosis genome makes this gene prone to deletions [134], and deletions of *katG* have been shown to confer high level resistance to isoniazid [135].

Other deletions have been reported to lead to the modification of the lipid composition of the mycobacterial wall, which can therefore alter the bacterial interaction with host cells. A deletion disrupting the locus of the polyketide synthase (Pks)15/1 in strains belonging to Lineage 4 results in a defective production of phenolic glycolipid (PGL) [136], thereby affecting the bacterial interaction with host immune cells [137]. Indeed, the absence of PGL has been linked to reduced MTBC virulence in infection models [138,139]. Similarly, a deletion specific to all Lineage 3 strains involves Rv1519, and has been linked to a decrease in the production of the anti-inflamatory cytokine IL-10 by in infected hosts [140]. Genomic deletions have also been associated with clinical phenotypes. For example, patients with extra-pulmonary TB were more likely to carry MTBC strains with a deletion in the phospholipase C-encoding gene *plcD* [141]. Moreover, a deletion of the *embR* locus has been detected in a particularly successful multidrug-resistant strain of MTBC that has successfully spread between continents (Coscolla et al., manuscript submitted). EmbR is a transcriptional regulator implicated in ethambutol resistance, as well as in the regulation of lipomannan/lipoarabinomannan ratio [142], which serves as immunomodulator crucial for mycobacterial virulence (reviewed in [143]). Finally, deletions in PE_PGRS33 have been associated with reduced induction of tumour necrosis factorα (TNF-α) by the host [144], reduced patient clustering (indicating reduced transmission), and absence of lung cavitation, supporting the view that PE_PGRS33 plays a role in the transmission success of MTBC in clinical settings [145].

3.3. New evidence supporting differences in virulence and immunogenicity between MTBC clinical strains

Following the 100 studies we reviewed four years ago [7], many new studies have explored the effect of MTBC strain variation on virulence and immunogenicity. Human-adapted MTBC is an obligate pathogen in the sense that i) it has no other animal or environmental reservoir, and ii) it has to cause (pulmonary) disease to transmit successfully. In other words, virulence is directly linked to transmission [85,146], which is unlike many other pathogens where transmission occurs independently of disease. “Virulence” in TB can be conceptualized as a composite comprising i) the ability of the bacteria to survive in face of the host immune responses, ii) their capacity to cause lung damage, iii) to survive the aerosolisation process outside of the host, and iv) successfully transmit to and infect a new host. In the following sections, we review recent studies that have generated evidence supporting strain differences with regards to these individual components of “virulence”.

Impact on host immune regulation—Studies in humans and animal models have shown that TNF α, interferon-g (IFN-γ), IL-12 and IL-17 are important mediators of a protective immune response against TB (reviewed in [1,147]). Other mediators such as IL-10 may play a role in limiting MTBC clearance during the early immune response [148]. Considering the link between virulence and transmissibility in MTBC, one could postulate that a more severe inflammatory response causing more lung damage would lead to more efficient transmission (further discussed in [147]). Intriguingly however, many studies have

found increased MTBC virulence associated with reduced and/or delayed inflammatory responses, perhaps by allowing a stronger bacterial proliferation early during the infection process, leading to increased virulence at a later stage. For example, studies in different infection models have found strain NH878 (belongs to Lineage 2/Beijing) constantly associated with a delayed inflammatory immune response and increased virulence [149–156].

At the lineage level, the so-called “modern” Lineages 2, 3 and 4, showed a lower early inflammatory response compared to Lineage 1 and Lineage 6 [157]. As discussed under Section 2.3. above, “modern” MTBC strains are more globally widespread than other lineages. Hence, the observation that “modern” strains are associated with a delayed inflammatory response (i.e. higher virulence) might be linked to the global success of these strains [6]. Indeed, a study by Reiling et al. [158] showed that “modern” strains were replicating faster *in vitro*, in human monocyte-derived macrophages, as well as in aerosol-infected mice. Several other studies have demonstrated the reduced/delayed pro-inflammatory response of the “modern” versus other lineages [159]. By contrast, Krishnan et al. [89] showed increased inflammatory responses in Lineage 1 strains compared to Lineage 4, but these were not significantly higher than in Lineage 2 strains.

Other studies have compared the different lineages within “modern” lineages. One study found that Lineage 3 exhibited a higher anti-inflammatory phenotype compared to Lineage 4 [140,160,161]. Wang et al. [162] reported that different Lineage 2/Beijing strains commonly induced lower levels of TNF α, IL-6, IL-10 and GRO-α compared to strain H37Rv in monocyte-derived macrophages and dendritic cells. Similarly, Sarkar et al. [87] showed that Lineage 2 exhibited a lower pro-inflammatory phenotype compared to Lineage 3 and Lineage 4. Some studies however, showed the contrary, Lineage 4 eliciting more TNF α than Lineage 2; this was particularly true when compared to a sub-lineage of Lineage 2 known as “modern Beijing” [163]. Moreover, Krishnan et al. [89] did not detect any differences between Lineage 2 and Lineage 4. These contradictory findings with respect to the host immune responses elicited by Lineage 2 and Lineage 4 could on the one hand be due to differences in experimental conditions. On the other hand, sub-lineages within the main lineages, as well as individual strains within any given lineage might vary in the inflammatory responses they elicit.

Because the main MTBC lineages are not genetically homogeneous (Figure 2), differences in immune regulation among strains are expected at sub-lineage level [157]. Such differences have been most widely studied in Lineage 2, where “modern Beijing” showed a lower inflammatory response compared to so-called “ancient Beijing” [159,163]. Yet, Wang et al. [162] did not find any differences between Beijing strains. Some of the variable cytokine profiles elicited by the distinct groups of strains within Lineage 2 have been linked to differential toll-like receptor recognition [164]. Only one study has investigated intra-lineage variation within Lineage 4 but found no difference between the so-called H and T families of strains [160].

Increased “virulence” surely entails more than a mere delay in early pro-inflammatory response. Indeed, Reiling et al. [158] reported lineage-specific differences in virulence

profiles based on variable bacterial uptake by host cells, differences in cytokine induction, and intracellular growth. In agreement with other studies, the “modern” Lineages 2 and 4 showed high replicative potential compared to “ancestral” Lineages 1 and 6. However, Lineage 2 was characterized by low uptake, and low cytokine induction, whereas Lineage 1 and Lineage 4 exhibited high uptake and higher cytokine induction.

To date, few studies have combined host and bacterial factors when studying the different immune response during MTBC infections in humans. Coussens et al. [165] showed that ethnicity plays an important role in the different inflammatory profiles in Africans versus Eurasians from a population in London, while MTBC lineages did not contributed significantly to the different immune responses in this patient population.

Impact on Disease Severity—An alternative way of measuring “virulence” is to look at the severity of the disease. Stavrum et al. [166] showed that TB patients from Tanzania infected with “modern” Lineage 4 strains showed more α1-acid glycoprotein and C reactive protein, higher neutrophils counts, and a lower body mass index than those infected with Lineage 1. De Jong et al. [167] showed that in the Gambia, individuals infected with modern Lineages 2 and 4 were more likely to progress to active disease compared to individuals infected with Lineage 6. The higher virulence of the modern Lineage 2 compared to “ancient” Lineage 6 was corroborated in a novel marmoset model of infection [168]. Infection with Lineage 2 induced more rapid weight loss, and led to a higher bacterial load in liver, spleen and lymph nodes [168]. Finally, several studies have found Lineage 2 associated with relapse [169–171], treatment failure [172], and fever early during treatment [173].

Within the “modern” lineages, Lineage 2 is generally associated with higher “virulence” than Lineage 4. For example, Orgarkov et al. [174] found that patients carrying a polymorphism previously associated with TB (CD209_336 A/G) who were infected with Lineage 2 were more likely to die of TB compared to patients infected with other strains. Nahid et al. [175] found during a clinical trial that patients infected with Lineage 2 strains and one sub-group of Lineage 4 were more likely to yield a positive culture at week eight after treatment initiation when compared to other Lineage 4 strains.

Similar to the observed differences in immune responses caused by different sub-lineages described above, within-lineage differences in disease presentation have also been reported. Kato-Maeda et al. [176] reported that a group of highly transmissible Lineage 2 strains from San Francisco were more virulent than other Lineage 2 strains in guinea pigs. Likewise, Aguilar et al. [177] demonstrated that highly transmissible Lineage 2/Beijing sub-lineages from South Africa were more virulent in mice than less transmissible strains. Taken together, the observation that highly transmissible strains tend to cause more severe disease in animal models is consistent with intrinsic bacterial features linking “virulence” with pathogenicity, transmission, and overall strain fitness.

Impact on disease presentation—Besides the classical pulmonary presentation, MTBC can cause a wide variety of extra-pulmonary manifestations affecting many organs of the human body [178]. Pulmonary and extra-pulmonary TB can also occur

simultaneously. Because efficient MTBC transmission relies on lung damage, bacterial genotypes that are more prone to cause pulmonary and cavitary disease will be able to transmit more efficiently. Experiments in the marmoset model have revealed more extra-pulmonary spread to the lymph node, liver and spleen in animals infected with a Lineage 6 clinical strain compared to animals infected with the Lineage 4 strain CDC1551 [168]. Importantly, these experiments were conducted using genetically identical individuals (i.e. marmosets always give birth to twins or triplets), suggesting that the different disease presentation associated with the different MTBC strains are due to strain rather host genetic differences.

In humans, epidemiological studies have been rather inconsistent with regards to MTBC strain or lineage effects on disease presentations. At least five studies have reported such effects. Kong et al. reported an associating between Lineage 2 and extra-pulmonary TB [179,180]. Another study found a similar association when comparing Lineage 2 to Lineages 3 and 4 [181]. By contrast, Click et al. found Lineages 1 and 3 associated with extra-pulmonary TB when compared to Lineage 2 [182]. A study in Vietnam found Lineage 4 associated with pulmonary rather than meningeal TB [183]. On the other hand, at least five studies failed to find any association between MTBC lineage and disease presentation [74,173,184–186].

Impact on transmission—Rates of MTBC transmission are generally inferred by comparing genotypic clustering between patient isolates from a given epidemiological setting [187]. Another proxy for inferring successful transmission of particular MTBC genotypes is measuring increases in the frequency of these genotypes over time in patient populations. Finally, genotypes associated with younger patient age have also been interpreted as reflecting successful transmission, because TB in young patients are more likely to reflect ongoing transmission as opposed to reactivation [188].

Based on these concepts, several epidemiological studies have supported the view that overall, strains from “modern” lineages are more transmissible than other MTBC strains. Buu et al. [189] reported higher genotypic clustering of Lineage 2 compared Lineage 1 in Vietnam. Similarly, a study in Shanghai found Lineage 2 associated with higher clustering and younger age compared to other strains [190]. Many other studies in various settings have reported a higher fitness of Lineage 2/Beijing strains reflected by increases in their frequency over time [69,159,191–196]. In some cases, the increase of Lineage 2 was associated with drug resistance [189,197]. Yet, several studies observed no such increase and no enhanced transmissibility of Lineage 2 strains [167,198–201]. Some of these contradictory findings could be due to differences in the study populations. For example Marais et al. [200] did not find any association between Lineage 2 and higher transmission among children of South Africa. Alternatively, if sub-lineages differ in transmissibility, the genetic heterogeneity within lineages would lead to inconsistent results among different study settings where sub-lineages differ in prevalence. In support of this notion, Kato-Maeda et al. showed that one sub-lineage within Lineage 2 showed higher genotypic clustering in San Francisco [202]. This sub-lineage was also more virulent in guinea pigs, supporting again an intrinsic bacterial role in increased pathogenicity linked to enhanced transmission [176]. Sub-lineages of other lineages also tend to differ in their transmissibility as observed

in San Francisco, USA for Lineage 4 [203] and Cotonu, Benin for Lineage 5 [72]. However, three studies have reported decreasing prevalence of *M. africanum* (i.e. Lineage 5 and 6) in Cameroon, Guinea-Bissau and Burkina Faso [204–206], supporting of the lower "virulence" of the "ancestral" lineages. Therefore, although sub-lineage differences might obscure general patterns, taken together, these studies support the view that on average "modern" MTBC lineages are more transmissible than other strains. Yet, Albanna et al. [199] observed reduced transmissibility of "modern" Lineage 3 compared to other lineages in Montreal, albeit with no difference between Lineage 1, Lineage 2 and Lineage 4.

4. Conclusions

MTBC has for many years been regarded as a "clone", harbouring too little genetic diversity to be worth considering. This paradigm has now changed, and many experimental and epidemiological studies have demonstrated the phenotypic impact of MTBC strain diversity. During the last four years alone, at least 52 new studies have explored phenotypic differences among strains and lineages of MTBC, in addition to the 100 studies we reviewed earlier [7]. Although clinical studies have reported discordant results regarding the propensity of MTBC lineages to cause pulmonary as opposed to extra-pulmonary disease, studies focusing on other proxies of "virulence" have detected common patterns showing that "modern" lineages are generally more virulent and more globally successful, compared to other more geographically restricted lineages (Figure 1). Differences in immunogenicity, severity of disease, and transmission consistently indicate that Lineage 2 and Lineage 4 are more virulent than Lineages 1 and 6. More work is needed with respect to Lineage 3, 5 and 7, but given the restricted geographical distribution of these latter lineages (Figure 1), chances are that at least Lineage 5 and Lineage 7 will prove to be low virulence variants. Consistent with previous reports [8,207], many studies have noticed an emergence of Lineage 2/Beijing over time. This increase of Lineage 2/Beijing might be partially driven by environmental factors such as antibiotic treatment (reviewed in [208]) or BCG vaccination [209–211]. In addition, this effect could also be driven by changes in host demography. Specifically, the human population expansions during the Neolithic- and Industrial Revolutions have been hypothetically linked to an increase in virulence of some MTBC lineages [6,75,76].

5. Outlook

In this review, we primarily focused on the genetic diversity of MTBC, i.e. the pathogen. As discussed in the contributions by Meyer et al. and Bustamante et al. to this special issue, human genetic variation plays an important role in TB. Considering the long-term association between MTBC and its human host [212], some degree of co-evolution is likely to have occurred [20,24,33–35,69,70,213,214]. Indeed, several recent studies have shown that the susceptibility to TB is influenced by variation in both the pathogen and the host [183,215–217]. Hence, the interactions between host and pathogen genomic diversity needs to be explored and better understood using novel approaches [218,219]. Moreover, the various environmental variables modulating these interactions will have to be considered [2]. A particular challenge will be predicting phenotypes or evolutionary responses in the context of changing environments [220]. MTBC faces many different environmental

conditions during its life cycle. These are encountered when penetrating into host cells, surviving within granuloma and cavities, persisting during aerosol transmission, and resisting drug pressure. Recently, a new environment has emerged to which MTBC has to adapt to: HIV co-infection [221]. Epidemiological studies have reported associations between HIV and Lineage 2 [222–224] and Lineage 6 in co-infected TB patients [225]. But how HIV co-infection as a selective pressure might impact the evolutionary trajectory of MTBC on the long run is unclear [146]. All of these complexities call for integrated approaches based on systems biology [226] and systems epidemiology [6] to improve our understanding of the role of variation in the host, the pathogen, and the environment in TB. Such an improved understanding will likely pave the way for novel tools and strategies to better control this important disease.

Acknowledgments

We thank the other members of our group for stimulating discussion. Research in our laboratory of is funded by the NIH grants R01 AI090928, the Swiss National Science Foundation (PP00P3_150750), the European Research Council (309540-EVODRTB), and SystemsX.ch.

References

1. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The Immune Response in Tuberculosis. Annu Rev Immunol. 2013; 31:475–527. [PubMed: 23516984]
2. Comas I, Gagneux S. The Past and Future of Tuberculosis Research. PLoS Pathog. 2009; 5
3. Casanova JL, Abel L. Genetic dissection of immunity to mycobacteria: The human model. Annu Rev Immunol. 2002; 20:581–620. [PubMed: 11861613]
4. 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. [PubMed: 12435764]
5. Lygizos M, Shenoi S, Brooks R, Bhushan A, Brust J, Zelterman D, et al. Natural ventilation reduces high TB transmission risk in traditional homes in rural KwaZulu-Natal, South Africa. BMC Infect Dis. 2013; 13:300. [PubMed: 23815441]
6. Comas I, Gagneux S. A role for systems epidemiology in tuberculosis research. Trends Microbiol. 2011; 19:492–500. [PubMed: 21831640]
7. Coscollá M, Gagneux S. Does *M. tuberculosis* genomic diversity explain disease diversity? Drug Discovery Today: Disease Mechanisms. 2010; 7:e43–e59. [PubMed: 21076640]
8. 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. [PubMed: 20113979]
9. Gagneux S. Genetic Diversity in *Mycobacterium tuberculosis*. 2013; 374:1–25.
10. 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. [PubMed: 19915672]
11. Stucki D, Gagneux S. Single nucleotide polymorphisms in *Mycobacterium tuberculosis* and the need for a curated database. Tuberculosis. 2013; 93:30–39. [PubMed: 23266261]
12. Jagielski T, van Ingen J, Rastogi N, Dziadek J, Mazur PK, Bielecki J. Current Methods in the Molecular Typing of *Mycobacterium tuberculosis* and Other Mycobacteria. Biomed Research International. 2014
13. Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, et al. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proc Natl Acad Sci USA. 2009; 106:17939–17944. [PubMed: 19815525]
14. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, et al. Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam

Resistance within Pneumococcal Mosaic Genes. PLoS Genet. 2014; 10:e1004547. [PubMed: 25101644]

15. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol. 1993; 31:406–409. [PubMed: 8381814]
16. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med. 1994:1703–1709. [PubMed: 7910661]
17. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med. 1994:1710–1716. [PubMed: 7993412]
18. Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, et al. SITVITWEB-- A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Inf Gen Evol. 2012; 12:755–766.
19. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucl Acids Res. 2010; 38:W326–W331. [PubMed: 20457747]
20. 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. [PubMed: 15041743]
21. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic Analysis Distinguishes *Mycobacterium africanum*. J Clin Microbiol. 2004; 42:3594–3599. [PubMed: 15297503]
22. 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. [PubMed: 14679230]
23. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, 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. [PubMed: 15024109]
24. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA. 2006; 103:2869–2873. [PubMed: 16477032]
25. Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere Y-O, Kreiswirth BN, van Soolingen D, et al. Genomic Deletions Classify the Beijing/W Strains as a Distinct Genetic Lineage of *Mycobacterium tuberculosis*. J Clin Microbiol. 2005; 43:3185–3191. [PubMed: 16000433]
26. Alland D, Lacher DW, Hazbon MH, Motiwala AS, Qi W, Fleischmann RD, et al. Role of Large Sequence Polymorphisms (LSPs) in Generating Genomic Diversity among Clinical Isolates of *Mycobacterium tuberculosis* and the Utility of LSPs in Phylogenetic Analysis. J Clin Microbiol. 2007; 45:39–46. [PubMed: 17079498]
27. Namouchi A, Didelot X, Schock U, Gicquel B, Rocha EPC. After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection. Genome Res. 2012; 22:721–734. [PubMed: 22377718]
28. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998; 393:537–544. [PubMed: 9634230]
29. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-Genome Comparison of *Mycobacterium tuberculosis* Clinical and Laboratory Strains. J Bacteriol. 2002; 184:5479–5490. [PubMed: 12218036]
30. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A. 2003; 100:7877–7882. [PubMed: 12788972]
31. Gutacker MM, Smoot JC, Migliaccio CAL, Ricklefs SM, Hua S, Cousins DV, et al. Genome-Wide Analysis of Synonymous Single Nucleotide Polymorphisms in *Mycobacterium tuberculosis*

Complex Organisms: Resolution of Genetic Relationships Among Closely Related Microbial Strains. Genetics. 2002; 162:1533–1543. [PubMed: 12524330]

32. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, et al. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J Infect Dis 2006. Jan 1.2006 :121–128.
33. Filliol I, Motiwala AS, Cavatore M, Qi W, Hernando Hazbon M, Bobadilla del Valle M, et al. Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and Recommendations for a Minimal Standard SNP Set. J Bacteriol. 2006; 188:759–772. [PubMed: 16385065]
34. Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis. 2004:1568–1577. [PubMed: 15498158]
35. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. PLoS Biology. 2008; 6:e311. [PubMed: 19090620]
36. Dos Vultos T, Mestre O, Rauzier J, Golec M, Rastogi N, Rasolofo V, et al. Evolution and Diversity of Clonal Bacteria: The Paradigm of *Mycobacterium tuberculosis*. PLoS ONE. 2008; 3:e1538. [PubMed: 18253486]
37. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, Nubel U, et al. High Resolution Discrimination of Clinical *Mycobacterium tuberculosis* Complex Strains Based on Single Nucleotide Polymorphisms. PLoS ONE. 2012; 7
38. Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, Yeboah-Manu D, et al. Two New Rapid SNP-Typing Methods for Classifying *Mycobacterium tuberculosis* Complex into the Main Phylogenetic Lineages. PLoS ONE. 2012; 7
39. Bergval I, Sengstake S, Brankova N, Levterova V, Abadia E, Tadumaze N, et al. Combined Species Identification, Genotyping, and Drug Resistance Detection of *Mycobacterium tuberculosis* Cultures by MLPA on a Bead-Based Array. PLoS ONE. 2012; 7:e43240. [PubMed: 22916230]
40. Sengstake S, Bablishvili N, Schuitema A, Bzekalava N, Abadia E, de Beer J, et al. Optimizing multiplex SNP-based data analysis for genotyping of *Mycobacterium tuberculosis* isolates. BMC Genom. 2014; 15:572.
41. Cowan LS, Diem L, Brake MC, Crawford JT. Transfer of a *Mycobacterium tuberculosis* genotyping method, spoligotyping, from a reverse line-blot hybridization, membrane-based assay to the Luminex multianalyte profiling system. J Clin Microbiol. 2004; 42:474–477. [PubMed: 14715809]
42. Abadia E, Zhang J, Ritacco V, Kremer K, Ruimy R, Rigouts L, et al. The use of microbead-based spoligotyping for *Mycobacterium tuberculosis* complex to evaluate the quality of the conventional method: Providing guidelines for Quality Assurance when working on membranes. BMC Infect Dis. 2011; 11
43. Gomgnimbou MK, Hernandez-Neuta I, Panaiotov S, Bachiyska E, Palomino JC, Martin A, et al. Tuberculosis-Spoligo-Rifampin-Isoniazid Typing: an All-in-One Assay Technique for Surveillance and Control of Multidrug-Resistant Tuberculosis on Luminex Devices. J Clin Microbiol. 2013; 51:3527–3534. [PubMed: 23966495]
44. Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, et al. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res. 2012; 22:735–745. [PubMed: 22294518]
45. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat Genet. 2014; 46:279–286. [PubMed: 24464101]
46. Roetzer A, Schuback S, Diel R, Gasau F, Ubben T, di Nauta A, et al. Evaluation of *Mycobacterium tuberculosis* typing methods in a four-year study in Schleswig-Holstein, Northern Germany. J Clin Microbiol. 2011 JCM-
47. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah LN, Brodkin E, et al. Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. N Engl J Med. 2011; 364:730–739. [PubMed: 21345102]

48. Walker TM, Clp CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis. 2013; 13:137–146. [PubMed: 23158499]
49. Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rüsch-Gerdes S, et al. Whole Genome Sequencing Reveals Complex Evolution Patterns of Multidrug-Resistant *Mycobacterium tuberculosis* Beijing Strains in Patients. PLoS ONE. 2013; 8:e82551. [PubMed: 24324807]
50. Perez-Lago L, Comas I, Navarro Y, Gonzalez-Candelas F, Herranz M, Bouza E, et al. Whole Genome Sequencing Analysis of Intrapatient Microevolution in *Mycobacterium tuberculosis*: Potential Impact on the Inference of Tuberculosis Transmission. J Infect Dis. 2014; 209:98–108. [PubMed: 23945373]
51. Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, Fortune S, et al. Variation among Genome Sequences of H37Rv Strains of *Mycobacterium tuberculosis* from Multiple Laboratories. J Bacteriol. 2010; 192:3645–3653. [PubMed: 20472797]
52. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet. 2012; 44:106–110. [PubMed: 22179134]
53. Pan Y, Yang X, Duan J, Lu N, Leung AS, Tran V, et al. The whole genome sequence of four BCG vaccine strains. J Bacteriol. 2011 JB-
54. Copin R, Coscolla M, Efstathiadis E, Gagneux G, Ernst JD. Impact of in vitro evolution on antigenic diversity of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG). Vaccine. 2014 IN PRESS.
55. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList-10 years after. Tuberculosis. 2011; 91:1–7. [PubMed: 20980199]
56. Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, et al. SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. Bioinformatics. 2012; 28:2991–2993. [PubMed: 23014632]
57. Steiner A, Stucki D, Coscolla M, Borrell S, Gagneux S. KvarQ: Targeted and direct variant calling from FastQ reads of bacterial genomes. BMC Genom. 2014 IN PRESS.
58. Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, et al. Ecotypes of the *Mycobacterium tuberculosis* complex. J Theor Biol. 2006; 239:220–225. [PubMed: 16242724]
59. Cousins DV, Peet RL, Gaynor WT, Williams SN, Gow BL. Tuberculosis in imported hyrax (Procavia capensis) caused by an ususual variant belonging to the *Mycobacterium tuberculosis* complex. Vet Microbiol. 1994; 42:135–145. [PubMed: 7886928]
60. Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, et al. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. Emerg Infect Dis. 2010:1296–1299. [PubMed: 20678329]
61. Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, et al. Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee. Emerg Infect Dis. 2013; 19:969–976. [PubMed: 23735084]
62. Parsons SDC, Drewe JA, van Pittius NCG, Warren RM, van Heiden PD. Novel Cause of Tuberculosis in Meerkats, South Africa. Emerg Infect Dis. 2013; 19:2004–2007. [PubMed: 24274183]
63. Canetti G. Infection caused by atypical mycobacteria and antituberculous immunity. Lille Med. 1970; 15:280–282. [PubMed: 5446090]
64. van Soolingen D, Hoogenboezem T, de Haas PEW, Hermans PWM, Koedam MA, Teppema KS, 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. [PubMed: 9336935]
65. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat Genet. 2013; 45:172–179. [PubMed: 23291586]
66. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, et al. Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005; 1:e5. [PubMed: 16201017]

67. Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, et al. Progenitor "*Mycobacterium canettii*" Clone Responsible for Lymph Node Tuberculosis Epidemic, Djibouti. Emerg Infect Dis. 2014; 20:21–28. [PubMed: 24520560]
68. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A. 2002; 99:3684–3689. [PubMed: 11891304]
69. Wirth T, Hildebrand F, Allix-Beguec C, Wölbeling F, Kubica T, Kremer K, et al. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. PLoS Pathog. 2008; 4:e1000160. [PubMed: 18802459]
70. Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, et al. Major *Mycobacterium tuberculosis* Lineages Associate with Patient Country of Origin. J Clin Microbiol. 2009; 47:1119–1128. [PubMed: 19213699]
71. 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
72. Gehre F, Antonio M, Fahun F, Odoun M, Uwizeye C, de Rijk P, et al. The First Phylogeographic Population Structure and Analysis of Transmission Dynamics of *M. africanum* West African Combining Molecular Data from Benin, Nigeria and Sierra Leone. PLoS ONE. 2013; 8:e77000. [PubMed: 24143198]
73. Blouin Y, Hauck Y, Soler C, Fabre M, Vong R, Dehan C, et al. Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching *Mycobacterium tuberculosis* Clade. PLoS ONE. 2012; 7
74. Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial Lineages Causing Pulmonary and Extrapulmonary Tuberculosis, Ethiopia. Emerg Infect Dis. 2013; 19:460–463. [PubMed: 23622814]
75. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014 advance online publication.
76. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat Genet. 2013; 45:1176–U311. [PubMed: 23995134]
77. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet. 2010; 42:498–503. [PubMed: 20495566]
78. 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 Bio Evol. 2013; 5:1849–1862. [PubMed: 24115728]
79. Achtman M. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. Annu Rev Microbiol. 2008; 62:53–70. [PubMed: 18785837]
80. Shih CH, Chang CM, Lin YS, Lo WC, Hwang JK. Evolutionary information hidden in a single protein structure. Proteins. 2012; 80:1647–1657. [PubMed: 22454236]
81. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. Nat Genet. 2013; 45:1183–U320. [PubMed: 23995135]
82. Zhang HT, Li DF, Zhao LL, Fleming J, Lin N, Wang T, et al. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet. 2013; 45:1255–U217. [PubMed: 23995137]
83. Osorio NS, Rodrigues F, Gagneux S, Pedrosa J, Pinto-Carbo M, Castro AG, 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. [PubMed: 23449927]
84. Lipsitch M, O'Hagan J. Patterns of antigenic diversity and the mechanisms that maintain them. Journal of The Royal Society Interface. 2007; 4:787–802.
85. Rodrigo T, Cayla JA, deOlalla PG, GaldosTanguis H, Jansa JM, Miranda P, et al. Characteristics of tuberculosis patients who generate secondary cases. Int J Tuberc Lung Dis. 1997; 1:352–357. [PubMed: 9432392]

86. Gehre F, Otu J, DeRiemer K, de Sessions PF, Hibberd ML, Mulders W, et al. Deciphering the Growth Behaviour of *Mycobacterium africanum*. PLoS Negl Trop Dis. 2013; 7:e2220. [PubMed: 23696911]
87. Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. Modern Lineages of *Mycobacterium tuberculosis* Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in Human Monocyte-Derived Macrophages. PLoS ONE. 2012; 7:e43170. [PubMed: 22916219]
88. Homolka S, Niemann S, Russell DG, Rohde KH. Functional Genetic Diversity among *Mycobacterium tuberculosis* Complex Clinical Isolates: Delineation of Conserved Core and Lineage-Specific Transcriptomes during Intracellular Survival. PLoS Pathog. 2010; 6:e1000988. [PubMed: 20628579]
89. Krishnan N, Malaga W, Constant P, Caws M, Thi Hoang Chau T, Salmons J, et al. *Mycobacterium tuberculosis* Lineage Influences Innate Immune Response and Virulence and Is Associated with Distinct Cell Envelope Lipid Profiles. PLoS ONE. 2011; 6:e23870. [PubMed: 21931620]
90. Khatri B, Fielder M, Jones G, Newell W, Abu-Oun M, Wheeler PR. High Throughput Phenotypic Analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* Strains' Metabolism Using Biolog Phenotype Microarrays. PLoS ONE. 2013; 8:e52673. [PubMed: 23326347]
91. Portevin D, Sukumar S, Coscolla M, Shui G, Li B, Guan XL, et al. Lipidomics and genomics of *Mycobacterium tuberculosis* reveal lineage-specific trends in mycolic acid biosynthesis. Microbiology Open. 2014
92. Vander Beken S, Al Dulayymi JR, Naessens T, Koza G, Maza-Iglesias M, Rowles R, et al. Molecular structure of the *Mycobacterium tuberculosis* virulence factor, mycolic acid, determines the elicited inflammatory pattern. Eur J Immunol. 2011; 41:450–460. [PubMed: 21268014]
93. Yuan Y, Lee RE, Besra GS, Belisle JT, Barry CE. Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA. 1995; 92:6630–6634. [PubMed: 7604045]
94. Muller B, Borrell S, Rose G, Gagneux S. The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. Trends Genet. 2013; 29:160–169. [PubMed: 23245857]
95. Trauner A, Borrell S, Reither K, Gagneux S. Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. Drugs. 2014:1063–1072. [PubMed: 24962424]
96. Ryndak M, Wang S, Smith I. PhoP, a key player in *Mycobacterium tuberculosis* virulence. Trends Microbiol. 2008; 16:528–534. [PubMed: 18835713]
97. Chesne-Seck ML, Barilone N, Boudou F, Asensio JG, Kolattukudy PE, Martin C, et al. A Point Mutation in the Two-Component Regulator PhoP-PhoR Accounts for the Absence of Polyketide-Derived Acyltrehaloses but Not That of Phthiocerol Dimycocerosates in *Mycobacterium tuberculosis* H37Ra. J Bacteriol. 2008; 190:1329–1334. [PubMed: 18065542]
98. Gonzalo-Asensio, Js; Malaga, W.; Pawlik, A.; Astarie-Dequeker, C.; Passemar, C.; Moreau, F., et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci USA. 2014
99. Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, et al. DNA Methylation Impacts Gene Expression and Ensures Hypoxic Survival of *Mycobacterium tuberculosis*. PLoS Pathog. 2013; 9:e1003419. [PubMed: 23853579]
100. Ando H, Miyoshi-Akiyama T, Watanabe S, Kirikae T. A silent mutation in mabA confers isoniazid resistance on *Mycobacterium tuberculosis*. Mol Microbiol. 2014; 91:538–547. [PubMed: 24354762]
101. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, et al. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science. 1994; 263:227–230. [PubMed: 8284673]
102. Guo H, Seet Q, Denkin S, Parsons L, Zhang Y. Molecular characterization of isoniazid-resistant clinical isolates of *Mycobacterium tuberculosis* from the USA. J Med Microbiol. 2006; 55:1527–1531. [PubMed: 17030912]
103. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001; 409:1007–1011. [PubMed: 11234002]

104. Bentley SD, Comas I, Bryant JM, Walker D, Smith NH, Harris SR, et al. The Genome of *Mycobacterium Africanum* West African 2 Reveals a Lineage-Specific Locus and Genome Erosion Common to the *M. tuberculosis*. Complex PLoS Negl Trop Dis. 2012; 6:e1552.

105. Wagner A. Gene duplications, robustness and evolutionary innovations. BioEssays. 2008; 30:367–373. [PubMed: 18348184]

106. Gey van Pittius N, Sampson S, Lee H, Kim Y, van Helden P, Warren R. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol. 2006; 6:95. [PubMed: 17105670]

107. Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R. Defining putative T cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. Vaccine. 2005; 23:1265–1272. [PubMed: 15652669]

108. Chaitra MG, Nayak R, Shaila MS. Modulation of immune responses in mice to recombinant antigens from PE and PPE families of proteins of *Mycobacterium tuberculosis* by the Ribi adjuvant. Vaccine. 2007; 25:7168–7176. [PubMed: 17709160]

109. Chaitra MG, Shaila MS, Nayak R. Detection of interferon gamma-secreting CD8+ T lymphocytes in humans specific for three PE/PPE proteins of *Mycobacterium tuberculosis*. Microbes and Infection. 2008; 10:858–867. [PubMed: 18653370]

110. Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L, Yang Z. Variation of the *Mycobacterium tuberculosis* PE_PGRS33 Gene among Clinical Isolates. J Clin Microbiol. 2005; 43:4954–4960. [PubMed: 16207947]

111. Karboul A, Gey van Pittius N, Namouchi A, Vincent V, Sola C, Rastogi N, et al. Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair. BMC Evol Biol. 2006; 6:107. [PubMed: 17163995]

112. Talarico S, Zhang L, Marrs CF, Foxman B, Cave MD, Brennan MJ, et al. *Mycobacterium tuberculosis* PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. Tuberculosis. 2008; 88:283–294. [PubMed: 18313360]

113. McEvoy C, van Helden P, Warren R, van Pittius N. Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis* PPE38 gene region. BMC Evol Biol. 2009; 9:237. [PubMed: 19769792]

114. Copin R, Coscolla M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G, et al. Sequence Diversity in the pe_pgrs Genes of *Mycobacterium tuberculosis* Is Independent of Human T Cell Recognition. mBio. 2014; 5

115. Dheenadhayalan V, Delogu G, Brennan MJ. Expression of the PE_PGRS 33 protein in *Mycobacterium smegmatis* triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes Infect. 2005

116. Yu G, Fu X, Jin K, Zhang L, Wu W, Cui Z, et al. Integrative analysis of transcriptome and genome indicates two potential genomic islands are associated with pathogenesis of *Mycobacterium tuberculosis*. Gene. 2011; 489:21–29. [PubMed: 21924330]

117. Domenech P, Kolly GS, Leon-Solis L, Fallow A, Reed MB. Massive Gene Duplication Event among Clinical Isolates of the *Mycobacterium tuberculosis* W/Beijing Family. J Bacteriol. 2010; 192:4562–4570. [PubMed: 20639330]

118. Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, Plikaytis BB, et al. Independent Large Scale Duplications in Multiple *M. tuberculosis* Lineages Overlapping the Same Genomic Region. PLoS ONE. 2012; 7:e26038. [PubMed: 22347359]

119. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST. Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. Yeast. 2000; 17:111–123. [PubMed: 10900457]

120. Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, et al. Genome plasticity of BCG and impact on vaccine efficacy. Proc Natl Acad Sci USA. 2007; 104:5596–5601. [PubMed: 17372194]

121. Fenner L, Malla B, Ninet B, Dubuis O, Stucki D, Borrell S, et al. Pseudo-Beijing: Evidence for Convergent Evolution in the Direct Repeat Region of *Mycobacterium tuberculosis*. PLoS ONE. 2011; 6:e24737. [PubMed: 21935448]

122. Julier C, Hyer RN, Davies J, Merlin F, Soularue P, Briant L, et al. Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. Nature. 1991; 354:155–159. [PubMed: 1944595]

123. Jordan P, Snyder L, Saunders N. Strain-specific differences in *Neisseria gonorrhoeae* associated with the phase variable gene repertoire. BMC Microbiol. 2005; 5:21. [PubMed: 15857514]

124. Alamro M, Bidmos FA, Chan H, Oldfield NJ, Newton E, Bai X, et al. Phase Variation Mediates Reductions in Expression of Surface Proteins during Persistent Meningococcal Carriage. Infect Immun. 2014; 82:2472–2484. [PubMed: 24686058]

125. Yindeeyoungyeon W, Likitvivatanavong S, Palittapongarnpim P. Characterization of alpha-isopropylmalate synthases containing different copy numbers of tandem repeats in *Mycobacterium tuberculosis*. BMC Microbiol. 2009; 9:122. [PubMed: 19505340]

126. Olsen I, Balasingham SV, Davidsen T, Debebe E, Rodland EA, van Soolingen D, et al. Characterization of the major formamidopyrimidine-DNA glycosylase homolog in *Mycobacterium tuberculosis* and its linkage to variable tandem repeats. FEMS Immunology & Medical Microbiology. 2009; 56:151–161. [PubMed: 19496823]

127. Tantivitayakul P, Panapruksachat S, Billamas P, Palittapongarnpim P. Variable number of tandem repeat sequences act as regulatory elements in *Mycobacterium tuberculosis*. Tuberculosis. 2010; 90:311–318. [PubMed: 20810321]

128. Perez-Lago L, Navarro Y, Herranz M, Bouza E, Garcia-de-Viedma D. Differences in gene expression between clonal variants of *Mycobacterium tuberculosis* emerging as a result of microevolution. Int J Med Microbiol. 2013; 303:674–677. [PubMed: 24189285]

129. Soto CY, Menendez MC, Perez E, Samper S, Gomez AB, Garcia MJ, et al. IS6110 mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol. 2004; 42:212–219. [PubMed: 14715755]

130. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, et al. IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. Mol Microbiol. 2004; 52:999–1012. [PubMed: 15130120]

131. Alonso H, Aguilo JI, Samper Sa, Caminero JA, Campos-Herrero MaI, Gicquel B, et al. Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. Tuberculosis. 2011; 91:117–126. [PubMed: 21256084]

132. Millan-Lou MI, Lopez-Calleja AI, Colmenarejo C, Lezcano MA, Vitoria MA, Del Portillo P, et al. Global Study of IS6110 in a Successful *Mycobacterium tuberculosis* Strain: Clues for Deciphering Its Behavior and for Its Rapid Detection. J Clin Microbiol. 2013; 51:3631–3637. [PubMed: 23985924]

133. Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr M. Genomic Deletions Suggest a Phylogeny for the *Mycobacterium tuberculosis* Complex. J Infect Dis. 2002; 186:74–80. [PubMed: 12089664]

134. Zhang Y, Young D. Strain variation in the katG region of *Mycobacterium tuberculosis*. Mol Microbiol. 1994; 14:301–308. [PubMed: 7830574]

135. Rouse DA, Morris SL. Molecular mechanisms of isoniazid resistance in *Mycobacterium tuberculosis* and *Mycobacterium bovis*. Infect Immun. 1995; 63:1427–1433. [PubMed: 7890405]

136. Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M, et al. Role of the pks15/1 Gene in the Biosynthesis of Phenolglycolipids in the *Mycobacterium tuberculosis* Complex. J Biol Chem. 2002; 277:38148–38158. [PubMed: 12138124]

137. Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, et al. The Phenolic Glycolipid of *Mycobacterium tuberculosis* Differentially Modulates the Early Host Cytokine Response but Does Not in Itself Confer Hypervirulence. Infect Immun. 2008; 76:3027–3036. [PubMed: 18443098]

138. Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, et al. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature. 2004; 431:84–87. [PubMed: 15343336]

139. Cambier CJ, Takaki KK, Larson RP, Hernandez RE, Tobin DM, Urdahl KB, et al. Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature. 2014; 505:218–222. [PubMed: 24336213]

140. Newton SM, Smith RJ, Wilkinson KA, Nicol MP, Garton NJ, Staples KJ, et al. A deletion defining a common Asian lineage of *Mycobacterium tuberculosis* associates with immune subversion. Proc Natl Acad Sci USA. 2006; 103:15594–15598. [PubMed: 17028173]

141. Kong Y, Cave MD, Yang D, Zhang L, Marrs CF, Foxman B, et al. Distribution of Insertion- and Deletion-Associated Genetic Polymorphisms among Four *Mycobacterium tuberculosis* Phospholipase C Genes and Associations with Extrathoracic Tuberculosis: a Population-Based Study. J Clin Microbiol. 2005; 43:6048–6053. [PubMed: 16333097]

142. Sharma K, Gupta M, Pathak M, Gupta N, Koul A, Sarangi S, et al. Transcriptional Control of the Mycobacterial embCAB Operon by PknH through a Regulatory Protein, EmbR, In Vivo. J Bacteriol. 2006; 188:2936–2944. [PubMed: 16585755]

143. Briken V, Porcelli SA, Besra GS, Kremer L. Mycobacterial lipoarabinomannan and related lipoglycans: from biogenesis to modulation of the immune response. Mol Microbiol. 2004; 53:391–403. [PubMed: 15228522]

144. Basu S, Pathak SK, Banerjee A, Pathak S, Bhattacharyya A, Yang Z, et al. Execution of Macrophage Apoptosis by PE_PGRS33 of *Mycobacterium tuberculosis* Is Mediated by Toll-like Receptor 2-dependent Release of Tumor Necrosis Factor-α. J Biol Chem. 2007; 282:1039–1050. [PubMed: 17095513]

145. Talarico S, Donald Cave M, Foxman B, Marrs CF, Zhang L, Bates JH, et al. Association of *Mycobacterium tuberculosis* PE_PGRS33 polymorphism with clinical and epidemiological characteristics. Tuberculosis. 2007; 87:338–346. [PubMed: 17475562]

146. Brites D, Gagneux S. Old and new selective pressures on *Mycobacterium tuberculosis*. Inf Gen Evol. 2012; 12:678–685.

147. Ernst JD. The immunological life cycle of tuberculosis. Nat Rev Immunol. 2012; 12:581–591. [PubMed: 22790178]

148. Redford PS, Murray PJ, O'Garra A. The role of IL-10 in immune regulation during *M. tuberculosis* infection. Mucosal Immunol. 2011; 4:261–270. [PubMed: 21451501]

149. Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, et al. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFNγ. Proc Natl Acad Sci U S A. 2001; 98:5752–5757. [PubMed: 11320211]

150. Li Q, Whalen CC, Albert JM, Larkin R, Zukowski L, Cave MD, et al. Differences in Rate and Variability of Intracellular Growth of a Panel of *Mycobacterium tuberculosis* Clinical Isolates within a Human Monocyte Model. Infect Immun. 2002; 70:6489–6493. [PubMed: 12379735]

151. Manca C, Reed MB, Freeman S, Mathema B, Kreiswirth B, Barry CE III, et al. Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. Infect Immun. 2004:5511–5514. [PubMed: 15322056]

152. Manca C, Tsenova L, Freeman S, Barczak AK, Tovey M, Murray PJ, et al. Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway. J Interferon Cytokine Res. 2005:694–701. [PubMed: 16318583]

153. Tsenova L, Ellison E, Harbacheuski R, Moreira A, Kurepina N, Reed M, et al. Virulence of Selected *Mycobacterium tuberculosis* Clinical Isolates in the Rabbit Model of Meningitis Is Dependent on Phenolic Glycolipid Produced by the Bacilli. J Infect Dis. 2005; 192:98–106. [PubMed: 15942899]

154. Chacón-Salinas R, Serafín-López J, Ramos-Payín R, Méndez-Aragón P, Hernández-Pando R, van Soolingen D, et al. Differential pattern of cytokine expression by macrophages infected in vitro with different *Mycobacterium tuberculosis* genotypes. Clinical & Experimental Immunology. 2005; 140:443–449. [PubMed: 15932505]

155. Rocha-Ramirez LM, Estrada-Garcia I, Lopez-Marin LM, Segura-Salinas E, Mendez-Aragon P, van Soolingen D, et al. *Mycobacterium tuberculosis* lipids regulate cytokines, TLR-2/4 and MHC

class II expression in human macrophages. Tuberculosis. 2008; 88:212–220. [PubMed: 18222732]

156. Subbian S, Bandyopadhyay N, Tsenova L, O'Brien P, Khetani V, Kushner N, et al. Early innate immunity determines outcome of *Mycobacterium tuberculosis* pulmonary infection in rabbits. Cell Commun Sign. 2013; 11:60.
157. Portevin D, Gagneux S, Comas I, Young D. Human Macrophage Responses to Clinical Isolates from the *Mycobacterium tuberculosis* Complex Discriminate between Ancient and Modern Lineages. PLoS Pathog. 2011; 7:e1001307. [PubMed: 21408618]
158. Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, et al. Clade-Specific Virulence Patterns of *Mycobacterium tuberculosis* Complex Strains in Human Primary Macrophages and Aerogenically Infected Mice. mBio. 2013; 4
159. Chen YY, Chang JR, Huang WF, Hsu SC, Kuo SC, Sun JR, et al. The Pattern of Cytokine Production *In Vitro* Induced by Ancient and Modern Beijing *Mycobacterium tuberculosis* Strains. PLoS ONE. 2014; 9:e94296. [PubMed: 24728339]
160. Mihret A, Bekele Y, Loxton AG, Aseffa A, Howe R, Walzl G. Plasma Level of IL-4 Differs in Patients Infected with Different Modern Lineages of *M. tuberculosis*. J Trop Med. 2012; 2012:518564. [PubMed: 23049571]
161. Tanveer M, Hasan Z, Kanji A, Hussain R, Hasan R. Reduced TNF-[alpha] and IFN-[gamma] responses to Central Asian strain 1 and Beijing isolates of *Mycobacterium tuberculosis* in comparison with H37Rv strain. Trans R Soc Trop Med Hyg. 2009; 103:581–587. [PubMed: 19375139]
162. Wang C, Peyron P, Mestre O, Kaplan G, van Soolingen D, Gao Q, et al. Innate Immune Response to *Mycobacterium tuberculosis* Beijing and Other Genotypes. PLoS ONE. 2010; 5:e13594. [PubMed: 21049036]
163. van Laarhoven A, Mandemakers JJ, Kleinnijenhuis J, Enaimi M, Lachmandas E, Joosten LAB, et al. Low Induction of Proinflammatory Cytokines Parallels Evolutionary Success of Modern Strains within the *Mycobacterium tuberculosis* Beijing Genotype. Infect Immun. 2013; 81:3750–3756. [PubMed: 23897611]
164. Carmona J, Cruz A, Moreira-Teixeira L, Sousa C, Sousa J, Osorio NS, et al. *Mycobacterium tuberculosis* Strains Are Differentially Recognized by TLRs with an Impact on the Immune Response. PLoS ONE. 2013; 8
165. Coussens AK, Wilkinson RJ, Nikolayevskyy V, Elkington PT, Hanifa Y, Islam K, et al. Ethnic Variation in Inflammatory Profile in Tuberculosis. PLoS Pathog. 2013; 9:e1003468. [PubMed: 23853590]
166. Stavrum R, PrayGod G, Range N, Faurholt-Jepsen D, Jeremiah K, Faurholt-Jepsen M, et al. Increased level of acute phase reactants in patients infected with modern *Mycobacterium tuberculosis* genotypes in Mwanza, Tanzania. BMC Infect Dis. 2014; 14:309. [PubMed: 24903071]
167. de Jong B, Hill P, Aiken A, Awine T, Antonio M, Adetifa I, et al. Progression to Active Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia. J Infect Dis. 2008; 198:1037–1043. [PubMed: 18702608]
168. Via LE, Weiner DM, Schimel D, Lin PL, Dayao E, Tankersley SL, et al. Differential Virulence and Disease Progression following *Mycobacterium tuberculosis* Complex Infection of the Common Marmoset (*Callithrix jacchus*). Infect Immun. 2013; 81:2909–2919. [PubMed: 23716617]
169. Lan NT, Lien HT, Tung IB, Borgdorff MW, Kremer K, van Soolingen D. *Mycobacterium tuberculosis* Beijing genotype and risk for treatment failure and relapse, Vietnam. Emerg Infect Dis. 2003:1633–1635. [PubMed: 14720411]
170. Sun YJ, Lee AS, Wong SY, Paton NI. Association of *Mycobacterium tuberculosis* Beijing genotype with tuberculosis relapse in Singapore. Epidemiol Infect. 2006:329–332. [PubMed: 16207386]
171. Huyen MNT, Buu TN, Tiemersma E, Lan NTN, Dung NH, Kremer K, et al. Tuberculosis Relapse in Vietnam Is Significantly Associated With *Mycobacterium tuberculosis* Beijing Genotype Infections. J Infect Dis. 2013; 207:1516–1524. [PubMed: 23408848]

172. Parwati I, Alisjahbana B, Apriani L, Soetikno R, Ottenhoff T, van der Zanden A, et al. *Mycobacterium tuberculosis* Beijing Genotype Is an Independent Risk Factor for Tuberculosis Treatment Failure in Indonesia. J Infect Dis. 2010; 201:553–557. [PubMed: 20064071]

173. van Crevel R, Nelwan RHH, de Lenne W, Veeraragu Y, van der Zanden AG, Amin Z, et al. *Mycobacterium tuberculosis* Beijing genotype strains associated with febrile response to treatment. Emerg Infect Dis. 2001; 7:880–883. [PubMed: 11747703]

174. Ogarkov O, Mokrousov I, Sinkov V, Zhdanova S, Antipina S, Savilov E. Lethal combination of *Mycobacterium tuberculosis* Beijing genotype and human CD209 – 336G allele in Russian male population. Infect Gen Evol. 2012; 12:732–736.

175. Nahid P, Bliven EE, Kim EY, Mac Kenzie WR, Stout JE, Diem L, et al. Influence of *M. tuberculosis* Lineage Variability within a Clinical Trial for Pulmonary Tuberculosis. PLoS ONE. 2010; 5:e10753. [PubMed: 20505778]

176. Kato-Maeda M, Shanley CA, Ackart D, Jarlsberg LG, Shang S, Obregon-Henao A, et al. Beijing Sublineages of *Mycobacterium tuberculosis* Differ in Pathogenicity in the Guinea Pig. Clin Vaccine Immunol. 2012; 19:1227–1237. [PubMed: 22718126]

177. Aguilar D, Hanekom M, Mata D, van Pittius NCG, van Helden PD, Warren RM, et al. *Mycobacterium tuberculosis* strains with the Beijing genotype demonstrate variability in virulence associated with transmission. Tuberculosis. 2010; 90:319–325. [PubMed: 20832364]

178. Zumla A, Raviglione M, Hafner R, Fordham von Reyn C. Tuberculosis. N Engl J Med. 2013; 368:745–755. [PubMed: 23425167]

179. Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, et al. Population-Based Study of Deletions in Five Different Genomic Regions of *Mycobacterium tuberculosis* and Possible Clinical Relevance of the Deletions. J Clin Microbiol. 2006; 44:3940–3946. [PubMed: 16957037]

180. Kong Y, Cave MD, Zhang L, Foxman B, Marrs CF, Bates JH, et al. Association between *Mycobacterium tuberculosis* Beijing/W Lineage Strain Infection and Extrathoracic Tuberculosis: Insights from Epidemiologic and Clinical Characterization of the Three Principal Genetic Groups of *M. tuberculosis*. Clinical Isolates J Clin Microbiol. 2007; 45:409–414.

181. Pareek M, Evans J, Innes J, Smith G, Hingley-Wilson S, Lougheed KE, et al. Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype. Thorax. 2013; 68:221–229. [PubMed: 23019255]

182. Click ES, Moonan PK, Winston CA, Cowan LS, Oeltmann JE. Relationship Between *Mycobacterium tuberculosis* Phylogenetic Lineage and Clinical Site of Tuberculosis. Clin Infect Dis. 2012; 54:211–219. [PubMed: 22198989]

183. Caws M, Thwaites G, Dunstan S, Hawn TR, Thi Ngoc Lan N, Thuong NTT, et al. The Influence of Host and Bacterial Genotype on the Development of Disseminated Disease with *Mycobacterium tuberculosis*. PLoS Pathog. 2008; 4:e1000034. [PubMed: 18369480]

184. Borgdorff MW, van Deutekom H, de Haas PEWP, Kremer K, van Soolingen D. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. Tuberculosis. 2004; 84:337–340. [PubMed: 15207809]

185. Nicol MP, Sola C, February B, Rastogi N, Steyn L, Wilkinson RJ. Distribution of Strain Families of *Mycobacterium tuberculosis* Causing Pulmonary and Extrapulmonary Disease in Hospitalized Children in Cape Town, South Africa. J Clin Microbiol. 2005; 43:5779–5781. [PubMed: 16272518]

186. Wampande E, Mupere E, Debanne S, Asiimwe B, Nsereko M, Mayanja H, 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. [PubMed: 24134504]

187. van der Spuy GD, Warren RM, Richardson M, Beyers N, Behr MA, van Helden PD. Use of Genetic Distance as a Measure of Ongoing Transmission of *Mycobacterium tuberculosis*. J Clin Microbiol. 2003; 41:5640–5644. [PubMed: 14662954]

188. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: what have we learnt from molecular epidemiology? Clin Microbiol Infect. 2013; 19:889–901. [PubMed: 23731470]

189. Buu TN, van Soolingen D, Huyen MNT, Lan NTN, Quy HT, Tiemersma EW, et al. Increased Transmission of *Mycobacterium tuberculosis* Beijing Genotype Strains Associated with

Resistance to Streptomycin: A Population-Based Study. PLoS ONE. 2012; 7:e42323. [PubMed: 22912700]

190. Yang C, Luo T, Sun G, Qiao K, Sun G, DeRiemer K, et al. *Mycobacterium tuberculosis* Beijing Strains Favor Transmission but Not Drug Resistance in China. Clin Infect Dis. 2012; 55:1179–1187. [PubMed: 22865872]
191. Anh DD, Borgdorff MW, Van LN, Lan NT, van Gorkom T, Kremer K, et al. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg Infect Dis. 2000:302–305. [PubMed: 10827122]
192. European Concerted Action on New Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerg Infect Dis 2006. May.2006 :736–743.
193. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CRE, Kidd M, et al. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family was associated with an increased ability to spread and cause disease. J Clin Microbiol. 2007 JCM-
194. Cowley D, Govender D, February B, Wolfe M, Steyn L, Evans J, et al. Recent and Rapid Emergence of W–Beijing Strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. Clin Infect Dis. 2008; 47:1252–1259. [PubMed: 18834315]
195. van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R, Marais BJ, et al. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. Tuberculosis. 2009; 89:120–125. [PubMed: 19054717]
196. Tuite AR, Guthrie JL, Alexander DC, Whelan MS, Lee B, Lam K, et al. Epidemiological evaluation of spatiotemporal and genotypic clustering of *Mycobacterium tuberculosis* in Ontario, Canada. Int J Tuberc Lung Dis 2013. Oct.2013 :1322–1327.
197. Kubica T, Rusch-Gerdes S, Niemann S. The Beijing genotype is emerging among multidrug-resistant *Mycobacterium tuberculosis* strains from Germany. Int J Tuberc Lung Dis. 2004; 8:1107–1113. [PubMed: 15455596]
198. Lillebaek T, Andersen AB, Dirksen A, Glynn JR, Kremer K. *Mycobacterium tuberculosis* Beijing genotype. Emerg Infect Dis. 2003; 9:1553–1557. [PubMed: 14720395]
199. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. PLoS One. 2011; 2006:e25075. [PubMed: 21949856]
200. Marais BJ, Hesseling AC, Schaaf HS, Gie RP, van Helden PD, Warren RM. *Mycobacterium tuberculosis* Transmission Is Not Related to Household Genotype in a Setting of High Endemicity. J Clin Microbiol. 2009; 47:1338–1343. [PubMed: 19261801]
201. Langlois-Klassen D, Senthilselvan A, Chui L, Kunimoto D, Saunders LD, Menzies D, et al. Transmission of *Mycobacterium tuberculosis* Beijing Strains, Alberta, Canada, 1991–2007. Emerg Infect Dis. 2013; 19:701–711. [PubMed: 23648234]
202. Kato-Maeda M, Kim EY, Flores L, Jarlsberg LG, Osmond D, Hopewell PC. Differences among sublineages of the East-Asian lineage of *Mycobacterium tuberculosis* in genotypic clustering. Int J Tuberc Lung Dis 2010. May.2010 :538–544.
203. Anderson J, Jarlsberg LG, Grindsdale J, Osmond D, Kawamura M, Hopewell PC, et al. Sublineages of lineage 4 (Euro-American) *Mycobacterium tuberculosis* differ in genotypic clustering. Int J Tuberc Lung Dis 2013. Jul.2013 :885–891.
204. Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, Sola C, et al. Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. J Clin Microbiol. 2003; 41:2547–2553. [PubMed: 12791879]
205. Godreuil S, Torrea G, Terru D, Chevenet F, Diagbouga S, Supply P, et al. First Molecular Epidemiology Study of *Mycobacterium tuberculosis* in Burkina Faso. J Clin Microbiol. 2007; 45:921–927. [PubMed: 17251410]
206. Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, et al. The Guinea-Bissau Family of *Mycobacterium tuberculosis* Complex Revisited. PLoS ONE. 2011; 6:e18601. [PubMed: 21533101]

207. Hanekom M, van Pittius NCG, McEvoy C, Victor TC, van Helden PD, Warren RM. *Mycobacterium tuberculosis* Beijing genotype: A template for success. Tuberculosis. 2011; 91:510–523. [PubMed: 21835699]

208. Borrell S, Gagneux S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis* [State of the art]. Int J Tuberc Lung Dis. 2009; 13:1456–1466. [PubMed: 19919762]

209. Lopez B, Aguilar D, Orozco H, Burguer M, Espitia C, Ritacco V, et al. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin Exp Immunol. 2003; 133:30–37. [PubMed: 12823275]

210. Abebe F, Bjune G. The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette–Guerin (BCG) vaccines: is there a link? Clinical & Experimental Immunology. 2006; 145:389–397. [PubMed: 16907905]

211. Tsenova L, Harbacheuski R, Sung N, Ellison E, Fallows D, Kaplan G. BCG vaccination confers poor protection against M. tuberculosis HN878-induced central nervous system disease. Vaccine. 2007; 25:5126–5132. [PubMed: 17241704]

212. Gagneux S. Host–pathogen coevolution in human tuberculosis. Phil Trans R Soc B: Biological Sciences. 2012; 367:850–859.

213. Brudey K, Driscoll J, Rigouts L, Prodinger W, Gori A, Al Hajoj S, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006; 6:23. [PubMed: 16519816]

214. Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, Battegay M, et al. HIV Infection Disrupts the Sympatric Host-Pathogen Relationship in Human Tuberculosis. PLoS Genet. 2013; 9:e1003318. [PubMed: 23505379]

215. Herb F, Thye T, Niemann S, Browne ENL, Chinbuah MA, Gyapong J, et al. ALOX5 variants associated with susceptibility to human pulmonary tuberculosis. Hum Mol Genet. 2008; 17:1052–1060. [PubMed: 18174194]

216. van Crevel R, Parwati I, Sahiratmadja E, Marzuki S, Ottenhoff T, Netea M, et al. Infection with *Mycobacterium tuberculosis* Beijing Genotype Strains Is Associated with Polymorphisms in SLC11A1/NRAMP1 in Indonesian Patients with Tuberculosis. J Infect Dis. 2009; 200:1671–1674. [PubMed: 19863441]

217. Intemann CD, Thye T, Niemann S, Browne ENL, Amanua Chinbuah M, Enimil A, 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.

218. Bartha IC. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. eLife Sciences. 2013; 2

219. Kodaman N, Sobota R, Mera R, Schneider BG, Williams SM. Disrupted human-pathogen co-evolution: a model for disease. Frontiers in Genetics. 2014; 5

220. Reed LK, Lee K, Zhang Z, Rashid L, Poe A, Hsieh B, et al. Systems Genomics of Metabolic Phenotypes in Wild-Type Drosophila melanogaster. Genetics. 2014

221. Kwan CK, Ernst JD. HIV and Tuberculosis: a Deadly Human Syndemic. Clin Microbiol Rev. 2011; 24:351–376. [PubMed: 21482729]

222. Caws M, Thwaites G, Stepniewska K, Lan NTN, Duyen NTH, Phuong NT, et al. Beijing genotype of *Mycobacterium tuberculosis* significantly associated with HIV and multi-drug resistance in tuberculous meningitis. J Clin Microbiol. 2006 JCM-

223. Middelkoop K, Bekker L, Mathema B, Shashkina E, Kurepina N, Whitelaw A, et al. Molecular Epidemiology of *Mycobacterium tuberculosis* in a South African Community with High HIV Prevalence. J Infect Dis. 2009; 200:1207–1211. [PubMed: 19764885]

224. Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, Gudo PS, et al. *Mycobacterium tuberculosis* Beijing Genotype Is Associated with HIV Infection in Mozambique. PLoS ONE. 2013; 8:e71999. [PubMed: 23940801]

225. de Jong BC, Adetifa I, Walther B, Hill PC, Antonio M, Ota M, 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. [PubMed: 20002176]

226. Young D, Stark J, Kirschner D. Systems biology of persistent infection: tuberculosis as a case study. Nat Rev Micro. 2008; 6:520–528.
227. Rose G. A genomic and transcriptomic study of lineage-specific variation in *Mycobacterium tuberculosis*. 2013
228. Galagan JE, Sisk P, Stolte C, Weiner B, Koehrsen M, Wymore F, et al. TB database 2010: Overview and update. Tuberculosis. 2010; 90:225–235. [PubMed: 20488753]
229. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucl Acids Res. 2014; 42:D581–D591. [PubMed: 24225323]
230. Coll F, Preston M, Guerra-Assunao JA, Hill-Cawthorn G, Harris D, Perdigao J, et al. PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. Tuberculosis. 2014; 94:346–354. [PubMed: 24637013]
231. Chernyaeva E, Shulgina M, Rotkevich M, Dobrynin P, Simonov S, Shitikov E, et al. Genome-wide *Mycobacterium tuberculosis* variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genom. 2014; 15:308.
232. Kohl TA, Diel R, Harmsen D, Rothgänger J, Meywald Walter K, Merker M, et al. Whole genome based *Mycobacterium tuberculosis* surveillance: A standardized, portable and expandable approach. J Clin Microbiol. 2014

Highlights

- Human-adapted MTBC comprises seven phylogenetic lineages
- MTBC strains belonging to separate lineages differ by 1,200 single nucleotide polymorphisms on average
- MTBC lineages are associated with different geographical regions
- Lineage 2 (includes Beijing) and Lineage 4 are more geographically widespread than other lineages
- Lineage 2 and Lineage 4 are more pathogenic and transmissible than geographically restricted lineages

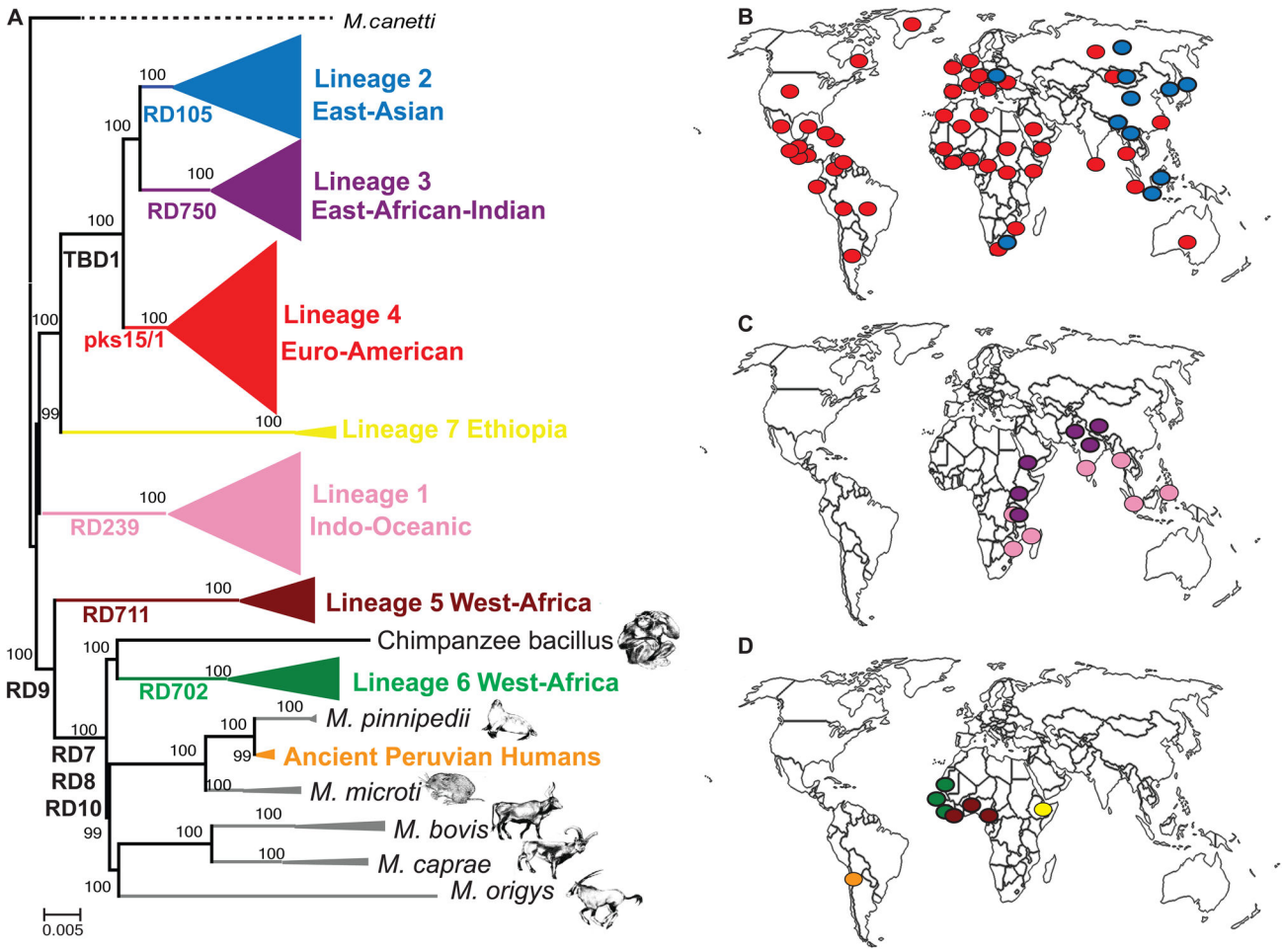

Figure 1.

A. Maximum likelihood phylogeny modified from Bos et al. [75]. Node support after 1000 bootstrap replications is shown on branches and the tree is rooted by the outgroup *M. canettii*. Large Sequence Polymorphisms (LSPs) described in [68] are indicated along branches. Scale bar indicates the number of nucleotide substitutions per site. **B, C and D.** Dominant MTBC lineages per country. Each dot corresponds to 1 of 80 countries represented in the 875 MTBC strains from the global strain collection analysed by Gagneux et al. [24]. The yellow and an orange dot represent Lineage 7 in Ethiopia [74] and the extinct MTBC strains from Peru, respectively [75]: panel **B** shows the most geographically widespread lineages, panel **C** the intermediately distributed lineages, and panel **D** the most geographically restricted lineages.

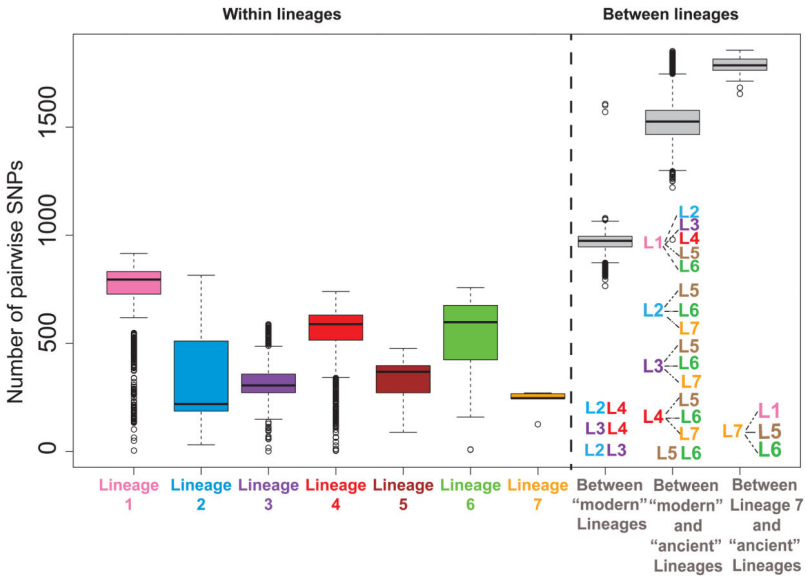

Figure 2.

Number of pairwise difference between MTBC strains. The alignment of 217 human-adapted MTBC clincial strains published previously [76] were used to calculate the number of SNPs between any two strains (i.e. the SNP-distance). We calculated the SNP distance among each pair of strains includeding 44 clinical strains belonging to Lineage 1, 37 strains of Lineage 2, 36 strains of Lineage 3, 64 strains of Lineage 4, 16 strains of Lineage 5, 16 strains of Lineage 6 and 4 strains of Lineage 7. The results are shown in a box-plot generated with R grouping pairwise SNP-distances within each lineage (number of pairwise comparisons were Lineage 1: N=946, Lineage 2: N=666, Lineage 3: N=630, Lineage 4: N=2,016, Lineage 5: N=120, Lineage 6: N=120, Lineage 7: N=6), within “modern” lineages (6,274 pairwise comparisons), between Lineage 7 and Lineages 1, 5, 6 (75 pairwise comparisons), and 12,825 other inter-lineage comparisons.

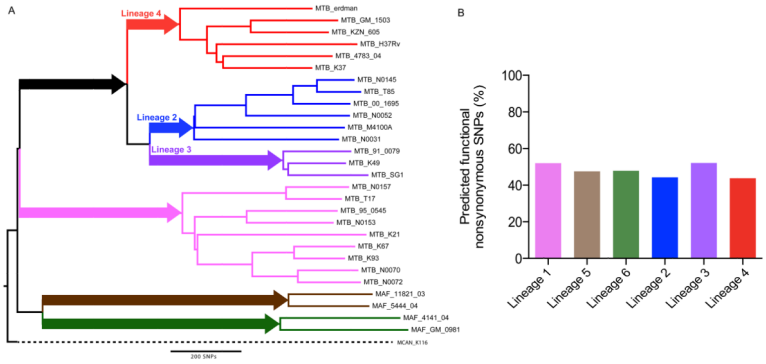

Figure 3.

Predicted functional impact of lineage-specific SNPs. **A.** Neighbour-joining phylogeny based on 28 globally representative MTBC strains, using 13,086 variable positions [227]. The six main lineages are named and branches coloured as reported previously [24,35]. The number of lineage specific SNPs are indicated along the main braches. **B.** Percentage predicted functional nonsynonymous SNPs per lineage based on the prediction algorithm SIFT [78].

Table 1

Databases and bioinformatics tools to explore MTBC genomic diversity

Tool type	Tool name	Input genomic data data	Variation investigated	Reference
Database	Tuberculist	H37Rv complete genome	Protein information, drug and transcriptome data, mutant and operon annotation, bibliography, structural views and comparative genomics	[55]
Database/tools	TBDB	9 complete genomes and 10 WGS	Genomic (SNPs, small indels), expression (RNA-seq and microarrays) and bibliography.	[228]
Database/tools	Patrik	101 complete genomes and 1899 WGS	Comparative genomics of mycobacteria, annotations, access to WGS data, metadata associated with genomes.	[229]
Database/tools	PolyTb	1500 GWS	SNPs, small indels and large deletions.	[230]
Database/tools	GMTV	1084 WGS	Clinical, epidemiological, microbiological and genome variations	[231]
Software	SpolPred	WGS:fastq	Spoligotyping	[56]
Software	SeqSphere	Gene sequences	Sequence types	[232]
Software	KvarQ	WGS:fastq	SNPs: drug resistant and phylogenetic markers	[57]