ANNUAL REVIEWS

Annual Review of Genetics

Genetic Diversity in *Mycobacterium tuberculosis* Clinical Isolates and Resulting Outcomes of Tuberculosis Infection and Disease

Julian S. Peters,[1] Nabila Ismail,[2] Anzaan Dippenaar,[2,3] Shuyi Ma,[4] David R. Sherman,[4] Robin M. Warren,[2] and Bavesh D. Kana[1]

[1]Department of Science and Innovation-National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand and the National Health Laboratory Service, Johannesburg 2000, South Africa; email: petersjulian8@gmail.com, bavesh.kana@wits.ac.za

[2]Department of Science and Innovation-National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg 7505, South Africa; email: nabilai@sun.ac.za, rw1@sun.ac.za

[3]Family Medicine and Population Health (FAMPOP), Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, 2000, Belgium; email: Anzaan.Dippenaar@uantwerpen.be

[4]Department of Microbiology, University of Washington School of Medicine, Seattle, Washington 98109, USA; email: shuyima1@uw.edu, dsherman@uw.edu

Annu. Rev. Genet. 2020. 54:511–37

First published as a Review in Advance on September 14, 2020

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-022820-085940

Keywords

Mycobacterium tuberculosis, lineages, principal genetic groups, epigenetics, drug resistance, disease outcome

Abstract

Tuberculosis claims more human lives than any other bacterial infectious disease and represents a clear and present danger to global health as new tools for vaccination, treatment, and interruption of transmission have been slow to emerge. Additionally, tuberculosis presents with notable clinical heterogeneity, which complicates diagnosis, treatment, and the establishment of nonrelapsing cure. How this heterogeneity is driven by the diversity of

clinical isolates of the causative agent, *Mycobacterium tuberculosis*, has recently garnered attention. Herein, we review advances in the understanding of how naturally occurring variation in clinical isolates affects transmissibility, pathogenesis, immune modulation, and drug resistance. We also summarize how specific changes in transcriptional responses can modulate infection or disease outcome, together with strain-specific effects on gene essentiality. Further understanding of how this diversity of *M. tuberculosis* isolates affects disease and treatment outcomes will enable the development of more effective therapeutic options and vaccines for this dreaded disease.

INTRODUCTION

Tuberculosis (TB) is an infectious respiratory disease caused by *Mycobacterium tuberculosis*, which spreads through infectious droplets that are expectorated by a diseased individual during the process of coughing, sneezing, singing, or even tidal breathing. It primarily affects the lungs and can lead to severe tissue damage resulting in chronic coughing, night sweats, weight loss, and malaise. In addition to the lungs, tubercle bacteria can also spread to the brain (causing TB meningitis), liver, spleen, and bones. These symptoms occur in a relatively small number of infected individuals, roughly 10%, whilst the majority of individuals carry tubercle bacteria with no symptoms but retain the potential to develop symptomatic disease. Regrettably, humanity has battled TB since before recorded history, and in the last two centuries, this disease is estimated to have caused more than a billion deaths, twice the number caused by malaria, small pox, the plague, cholera, influenza, and acquired immunodeficiency syndrome (AIDS) combined (57). This growing public health threat was met with a declaration of a state of emergency in 1993 by the World Health Organization, and the resulting programmatic efforts have saved millions of lives (45, 134). This has, however, been insufficient to turn the tide of the pandemic, as numerous socioeconomic drivers, concomitant human immunodeficiency virus (HIV) infection, and the emergence of drug resistance have thwarted elimination efforts globally. As a result, there is growing comprehension that focusing only on treatment or treatment outcomes will not eliminate TB. Rather, attention needs to be given to other aspects that drive high TB incidence, including poverty, overcrowding, malnutrition, continued transmission, and weak health systems (84). Recent efforts directed at identifying host immune biomarkers of protection, disease progression, treatment response, and disease recurrence suggest that the study of human genetic determinants of TB infection outcome holds promise for new vaccines and therapeutic options (see the sidebar titled TB Infection Resistance). This, combined with the improved protective efficacy reported in the recent M72/AS01E and Bacille Calmette-Guérin (BCG) revaccination clinical trials (85, 116, 124), has bolstered enthusiasm for the development of a new TB vaccine. In contrast, the identification of bacterial biomarkers that drive TB infection and disease outcomes is largely lacking.

TB INFECTION RESISTANCE

The existence of individuals who are exposed to *Mycobacterium tuberculosis* but never develop an infection (termed the resistor phenotype) has been well documented, in addition to those individuals who get infected and do not progress to full-blown TB disease (**Figure 3**). These phenomena confirm that, in addition to bacterial genetic determinants, host genetics most likely play a central role in modulating infection resistance and disease outcome. These genetic factors are now the focus of intense study, as they hold the key to identifying correlates of protection that can be used to develop new vaccines or host-adjunctive therapies to accelerate bacterial clearance and reduce lung damage (127).

M. tuberculosis is an intracellular pathogen, and a multiplicity of factors relating to its fundamental metabolism and physiology can be associated with ultimate transmission potential, virulence, and pathogenesis. Studies that have described these aspects of microbial physiology have used laboratory strains of the tubercle bacillus characterized by limited genetic variability. However, the introduction of various molecular typing and whole-genome sequencing (WGS) techniques has revealed that the global TB epidemic is driven by a heterogeneous collection of *M. tuberculosis* strains that have phylogeographically evolved with certain populations. How this diversity in strain type contributes to the heterogeneous presentation of infection and disease in the clinical setting, or whether any potential bacterial biomarkers exist within these distinct isolates, remains relatively unexplored. In this review of selected literature, we describe the genetic heterogeneity in *M. tuberculosis* strains; explore how particular clinical isolates are able to modulate immune responses, pathogenesis, and disease outcome; and describe how host metabolism is modulated by epigenetic mechanisms in the tubercle bacillus.

WGS: whole-genome sequencing

Phylogeographically: describes geographic distribution of genetic traits or distinguishing markers in individuals and particular bacterial strains

***Mycobacterium tuberculosis* complex (MTBC):** genetically related bacterial species comprising numerous pathogens that cause TB disease in humans and animals

Lineages: comprise a classification system that further delineates principal genetic groups into eight subgroups based on differences in long sequence polymorphisms

Principal genetic groups (PGGs): a classification system for mycobacterial strains that is based on single nucleotide polymorphisms in distinct structural genes in mycobacteria

Single nucleotide polymorphism (SNP): the mutation of a single nucleotide in DNA

Large sequence polymorphism (LSP): large changes in genetic material, including insertion and deletion of genomic regions

PHYLOGENY OF *MYCOBACTERIUM TUBERCULOSIS*

The genomes of the members of the *M. tuberculosis* complex (MTBC), including *M. tuberculosis*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium pinnipedii*, *Mycobacterium microti*, and *Mycobacterium africanum*, were previously thought to be highly conserved, suggesting that the occurrence of any variation would be of no clinical significance (31). However, guinea pig infection studies as early as 1960 demonstrated that strains obtained from India were less virulent than strains obtained from the United Kingdom (69). These early studies highlighted strain-specific effects on transmissibility, immunogenicity, and virulence, thus providing evidence that the genetic variation of the MTBC was of clinical relevance. The progenitor of the human-adapted lineages are hypothesized to have evolved in Africa approximately 70,000 years ago and subsequently moved out of Africa due to pastoral migration (30, 41). Over time, the progenitor evolved into at least eight different lineages described to date (**Figure 1**), possibly due to geographical isolation in different human populations. Following the European discovery of the Americas, members of these lineages became dispersed to reflect the current global population structure of the MTBC (13).

Phylogenetic studies have made use of unambiguous, unique, and irreversible genetic markers to classify *M. tuberculosis* strains into principal genetic groups (PGGs). Five major global phylogenetic classification studies have grouped genetically related classes of mycobacterial strains (**Figure 1**). One of the earliest phylogenetic studies used single nucleotide polymorphism (SNP) analysis of structural genes to group *M. tuberculosis* into three principal genetic groups using synonymous and nonsynonymous SNPs (111). Analysis of long sequence polymorphs (LSPs) was used to classify 875 strains into four lineages that were initially defined by Baker et al. (8) using SNP analysis of drug resistance genes. Furthermore, two additional lineages were observed, which are traditionally referred to as *M. africanum* (42).

Using an in silico approach, Gutacker et al. (52) screened 5,069 isolates for 36 SNPs, yielding a classification of 9 genetic clusters, all of which were congruent with 2 previous classifications (**Figure 1**) (8, 111). A PCR-based reverse hybridization technique that assesses the diversity of the direct repeat loci was evaluated using spoligotyping, revealing six major lineages that were congruent with previously described groupings. These studies are well summarized by Gagneux & Small (43), and the general theme emerging is that the phylogenetic classification of isolates is generally conserved regardless of the method used. The phylogenetic framework laid by these early studies paved the way for WGS technologies that have enabled comprehensive evaluation

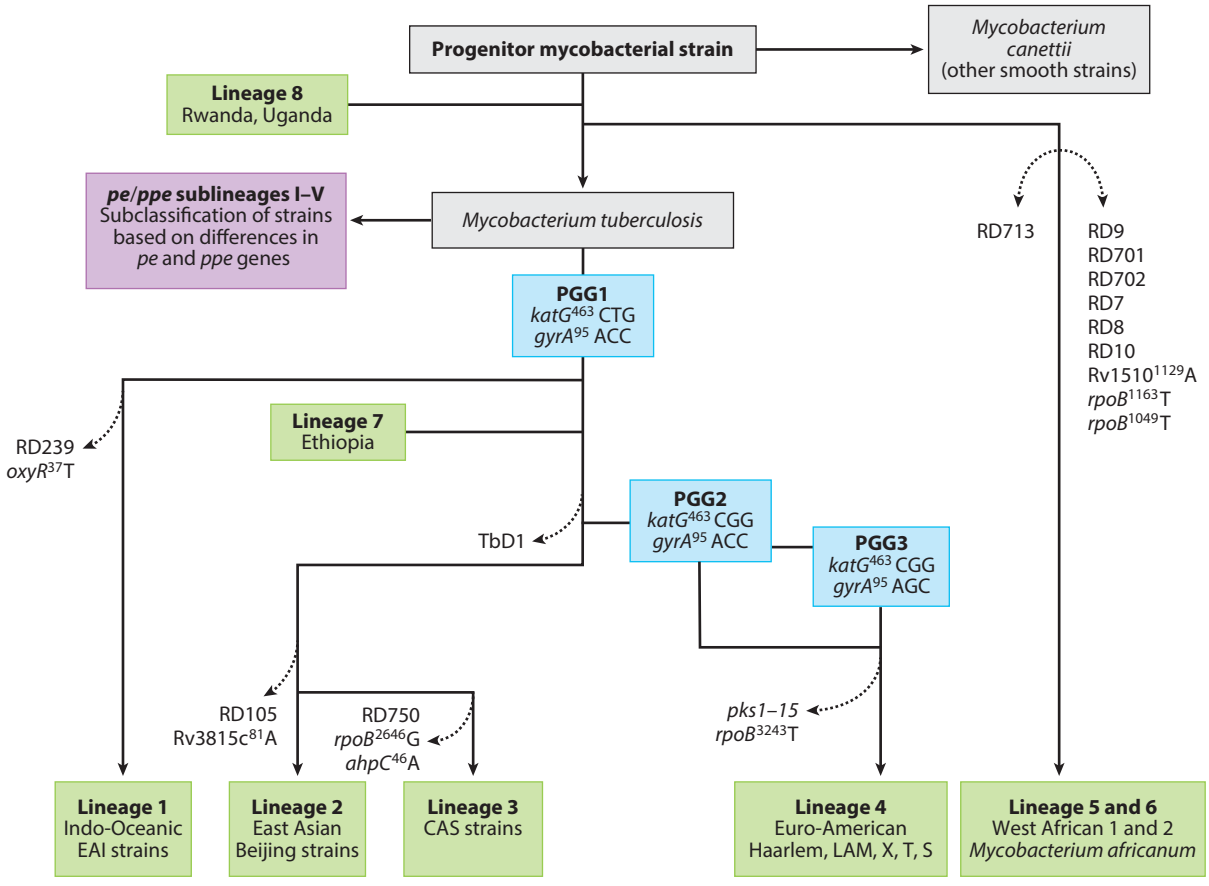


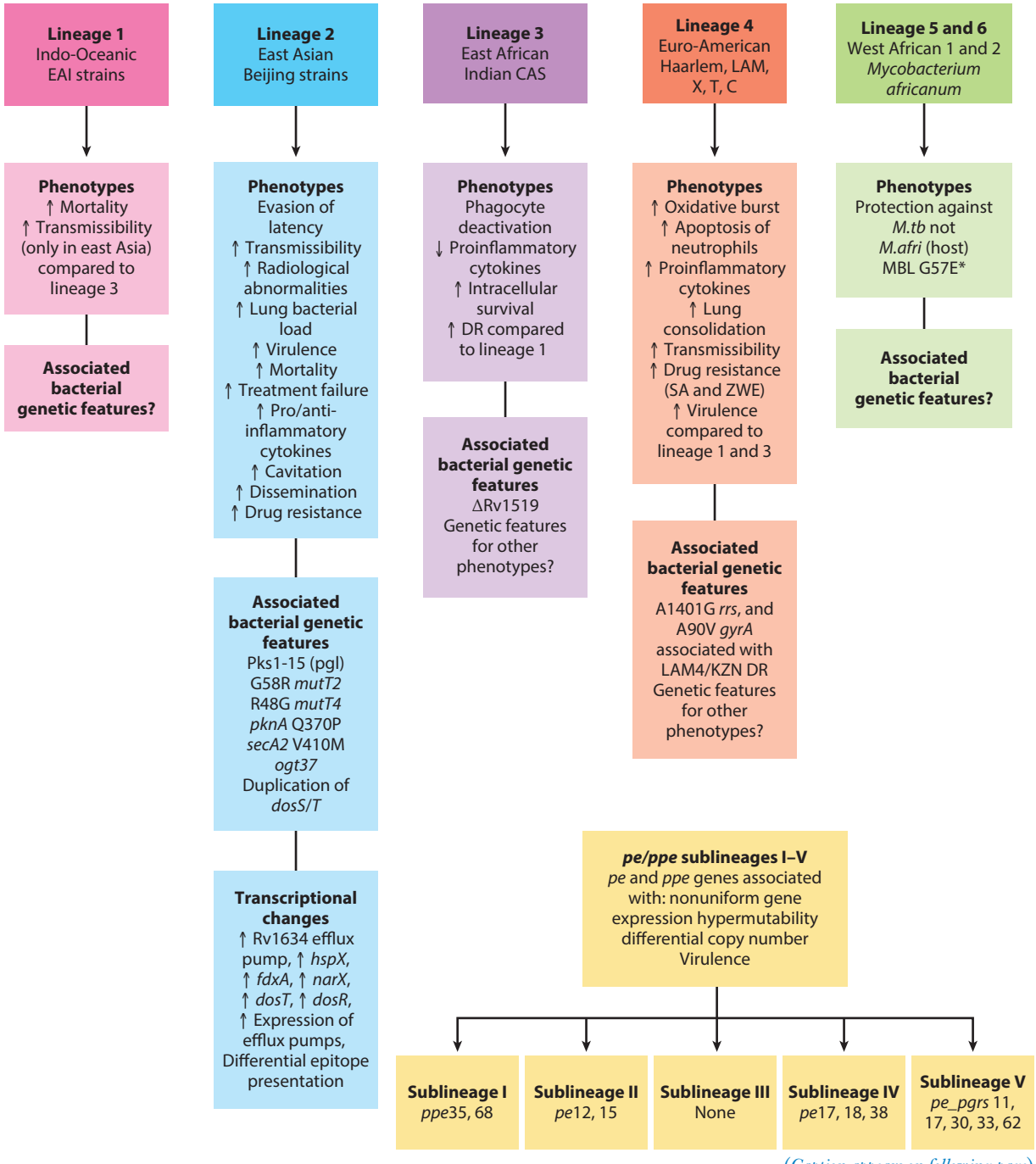
Figure 1

Evolution and phylogeny of *Mycobacterium tuberculosis* clinical isolates. Current circulating strains of *M. tuberculosis* comprise a heterogeneous distribution of strains with distinct genetic differences that allow for the classification of most strains into three principal genetic groups (PGGs; *blue boxes*) and lineages 1–8 (*green boxes*). This classification is based on regions of difference (RDs) or specific changes in certain loci, both shown as dashed lines. Shown also is the common ancestor of modern-day tubercle bacteria and *Mycobacterium canettii*, which has a smooth colony morphology compared to the characteristic surface cording seen on *M. tuberculosis* colonies. Lineage 4 has members of both PGG2 and PGG3. Strains can also be classified based into five sublineages based on copy number or the expression of their *pe/ppe* genes (*purple box*).

of strain diversity, capturing discreet changes at the single-nucleotide level in the genome. The publication of the whole-genome sequence of the laboratory strain *M. tuberculosis* H37Rv has become the blueprint for the resequencing of clinical isolates (28).

Human disease is primarily caused by lineages 1–8, each lineage being composed of multiple sublineages (4, 86). Among the human-adapted MTBC lineages, some occur globally and others are geographically restricted, suggesting generalist and specialist phenotypes (31, 112). Lineages 5 and 6 are geographically limited to West Africa (*M. africanum*) (135). Lineage 4 is primarily found in Latin America and Europe, whereas lineage 2 is the predominant genotype in East Asia (17, 109). Lineage 2 (Beijing strain family) has been intensely studied, given its importance in driving the TB epidemic in Asia, Russia, and South Africa. It has been strongly linked to increased virulence, increased transmissibility, and drug resistance (53, 102). This lineage is thought to have

evolved 6,600 years ago and can be classified into 7 different clonal complexes (sublineages) (53, 82, 112). How this genetic diversity, and that seen in other lineages, impacts disease outcome is an important area of research, and key studies are summarized below, and in **Figure 2**, in the context of TB infection and pathogenesis.



(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

Phenotypic and genotypic traits of *Mycobacterium tuberculosis* lineages. Where available, phenotypic traits for clinical isolates are shown, followed by associated possible genetic determinants. Lineages 5 and 6 (*green*) represent lineages where differential phenotypic traits or bacterial determinants of said traits have not been intensively studied; however, host determinants have been investigated and are noted (*asterisk*). Lineages 1, 3, and 4 (*pink*, *purple*, and *orange*, respectively) represent those strains where notable phenotypic differences have been described but the genetic basis of these in tubercle bacteria have not been studied in detail. The most well-studied is lineage 2 (*blue*), with multiple clinical observations compared to other strains and multiple links to bacterial genetic determinants. Within the boxes, arrows pointing up denote an increase in a particular trait or gene expression. Examples of *pe/ppe* lineages, and proteins within those lineages, that have a bearing on infectiousness, pathogenicity, virulence, and possible course of infection are shown in yellow. Abbreviations: CAS, Central Asian strain; DR, drug resistance; EAI, East African Indian; KZN, KwaZulu Natal; LAM, Latin American–Mediterranean; *M.afri*, *Mycobacterium africanum*; *M.tb*, *Mycobacterium tuberculosis*; SA, South Africa; ZWE, Zimbabwe.

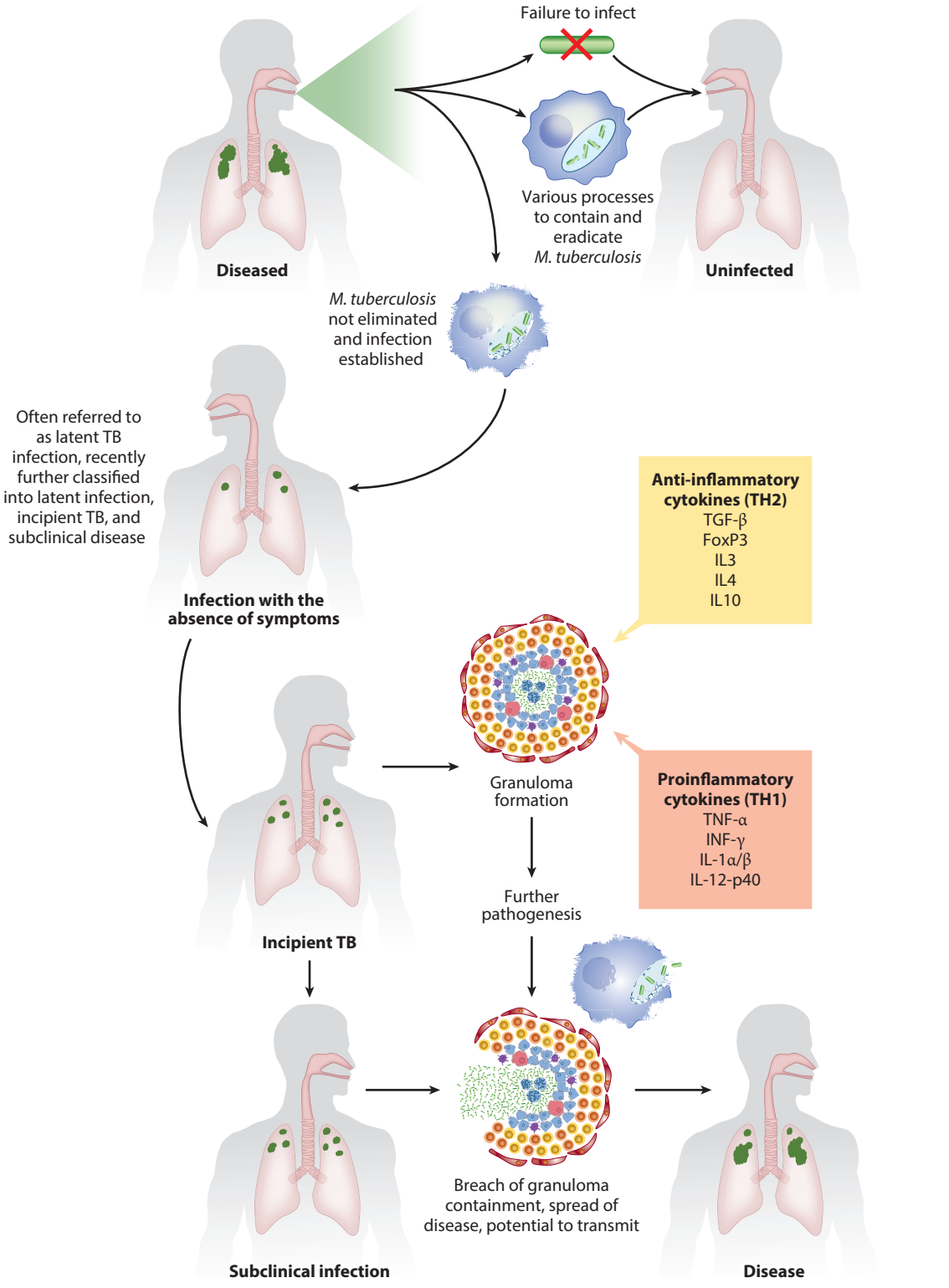
TB INFECTION AND PATHOGENESIS

TB infection occurs through the sharing of air containing aerosolized *M. tuberculosis* between diseased and uninfected susceptible hosts, leading to a number of scenarios. Infection can be contained in an asymptomatic state or followed by incipient TB characterized by an evolving immune biomarker signature that is predictive of progression from infection to disease (37). If further immunological containment is lost, radiologically and/or microbiologically detectable subclinical disease occurs, and progression from this state ultimately results in the establishment of full-blown TB disease. Subsequent chemotherapy can lead to nonrelapsing cure, cure followed by disease recurrence, or treatment failure associated with drug resistance (**Figure 3**). Bacterial biomarkers or genetic determinants that allow for distinguishing these clinical phenomena are poorly described. However, there is evidence that points to salient features of the tubercle bacillus that enable successful transmission and colonization of the human host.

As an example, *M. africanum*, also a causative agent of a substantive amount of TB in West Africa, has equivalent transmissibility as *M. tuberculosis*; however, progression to active disease was significantly lower for the former (32). In contrast, severity of disease was significantly worse for *M. africanum*-associated disease, suggestive of the presence of host genetic drivers that modulate disease outcome. In this regard, it was shown in a Ghanaian population that the G57E variant of the mannose-binding lectin confers a protective role only to disease caused by *M. africanum* and not by *M. tuberculosis* (118). In a separate study, X-rays showed that individuals infected with Euro-American lineages (LAM) presented with more consolidations when compared to other lineages, while individuals with Beijing strains had significantly more cavitation. Furthermore, disease caused by the Beijing strains prior to treatment initiation was associated with a shorter duration to presentation, meningitis, and drug resistance at the commencement of therapy (117). Other distinctive strain-specific effects are summarized in **Table 1** and discussed in further detail hereafter.

GENETIC EVOLUTION OF PATHOGENICITY

M. tuberculosis lacks the classic virulence factors that other organisms encode to enable the invasion of host tissues, such as pili, flagella, fimbriae, and classic toxins (18). In order to understand the genetic determinants of *M. tuberculosis* infection outcomes, it is important to map out the evolutionary history of the genome. This allows for the identification of genes involved in the emergence of virulence and those genetic elements that contribute to pathogenicity, wherein variation is likely to affect the outcome of infection. The emergence of pathogenic mycobacteria has been described as a biphasic evolutionary process involving the acquisition of genetic material, gene duplication, and, finally, reductive evolution as the organism settled into a restricted niche (100, 129). However, careful scrutiny of the MTBC suggests that no recent horizontal gene transfer has



(*Caption appears on following page*)

Figure 3 (*Figure appears on preceding page*)

Transmission and pathogenesis of *Mycobacterium tuberculosis*. Shown is a spectrum of infected and diseased states that may possibly prevail upon exposure of a susceptible host to tubercle bacteria. During the cycle of pathogenesis and granulomatous disease, a multiplicity of cytokines are required. Those shown in boxes have been demonstrated to be differentially modulated by clinical isolates, thereby influencing the outcome of disease. Abbreviation: TB, tuberculosis.

occurred, hence genomic islands that are exclusive to the MTBC most likely represent ancient acquisitions from α-, β-, and γ-proteobacteria (100).

The MTBC evolved as a single clonal group from a pool of recombinogenic mycobacterial ancestors that resembled *Mycobacterium canettii*, the closest neighbor to the MTBC (88, 114). *M. canettii* is an opportunistic species exhibiting complete genomic isolation from the MTBC but sharing 98% sequence identity. This allows for investigation of how virulence emerged and identification of genes that are relevant for the transition from an emerging opportunistic pathogen to an obligate intracellular pathogen. *M. canettii* harbors up to 366 genes that are absent from the MTBC and lacks 51 genes, including whole operons, that are present in the MTBC. Missing genes in *M. canettii* mostly include those involved in lipid metabolism, protein transport, and proteins of unknown function. In addition, differences in genes related to lipooligosaccharide metabolism could explain the formation of smooth colonies in *M. canettii* as opposed to the cording phenotype observed for colonies produced by members of the MTBC (12, 115).

Analysis of the regions of difference (RDs) (**Figure 1**) used to classify *M. tuberculosis* into different lineages possibly provides clues to the basis of phenotypic attenuation of some strains. For example, loss of RD1 in *M. bovis* BCG strains results in severe attenuation, hence its usefulness as a vaccine strain. RD1 encodes the ESX-1 secretion system, which is used to secrete proteins such as CFP10, ESAT-6, and other virulence determinants into the cytosol of macrophages, enabling the escape of the pathogen from the phagosome (56, 72). The demonstrated essentiality of RD for the virulence of *M. tuberculosis* represents an important example of evolutionary loss of genomic regions affecting virulence.

SELECT GENETIC DETERMINANTS THAT MODULATE TRANSMISSION AND DISEASE OUTCOME

Distinctive genomic features of *M. tuberculosis* (or MTBC), which are absent in other organisms, are likely mediators of the pathogenic process, including members of the *pe/ppe* and *pe_pgrs* gene families. While the function of most of these proteins remains unknown, emerging evidence suggests that they are required for epitope variation to enable host evasion (40, 105). For a comprehensive summary of the *pe/ppe* genes, the reader is referred to previous literature (14, 81, 105). Of particular note, McEvoy et al. (81) demonstrated that nonsynonymous SNPs occur more frequently in *pe/ppe* genomic regions when compared to other regions, suggesting plasticity within the genome that allows for the ability to quickly adapt. In-depth analysis of whole-genome sequences of members of the MTBC has highlighted that T cell epitopes are highly conserved, suggesting evolutionary pressure on these genomic regions (29). Similarly, genomic regions encoding components in cell wall biosynthesis, transcriptional regulation, and DNA repair pathways were shown to be under convergent positive selection in *M. tuberculosis* (26, 87). With regard to the ability of tubercle bacilli to adopt a dormant state, duplication of the dormancy regulon (*dos*) in lineage 2 and lineage 4 isolates of the MTBC has been reported, although the biological advantage of this duplication and its role in disease phenotype are not fully understood (36, 133). **Figure 2** contains a summary of these and other lineage-specific genetic differences that have been associated with outcomes. Collectively, these observations suggest a host-induced unidirectional evolutionary process, which allows the pathogen to adapt to changing microenvironments encountered during pathogenesis.

Region of difference (RD): a large section of DNA that has been lost or changed through the evolution of mycobacterial strains

***pe/ppe* genes:** a group of genes that encode proteins containing the proline-glutamine (PE) motif at the N terminus

Table 1 Genes/molecules contributing to strain-specific differences in tuberculosis infection and disease

Gene/protein/lipid	Strains compared	Possible effect on tuberculosis infection or disease	Reference(s)
Pgl, Rv2952	CDC1551 and HN878	Decrease in inflammatory response; differences in virulence, bacterial load, TNF-α production, necrosis, and rate of bacterial clearance in CSF	59, 122
α-crystallin, *hsp55*, PstP1, 47KDa antigen	Beijing, F23, and H37Rv	Increased expression of α-crystallin and decreased expression of Hsp65, PstS1, and the 47-kDa protein in Beijing compared to the other strains	92
Rv1160 (*mutT2*), Rv3908 (*mutT4*), *ogt37*	LAM, Beijing, *Mycobacterium bovis*	*M. bovis*–pyrazinamide resistance, LAM10-isoniazid resistance, Beijing-multidrug resistance	16
accE5, *kdpD*, UT205, ↑ESX-1, ↑*mbt*, ↑*moa*	Lineage 4 LAM strains UT127 and UT205	Increased activation of virulence systems such as the ESX-1, polyacyltrehalose, and sulfolipids in UT205; increased expression of genes involved in DNA replication, cell division, and lipid biosynthesis in UT127	7
Rv0178 D150E, Rv0759c-Rv0760c 2-bp deletions affecting *phoP* regulation	Lineage 2 Beijing strains H54 and H112	Strain H112 exhibits significantly better intracellular survivability and lower levels of TNF-α	95
sigG, *pks5*, *pks7*, *mce-Ib*, Fad26, *virS*	Lineage 2 and lineage 4	Nonsynonymous mutations in lineage 2 strains (*sigG*, *pks5*, *pks7*) and lineage 4 strains (*mce-Ib*, Fad26, *virS*) affect virulence and transmission	68
dosR, *dosT* (Rv3130c, Rv3133c), *hspX*, Rv2031c, *fdxA*, Rv2007c, *narX*	Beijing and H37Rv	Increased DosR regulon in Beijing; differential acetylation of DosR; accumulation of triacylglycerides in aerobic culture	99
S219L (PhoP), A219E (MazG), I228M (EspK)	H37Rv and H37Ra	Attenuation of H37Ra	62
Deletion of pe35, *ppe68*, *esxB*, *esxA*	*M. bovis*; *M. bovis* BCG; *Mycobacterium microti*; *Mycobacterium africanum*; *M. tuberculosis* H37Rv; H37Ra, and six *M. tuberculosis* clinical isolates	Attenuation of animal isolates in human hosts	62
glgP, *linB*	*M. tuberculosis* K strain versus H37Rv and *M. bovis*	Differences in macrophage activation	103
eccD3a 76 S→N, *eccD3* 95 A→T, *mmpL10* 408 T→A, *plcA* 446 T→A, *mbtB* 674 V→L, *ppsA* 1194 liters→R, *mas* 2005 T→P, Rv2952 176 G→R, *kefB* 102 T→A, *lipF* 233 R→C, *papA2* 466 P→L, *fadD23* 422 E→Q, *espK* 44 D→N, *espK* 660 E→A, *eccC2* 650 D→G	Lineage 2 versus lineage 3, 4, and 6 and animal strains	Nonsynonymous SNPs in lineage 2 strains, absent from lineage 3, 4, 6, and 8 that are located in known virulence genes that have been validated across other lineage 2 strains as having an effect on the invasion of host cells, evasion of immune responses, and bacterial proliferation	62
T cell cluster 9 epitopes (0579, 0591, 1737, 1829)	Lineage 2 versus lineage 3, 4, and 6 and animal strains	Loss of antigens in the lineage 2 strains resulting in the evasion of the immune response	62

Abbreviations: BCG, Bacille Calmette-Guérin; CSF, cerebrospinal fluid; LAM, Euro-American lineages; pgl, phenolic glycolipid; SNP, single nucleotide polymorphism; TNF-α, tumor necrosis factor-α.

INSERTION ELEMENTS AND GENOMIC VARIATION

Insertion sequence *6110* (IS*6110*): a naturally occurring transposable element in mycobacteria where the number and location of insertion in mycobacterial genomes have been valuable in further classifying mycobacterial strains

Analysis of insertion sites for insertion sequence *6110* (IS*6110*), a natural transposon in *M. tuberculosis*, has also been used to identify essential genes in a total of 161 clinical isolates (140). A genome-wide survey characterizing the precise base-specific insertion sites of this naturally occurring transposable element demonstrated that 180 insertions were intragenic, affecting 100 open reading frames. The number of genes carrying a disruptive insertion of IS*6110* in clinical strains is much lower when compared to laboratory strains, thus suggesting that relatively few genes (<300) can readily accept an IS*6110* insertion in the clinical setting. Hence, more genes appear to be essential in vivo in clinical strains compared to laboratory strains, consistent with the notion that repeated human-human passage likely exerts substantial selective pressure to retain genes required for fitness, transmission, and persistence (140). Similarly, Gonzalo-Asensio et al. (51) assessed the dynamic distribution of IS*6110* between MTBC lineages by genetic and IS*6110* messenger RNA (mRNA) analysis of 2,236 clinical isolates. This study reported that modern *M. tuberculosis* lineages (Beijing, CAS, LAM, and T) are more widespread globally and associated with high transmissibility, drug resistance, and virulence. These strains carry high copy numbers of the IS*6110* element in comparison to ancient strains (*M. africanum*, East African Indian (EAI), X-strains, and *M. bovis*).

The potential of IS*6110* elements to drive virulence is further demonstrated through an outbreak of virulent *M. bovis*, which can infect but is normally unable to transmit in immune-competent human hosts. This strain carried an IS*6110* element upstream of the *phoPR* operon acting as an exogenous promoter and increasing virulence (50). Beijing strains have the highest copy number of the IS*6110* transposon element inserted in their genomes, possibly contributing to aspects such as host immune evasion. Consistent with this notion, these strains carry IS*6110* insertions in *ppe38–71*, resulting in the lack of secretion of several PE_PGRS and PPE-MPTR proteins (51). Notably, 80 PE_PGRS and PPE-MPTR proteins that are secreted by the ESX-V secretion system are affected by this deletion, including PE_PGRS30, PE_PGRS33, and PE_PGRS47 that have all been experimentally shown to be involved in virulence in animal models (3, 5, 6, 60, 104).

Gene expression studies of the IS*6110* element revealed that a transcriptionally active IS*6110* derivative (pIS*6110*-FS) exhibited 20-times-higher transposition rates than the wild-type IS*6110*, which was more pronounced during stationary phase and starvation (51). Furthermore, comparison of mRNA quantities of IS*6110* between MTBC members showed that *M. tuberculosis* strains exhibit higher IS*6110* expression than *M. bovis* strains (51). The expression of IS*6110* was upregulated in murine macrophage models, indicating a potential role of IS*6110* in adaptation to the host (51).

GENETIC BASIS FOR THE SUCCESS OF LINEAGE 2/BEIJING STRAINS

Since their identification by van Soolingen et al. (126), Beijing strains have been widely associated with a number of phenotypes epidemiologically, clinically, and experimentally, including high frequency of transmission, poor treatment outcome, severe host immune modulation, higher propensity to develop drug resistance, higher bacterial load, severe histopathology, and extrapulmonary presentations (54). Epidemiologically, Beijing strains are the most widespread strains globally, indicating that this lineage may have evolved unique properties allowing effective spread. Beijing strains demonstrate high adaptability to different human populations; for example, Beijing strains that are found in Peru are unique to Peru, suggesting that they emerged as a distinct endemic clone to that host population (61). Beijing strains are identified by several genetic markers, including the deletion of spoligotype spacers 1–34 and the IS*6110* insertion between the *dnaA* and the *dnaN* genes (125). These crude strain identifiers are unlikely to lead to specific phenotypes; however, other unique molecular identifiers of the Beijing lineage exist and are likely

to result in subtle or major phenotypic variations. These include the deletion of several RDs (e.g., 207, 105, 181, 150, and 142), deletion of individual genes (e.g., Rv0729c and Rv0927c), and mutation of genes (e.g., Rv2629, Rv1160, and Rv3908) (54). However, the effects of these genetic changes in TB infection and disease outcome have not been firmly established.

TH1 immune response: a response to infection that is characterized by the production of interferon-γ aimed at inducing the activation of macrophages to kill intracellular pathogens

Evidence of hypervirulence of the Beijing lineage has previously been assessed in vitro and in animal models by assessing cytokine production, mortality, rate of proliferation, in vivo growth using competition assays, extent of lung damage, and dissemination (10, 78). While hypervirulence has been associated with the Beijing lineage in general, it is not ubiquitous to all Beijing strains. Rather, this lineage exhibits a series of phenotypes from hypo- to hypervirulent as evidenced by a number of epidemiological studies on outbreaks of different Beijing isolates (33, 53). Molecular mechanisms that underpin the success of the Beijing strains can be divided into genetically encoded discrepancies, epigenetic modifications, and natural variations in protein expression. Furthermore, differential immune modulation by Beijing strains is possibly associated with disease progression and infection outcome.

The TH1 immune response is predominantly driven by the proinflammatory cytokines tumor necrosis factor-α (TNF-α) and interferon-γ (IFN-γ), which restrict MTBC growth using reactive oxygen/nitrogen intermediates and autophagy. In contrast, TH2 cytokines such as transforming growth factor-β (TGF-β) result in an anti-inflammatory immune response as they are associated with a decrease in activated CD4 cells and failure to restrict *M. tuberculosis* growth. The balance between pro- and anti-inflammatory cytokines results in granuloma formation (**Figure 3**). Some lineage 2 strains have been associated with a dampened proinflammatory immune response driven by a phenolic glycolipid, resulting in extensive dissemination of the pathogen. Furthermore, differences in monocyte activation have been identified as a mechanism of differential virulence, illustrated by the fact that the CDC1551 outbreak strain induced higher levels of proinflammatory TH1 (IL-1α/β) cytokines. In contrast, analysis of another outbreak Beijing strain HN878 revealed induction of higher levels of IL3 and IL4 that are characteristic of TH2-type immunity (98). This phenotype was subsequently genetically linked to the *pks1–15* loci (98). Phenolic glycolipids, encoded by the *pks* gene cluster, form part of the surface lipid repertoire, and studies have shown that these specific lipids are used by *M. tuberculosis* to engage the host by enhancing the recruitment of growth-permissive macrophages. However, further studies have shown that this surface lipid is not present in all hypervirulent Beijing strains (99). For example, a comparison of an outbreak Beijing strain to a nonoutbreak Beijing strain in California identified a 7-bp frameshift mutation on the *pks* gene in the outbreak strain, while the nonoutbreak strain had an intact *pks* gene cluster that is typical of other virulent Beijing strains (19).

The strain-specific repertoire of induced cytokines likely helps to determine whether or not a strain will be successful, but currently no specific cytokines have been consistently associated with particular strains and outbreaks. Furthermore, there are some inconsistencies in the immunological profiles, even with subspecies of the same strain. In this regard, modern and ancient lineages of the Beijing genotype differ distinctly in their induction of proinflammatory cytokines, with ancient lineages inducing significantly higher levels of IL-1b, IL-6, IL-8, IL-10, GM-CSF, TNF-α, GRO-a, and RANTES signatures (22). These observations may be the result of the differences in the RDs in these two groups, where ancient sublineages have intact RD181, RD150, and RD142.

In an attempt to link specific genes in Beijing strains to differences in outcome of infection within the lineage, comparison of previously published sequences of 1,082 strains from the ancient and modern lineages were analyzed post hoc (76). Nineteen nonsynonymous SNPs were reported in conservative codons with significant enrichment for the category of regulatory networks (*pknA*, Rv0452, Rv0890c, Rv2488c, and Rv3173c). Three of the SNPs were located in promoter regions and were verified to alter gene expression (Rv0603, Rv3713, and Rv0308). Furthermore, four of

the genes had frameshift mutations as a result of deletions or premature stop codons (Rv1730c, Rv2147c, Rv2148c, and Rv2180c). Such differences between members of the same lineage may provide clues to explain the molecular mechanisms of success for particular clinical isolates, but further work is required.

Sympatric: the close evolution of two species within the same geographic region

Allopatric: the evolution of species in separate (or separated by) geographic regions

Mycolic acid (MA): a long-chain fatty acid that is an integral component of the mycobacterial cell envelope and limits the diffusion of drugs and small molecules

STRAIN-SPECIFIC ASSOCIATIONS WITH HIV COINFECTION

TB and HIV occur in lethal synergy, particularly in Southern Africa, due to a number of reasons including socioeconomic and molecular aspects of the two infections that allow them to drive each other, as described in the literature (39, 121). Very few studies have sought to determine whether specific *M. tuberculosis* genotypes are associated with HIV infection. Rather, associations between HIV status and lineages are mostly secondary observations of larger studies, the results of which are dependent on the design of the parent study. An analysis of an HIV-positive cohort in a Vietnamese population revealed that 49% of the isolates were Beijing strains and 35% were Indo-Oceanic strains (80). However, this observation more likely represents the phylogenetic structure of the bacterial population rather than an inclination of Beijing strains to infect HIV-positive individuals. In a pediatric cohort in Kampala, there were no significant associations between lineages and HIV status. It should be noted that a single strain (lineage 4_U; Uganda strain) is predominant in the study population, possibly resulting in a bias (131). In contrast, a recent study by Konstantynovska et al. (66) reported that a Beijing strain was significantly associated with HIV status in Ukraine, despite the fact that Beijing strains represent the minority of sampled strains in the study. Another study in Mozambique also demonstrated a significant association between HIV serostatus and Beijing strains when compared to non-Beijing strains in a multivariable analysis adjusted for age, sex, and province (130). The correlation of specific *M. tuberculosis* strains with HIV status may be associated with a particular outbreak rather than the whole lineage. The lineage 4 LAM4/KZN outbreak DR strain in Durban, South Africa, was significantly associated with HIV coinfection (44). In contrast, in a study in Panama, the most prevalent strain was found to be the LAM9, with 78% of these strains being drug resistant and 75% of the sample population being HIV negative (70).

Fenner et al. (39) revealed that HIV coinfection with TB disrupts the normal sympatric host-pathogen evolutionary coexistence that tubercle bacteria have with humans, leading to an allopatric relationship between the pathogen and host. In this case, a sympatric infection was defined as a combination of a particular strain lineage and its corresponding natural host population, such as a Euro-American lineage infecting a Euro-American host. Thus, an allopatric infection would be exemplified by an East Asian lineage infecting a Euro-American host. Using a second validation set of 1,642 isolates belonging to lineages 1–6 obtained in Switzerland from 1991 to 2011, it was observed that the proportion of HIV infection was 4.5 times higher in patients with an allopatric strain compared to patients with a sympatric strain. These findings support a model where the stable relationship between the human host and its locally adapted *M. tuberculosis* lineage is disrupted by the differential selective pressure exerted by the evolutionary newcomer in the form of HIV infection (39). Similarly, different strain types of *M. tuberculosis* have been observed to perturb the HIV infection process, as evidenced by CDC1551, which induces greater HIV replication when compared to HN878 in coinfected peripheral blood mononuclear cells (97).

STRAIN-SPECIFIC DIFFERENCES IN THE MYCOBACTERIAL CELL ENVELOPE

Mycolic acids (MAs) are highly hydrophobic wax-like fatty acids that occur in the cell wall of mycobacteria (128). These α-alkyl β-hydroxy fatty acids constitute a major component of the

mycobacterial cell envelope, giving it structure and providing protection against detergents, antibiotics, and dehydration. MAs are heterogeneous in their chemical structure, forming distinct classes that are composed of fatty acids of different lengths and variable side chains. Three major classes of MAs are described in the literature and include the α-, methoxy-, and keto-MAs, which vary in the number of cyclopropane rings as well as their configuration (*cis* or *trans*) (94, 96). The relative representation of each of these classes varies between strains and also seems to directly influence virulence. An in vitro study by Vander Beken et al. (128) investigated synthetic purified single monomers to study the variations in the immune responses of different MAs in a mouse model. The α-MA appeared to be inert, whereas keto-MA with cyclopropane rings in the *cis* orientation elicited mild inflammatory responses, and oxygenated methoxy-MA resulted in strong immune responses (128). This differential perturbation of the immune system by variant MAs provides a mechanism for different strains of *M. tuberculosis* to affect the outcome of infection (128). A subsequent mass spectrometry-based study assessed the relative abundance of 80 MAs across 36 clinical isolates covering 4 phylogenetic lineages (94). Significant variations were observed between the MA content of ancient and modern lineages, with significantly lower amounts of α-MAs among lineage 6 isolates and an inversion of the methoxy:keto mycolates ratio in lineage 1 isolates. Furthermore, the ratio of oxygenated MA and α-MA was found to be significantly higher for lineage 6 compared to modern lineages and lineage 1 strains (**Figure 4**). WGS of these isolates identified relevant SNPs that may be responsible for the differences in the MA content. A total of 97 SNPs in the MA pathway were observed across all four lineages studied, of which 17 were nonsynonymous and observed in a single lineage. Three of these nonsynonymous SNPs were shared by two modern lineages, with lineages 3 and 4 displaying lineage-specific SNPs. Nine SNPs were predicted to have an effect on protein function [Rv0642c (F95L), Rv0644c (R114L), Rv0904c (G158E), Rv1348 (A653T), Rv1484 (V78A), Rv1686c (V20F), Rv1687c (D102N), Rv3800c (A523G), and Rv3082c (L316R)], and they were only found in the ancient lineages (94). The functional consequences of these findings require further study.

Mutagenesis: the process whereby stable mutations are introduced into DNA

DRUG RESISTANCE AND DISEASE PROGRESSION OR OUTCOME

Early studies on the population structure of *M. tuberculosis* revealed that drug resistance was more commonly associated with Beijing strains, suggesting lineage-specific mechanisms of adaptation to chemotherapeutics (77, 119, 120). Mutations in DNA repair genes can result in increased mutagenesis, and consistent with this, analysis of Beijing isolates revealed mutations in *mutT2*, *mutT4*, and *ogt*, suggesting that these may drive mutagenesis. However, subsequent studies reported that the spontaneous mutation rate in the Beijing family is not significantly different from that of non-Beijing strains, suggesting that while certain mutations in DNA metabolism genes are unique to the Beijing strains, they do not directly cause drug resistance (38, 71).

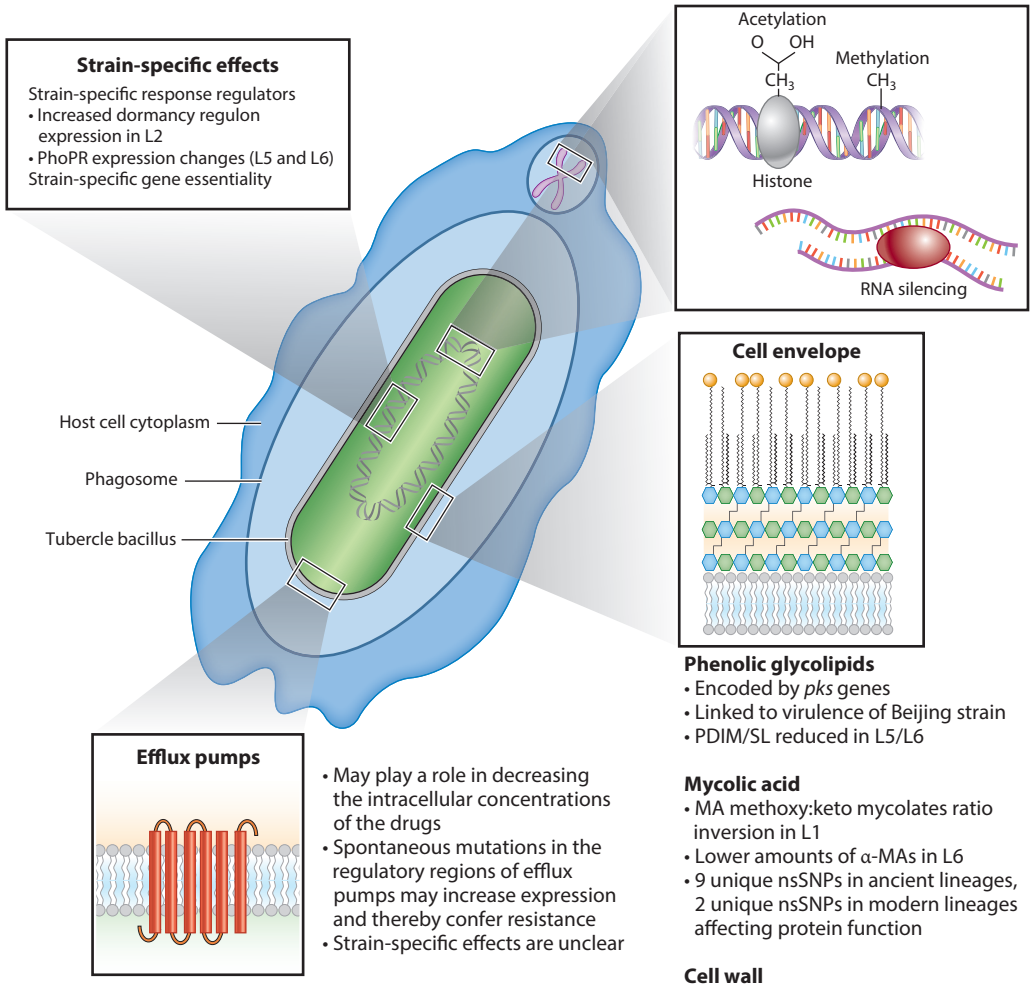
Using H37Rv as a reference genome, a recent study identified 12,802 previously undescribed nucleotide variations in strains from India, of which 38 were located in genes known to be associated with drug resistance (2). Mutations in *rpoB* (S450L) and *katG* (S315T) were dominant in patients presenting with multidrug resistance. A cluster of SNPs associated with ethambutol resistance on the *emb* genes was also identified, among which two were previously undescribed (*embB* G406A, M306I). The study also identified some co-occurring mutations, such as the *rpoC* (A172V) that occurred with the *rpoB* (S450L) mutation in up to 74% of the isolates from northern India. Some co-occurring mutations were found only in specific lineages; for example, the D229G mutation in the Rv2247 (*accD6*) gene occurred with *katG* R463L in all Beijing multidrug-resistant (MDR) and pre-extensively drug-resistant (XDR) isolates. Furthermore, the *gid* E92D SNP was

Epigenetic effects

Glycosylation
- *O*-mannosylation gene Rv0616 overexpression increases virulence
- Differential glycosylation of 67 proteins among different strains on virulence genes

Methylation
- Bacterial genes: *M. tuberculosis* methyltransferases (Rv1988 and Rv2966), *mamA*, *mamB*, and *hsdM*
- Host genes methylated: Histone 2, 3 and 4: Modulating the transcription of genes including NOX1, NOX4, and NOS affecting the inflammatory response
- L1 hypomethylation due to *mamB* S253L mutation resulting in differential expression of 44 genes compared to L4

RNA silencing
- Bacterial miRNA 29, 144, 155 supress expression of host IFN-γ, dampening immune response due to immunosuppression



- May play a role in decreasing the intracellular concentrations of the drugs
- Spontaneous mutations in the regulatory regions of efflux pumps may increase expression and thereby confer resistance
- Strain-specific effects are unclear

Phenolic glycolipids
- Encoded by *pks* genes
- Linked to virulence of Beijing strain
- PDIM/SL reduced in L5/L6

Mycolic acid
- MA methoxy:keto mycolates ratio inversion in L1
- Lower amounts of α-MAs in L6
- 9 unique nsSNPs in ancient lineages, 2 unique nsSNPs in modern lineages affecting protein function

Cell wall
- Thicker cell walls decrease permeability to drugs

(*Caption appears on following page*)

Figure 4 (*Figure appears on preceding page*)

Epigenetic effects and other determinants in *Mycobacterium tuberculosis* that affect pathogenesis. Epigenetic changes include glycosylation, methylation, and RNA silencing. DNA can be methylated and acetylated. Possible changes in the cell envelope and efflux pumps can also affect metabolism and resistance. Abbreviations: α-MAs, α-mycolic acids; IFN-γ, interferon γ; L, lineage; miRNA, microRNA; nsSNPs, nonsynonymous single nucleotide polymorphisms.

observed to occur concurrently with the K43R *rpsL* mutation in East Asian Beijing genotypes of MDR and pre-XDR isolates (2).

Potentially important phenotypic differences between clinical strains are also likely associated with strain-specific transcriptional responses. Colangeli et al. (27) profiled drug-sensitive clinical *M. tuberculosis* strains and found that patients infected with pretreatment isolates exhibiting approximately twofold lower minimum inhibitory concentrations (MICs) in isoniazid and rifampicin were significantly more likely to cure following standard therapy. Notably, these subresistance-breakpoint differences in MIC were as predictive of cure versus relapse as all other standard clinical demographics combined but could not be associated with point mutations in genes that had been previously associated with conferring resistance to the drugs profiled (27). These subtle shifts in MIC could potentially be linked to *M. tuberculosis* lineage-dependent induction of drug tolerance during host infection, which could contribute to differential treatment response in vivo (1).

Another potential contribution to subresistance-breakpoint MIC shifts between strains might be the strain-specific gene essentiality differences that have been identified using transposon insertion sequencing (Tn-seq) profiling (20). Among the genes conferring strain-specific fitness costs is the *katG* gene, which catalyzes the conversion of isoniazid into its active form. Notably, the degree of fitness defect associated with *katG* disruption in different clinical strains was not correlated to *katG* SNPs but was associated with the degree to which the strains were susceptible to isoniazid (20). Collectively, these phenotypic differences between strains, in the absence of obvious functional SNP differences in the directly relevant genes, suggest that strain-specific differences in transcriptional responses could contribute to these phenotypes.

STRAIN-SPECIFIC TRANSCRIPTIONAL RESPONSES

Several studies have reported strain-specific differential transcriptional responses in clinical strains during exposure to stresses, including drug treatment and host infection (49, 67, 123). Comparing 17 *M. tuberculosis* strains from 5 different phylogenetic lineages, Homolka and colleagues (58) found that the strains elicited transcriptional profiles that clustered by lineage, during both log-phase growth in broth culture and infection in bone marrow–derived macrophages (BMDMs). Among the hundreds of genes that exhibited lineage-specific expression patterns were lipid utilization and cell wall biosynthesis genes. For example, strains from the Indo-Oceanic lineage had lower levels of *mymA* operon genes, which modify MA components of the cell wall, and strains from the West African lineage 6 had lower expression of sulfolipid biosynthesis and phthiocerol dimycocerosate (PDIM) biosynthesis genes (58). The differential expression of genes in these pathways could modulate remodeling of the mycobacterial cell wall and thus contribute to the decreased fitness observed in these strains during BMDM infection. Lipid metabolism genes were also found to be differentially expressed when comparing BMDM infection responses between the hypervirulent East Asian lineage 2 (Beijing) strain, HN878, and the immunogenic Euro-American lineage 4 strain, CDC1551 (67). Interestingly, the strain-specific pathogen regulation of lipid metabolism is coupled to differential expression of host lipid metabolism genes, which respond in a strain-specific way (67).

Evidence suggests that strain-specific activities of transcriptional regulators as well as strain-specific differences in transcriptional start site sequences and epigenetic modifications could

Minimum inhibitory concentration (MIC): the lowest concentration of drug required to prevent growth of bacteria

Transposon insertion sequencing (Tn-seq): saturating transposon mutagenesis followed by whole-genome sequencing to map transposon insertion sites

potentially contribute to these differential phenotypes (24, 25). Based on genome-wide SNP and indel data from a set of 219 MTBC strains, Chiner-Oms et al. (25) identified 28 transcription factors annotated in H37Rv as either missing or harboring a dysfunctional mutation in one or more of these strains. For example, the deletion of RD743 and RD715 in West African lineage 5 strains affects the transcription factors Rv1994c and Rv2478c, which may in turn affect the expression of associated genes (25). Other studies have identified strain-specific differences in sequence and expression of stress-responsive regulators, including *dosRS* and *phoPR* (15, 34, 35). SNPs in the *phoP* promoter and in the *phoR* kinase, which have been associated with West African lineages 5 and 6 strains, could potentially explain the reduced expression of sulfolipid and PDIM biosynthesis genes observed during macrophage infection, as these genes are regulated by *phoP* (15, 58). Other studies have found strain-specific differences in *dosR* expression under log-phase broth culture and during macrophage infection and have also reported differential overall transcriptional response during exposure to hypoxia in vitro (34, 35). Some of this strain-specific differential regulation could arise from strain-specific rewiring of regulatory subnetworks (24). For example, the *dosR*-associated overexpression in some East Asian lineage 2 (Beijing) strains can be associated with a point mutation upstream of the gene, which creates a novel *sigA* recognition motif (24). Collectively, these data suggest that *M. tuberculosis* harnesses a broad range of mechanisms to coordinate gene expression, many of which exhibit strain-specific diversification.

Epigenetics: describes mechanisms that drive differences in metabolism, physiology, and phenotype that are not based on DNA sequence

EPIGENETICS AND POSTTRANSLATIONAL MODIFICATIONS

Epigenetics has gained much attention as a major contributor to the shaping of host–pathogen interactions through changes in gene expression arising from mechanisms independent of the underlying DNA sequence (47, 65). These epigenetic changes influence cellular functions and are highly complex, occurring through various modification mechanisms, including the formation of secondary structures (e.g., G-quadruplexes), noncoding RNA, and posttranslational modifications of proteins [acetylation, phosphorylation, and changes in histone-like proteins (90, 101)] (**Figure 4**). Additionally, DNA methylation of *M. tuberculosis* has received considerable attention since the discovery of a critical methyltransferase, MamA (108). *M. tuberculosis*-driven epigenetic processes can affect the expression of both bacterial and host genes and are summarized in **Table 2**.

Regulation of gene expression by methylation in TB has been demonstrated to be essential for survival in vivo. The DNA methyltransferase MamA attaches to a 6-bp recognition sequence in approximately 2,000 mycobacterial genes and, in doing so, possibly regulates the expression of over half of the *M. tuberculosis* genome (108). Mutants of *mamA* were shown to grow normally in vitro; however, they were attenuated under hypoxic conditions, suggesting that methylation is required for survival under stringent conditions. Several genes had reduced expression upon mutation of *mamA*, including *whiB7*, Rv0102, Rv0142, and *corA*. In light of the differences in virulence and outcomes of infection among different lineages of *M. tuberculosis*, the sequence of *mamA* was assessed across different lineages. It was observed that Euro-American lineages contain a functional *mamA* gene, while Beijing strains have a point mutation that renders the protein inactive. However, Beijing strains have another functional methyltransferase, HsdM, which is inactivated by a point mutation in Euro-American lineages (108). It remains to be clarified whether these two methyltransferases have different methylation capacities with different effects on virulence.

Phelan et al. (93) characterized the methylome of *M. tuberculosis* using 16 samples that include strains from lineages 1, 2, 4, 5, and 6, revealing three methylated motifs that were detected across most of the isolates irrespective of lineage (93). They also identified partner methylation motifs on both the forward and reverse strands as well as single methylation motifs for which no partner was found. To further investigate differences in methylation patterns, mutations in

Table 2 Epigenetic effects contributing to tuberculosis infection and pathogenesis

Nature of modification	Target in host	Bacterial effector	Possible effect on tuberculosis infection or disease	Reference(s)
Methylation	Hypermethylation of CpG islands of various regulatory regions	Not identified	Dysregulation of gene expression, strain-dependent changes in methylation of inflammatory genes, differential ethnic-based susceptibility	65, 136
Methylation	Histone 3K4	ESAT-6	Reduction of IFN-γ H3K4 methylation	65, 136
Methylation	Histone 3R42	Rv1988 (mycobacterial methyltransferase)	Repression of NOX1 and NOX4, increased bacterial load	65, 136
Methylation	5-Methylcytosine-specific DNA methyltransferase binding to histone 3 and histone 4	Rv2966c	Repression of host *grk5* gene impairing inflammatory cells from migrating to the site of infection	107
Acetylation	Histone 3K5	ESAT-6	Reduction of CIITA pI loci of H3K4 acetylation	63
Acetylation	DUSP16 and free histones	Eis protein	Suppresses the host immune response, increased intracellular survival of bacteria	63
Deacetylation	Histone 3 hypoacetylation	Not identified	Suppression of TH1 responses, knockdown of essential transcriptional regulators, inhibiting crucial protein kinases required for regulating HDAC1, increased bacterial proliferation	21
Deacetylation	Histone 2 MHC2TA and CIITA 2	19-kDa lipoprotein	Reduction of antigen presentation to T cells, immune evasion	63
microRNA				
NA	IFN-γ, TNF-α, IL-6, IL-12	miRNA 29, 144, 155, 99b, 125b, 147, 21	Reduced IFN-γ and TNF-α, modulation of IL-6 and IL-12	55, 63
NA	CAC-NA2D3	miRNA 27a	Downregulation of calcium signaling, inhibition of autophagosome formation	75
NA	NF-κβ	miRNA 27b	Suppression of NF-κβ expression, increase in p53 dependent apoptosis thus enabling bacterial containment	73
NA	IRAK 4	miRNA 27a	Reduction in levels if IRAK 4, dampened immune response	132
NA	UVRAG, N-WASP, FOXO1	miRNA 125a, 142–3p, 582 5p	Dysregulates autophagy, impairs phagocytosis, promotes pathogen survival	9

Abbreviations: IFN-γ, interferon-γ; miRNA, microRNA; NA, not applicable; NOX, NADPH oxidase; TNF-α, tumor necrosis factor-α.

methyltransferase genes associated with each motif (GATN4RTAC: hsdS.1, *hsdM*, and *hsdS*; CTCCAG: *mamA*; CACGCAG: *mamB*), which could putatively explain the loss of function, were identified (93). Three mutations were found for the GATN4RTAC motif in the isolates that lacked methylation, including *hsdM* P306L, *hsdM* G173D, and *hsdS* L119R. Three isolates did not exhibit any methylation at the CTCCAG sites, and the mutations associated with this were *mamA* E270A, which resulted in a frameshift at position 1257 and a novel A460T mutation. The lack of methylation of the CACGCAG motif was associated with a mutation in the *mamB* gene that resulted in a major truncation. The study also revealed lineage-specific methylation effects, possibly related to SNPs, such as a unique mutation S253L in the *mamB* gene in lineage 1 (93). Using lineage 1 and lineage 4 strains, methylation status has been experimentally shown to have a role in transcription, resulting in differential gene expression. A total of 44 genes were differentially expressed based on their methylation status in lineage 4 strains, while this was not observed with lineage 1 strains (48). How this translates to differences in the virulence and transmissibility of these epidemiologically distinct strains remains to be unraveled. While it may be plausible that the methylation status is responsible for these different transcription levels, more recently, mutations in transcription start sites have also been identified as a likely cause for these differences (24).

The role of epigenetics in TB has been recently reviewed by Kathirvel et al. (63). *M. tuberculosis* exports a methyltransferase (Rv1988) which methylates arginine at position 42 of histone 3, thus modulating the transcription of a number of host genes, most of which are involved in evading or dampening host immune responses. These genes include NOX1, NOX4, and NOS, which play a major role in the production of reactive oxygen and nitrogen species. Rv1988 is exclusive to pathogenic *M. tuberculosis* strains; hence, a mutant of *Mycobacterium smegmatis* engineered to carry Rv1988 displayed increased survival in the mice, with high bacterial loads thus confirming this gene as a virulence factor (65, 139). Another methyl transferase (Rv2966c) methylates cytosines of specific DNA sequences in the nucleus in a phosphorylation-dependent manner. Phosphorylation of Rv2966c was demonstrated to strengthen DNA binding capability, as well as significantly increasing methyltransferase activity. Localization experiments confirmed the secretion of Rv2966c into the nucleus of macrophages with specific C-terminal residues required for the nuclear localization. Interaction studies demonstrated that Rv2966c interacts with histone 3, histone 4, and nucleophosmin 1, a histone chaperone used to transcriptionally regulate several genes (107).

Posttranslational modifications of proteins also play a major role in determining the virulence of *M. tuberculosis*. Inhibition of glycosylation via the mutation of genes responsible for *O*-mannosylation in *M. tuberculosis* resulted in the strong attenuation of pathogenicity in mouse models (74). In a recent study, Su et al. (113) demonstrated that the overexpression of Rv1016c, a mannosylated *M. tuberculosis* protein, in *M. bovis* BCG resulted in the loss of protection and increased virulence of the recombinant vaccine strain in mouse and zebrafish models by decreasing the production of IL-2, IL-12-p70, TGF-β, and IL-6, which reduce the protective immune response. Glycosylation patterns across isolates of different lineages have recently been assessed by Birnhanu et al. (11). This study identified over 2,500 glycosylation events in 1,325 proteins that were enriched for cell envelope biosynthesis, fatty acid and lipid metabolism, two-component systems and host–pathogen-interacting molecules. Further analysis revealed that 67 proteins essential for survival in the host are differentially glycosylated among different lineages, suggesting that these may contribute to phenotypic variability among the different lineages (11).

Acetylation status is another posttranslational modification that allows for strain-specific effects. The total acetylome of different lineages of *M. tuberculosis* revealed that lineage 4 and lineage 7 strains displayed differential acetylation in 165 proteins, with lineage 7 exhibiting hypoacetylation in 161 proteins involved in virulence, host interaction, and stress response (46). Comparison of the global acetylation status of *M. tuberculosis* under normal oxygen and hypoxic

conditions revealed that the acetylation of 269 proteins changed during hypoxia. Of particular note, the DosR regulator was deacetylated at position K182, which was subsequently followed by the transcription of the DosR target genes. This highlights deacetylation as an important epigenetic mechanism for the regulation of the dormancy regulon. Further experiments using mouse models demonstrated that defective acetylation of DosR results in lower bacterial counts and reduced pathology, confirming that epigenetic modifications impact the outcome of infection (137). *M. tuberculosis* infection results in the dysregulation of several microRNAs in the serum of TB patients, as well as within human macrophages; some of these are summarized in **Table 2**. However, their association with outcomes, and any strain-specific effects, have not been clearly described.

microRNA: small noncoding RNA that regulates gene expression

Transcriptional biomarkers: a collection of transcriptional patterns that provide a predictive indicator of the progression of disease, treatment response, or risk of disease recurrence

CONCLUSION AND FUTURE OUTLOOK

For decades, researchers assumed that differences in *M. tuberculosis* clinical strains were negligible. In recent years, systems biology tools have helped to reveal meaningful molecular and phenotypic differences between strains that result in functional consequences for infection progression and treatment outcome (see the sidebar titled Phenotypic Heterogeneity in Tubercle Bacteria). These insights collectively suggest that, to be maximally effective, the development of antitubercular interventions (both drugs and vaccines) will need to account for interstrain differences. There are several strategies to address strain-specific differences in response to interventions. First, drug and vaccine development efforts could prioritize targeting processes that promote pathogen clearance consistently across bacterial strains. This would call for expanded strain-specific testing during the discovery phase, exploiting experimental tools including strain-specific Tn-seq profiling (20). Computational models that predict core essential genes and pathways, together with strain-specific response differences, could potentially help streamline the discovery process by informing the selection of representative reference strains. Recent efforts at modeling condition-specific and strain-specific responses in *M. tuberculosis* and other pathogens based on genomic, transcriptomic, and metabolomic data suggest the potential utility of these computational tools (79, 83, 89, 91, 106). Secondly, the development of molecular and functional biomarkers that could rapidly predict strain-specific drug responses could help tailor the application of existing interventions to optimize outcomes. The observation that pretreatment drug susceptibilities correlate significantly to treatment outcome in drug-sensitive strains suggests the potential utility of this as an indicator for whether patients should receive longer therapy (27). These phenotypic biomarkers, paired with genotypic or transcriptional biomarkers reporting on bacterial drug resistance patterns or treatment progression, could help inform regimen selection and duration of treatment while accounting for strain-specific differences (64, 110, 138, 141). The combined effect of these approaches will optimize treatment outcomes.

PHENOTYPIC HETEROGENEITY IN TUBERCLE BACTERIA

In addition to the genetic variability of circulating bacterial strains, a single strain of the tubercle bacillus has the ability to adopt numerous phenotypically distinct drug-tolerant or nonreplicative states. This is best illustrated by the description of differentially culturable tubercle bacteria (DCTB) that have been identified in the sputum of treatment-naive individuals (23). These bacteria are unable to form colonies on agar plates but can be recovered in liquid media supplemented with growth factors. It has been demonstrated that sputum resident DCTB display differential reliance on growth factors, and when combined with the genetic variability in strains, these observations underscore a remarkable degree of heterogeneity and adaptability in tubercle bacteria.

SUMMARY POINTS

1. Tuberculosis (TB) disease in humans is primarily caused by a diversity of *Mycobacterium tuberculosis* strains that can be classified into nine distinct lineages, some of which occur globally, while others are confined to particular geographic regions.
2. The *M. tuberculosis* complex of disease-causing bacteria evolved from a pool of ancestral bacterial strains resembling *Mycobacterium canettii*, with ancient acquisitions of genomic islands from α-, β-, and γ-proteobacteria. Insertion elements and the selective loss of regions of difference contribute to the current variation seen in mycobacterial strains.
3. Lineage 2 (Beijing) strains appear to be associated with significantly severe disease pathology, shorter duration to presentation, meningitis, high frequency of transmission, poor treatment outcomes, higher propensity to develop drug resistance, and extrapulmonary presentations.
4. Numerous genetic determinants in pathogenic mycobacteria have been the subject of positive selection, including, but not limited to, *pe/ppe* and *pe_pgrs* genes, T cell epitopes, genes associated with cell wall biosynthesis, transcriptional regulation, and DNA repair.
5. Coinfection of TB-infected/diseased individuals with HIV perturbs the well-established sympatric host-pathogen relationship that tubercle bacteria have with humans, leading to an allopatric association.
6. Strain-specific virulence, transmissibility, and pathogenic properties are most likely associated with differential activities of transcriptional regulators, differences in transcriptional start site sequences, and epigenetic modifications in both the host and infecting bacteria.

FUTURE ISSUES

1. While the genetic variability in globally circulating *M. tuberculosis* strains has been recognized, how does this diversity contribute to transmission, infection, and disease outcome? Moreover, are there strain-specific biomarkers that differentially predict disease progression, treatment response, and risk of disease recurrence?
2. What are the genetic factors that distinguish successful *M. tuberculosis* strains versus those that display less penetrance in the human population, and how can these be used to design effective vaccination and treatment strategies?
3. Do current animal model systems allow for the dissection of strain-specific effects on pathogenesis, or can these questions only be addressed by studying disease presentation in humans?
4. Given that several successful clinical isolates demonstrate differences in antigen content and the ability to vary T cell epitopes, will a single vaccine be sufficiently protective against all circulating isolates or should strain-specific vaccines be developed for deployment based on geographical distribution of strain types?
5. Are specific *M. tuberculosis* strain types preferentially associated with HIV coinfection, which is a major driver of TB in endemic regions?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by grants from the Bill and Melinda Gates Research Foundation (grant number OPP1100182 to B.D.K. and J.S.P.), South African Medical Research Council with funds from the Department of Health (to B.D.K., J.S.P., R.M.W., A.D., and N.I.), the Department of Science and Innovation (to B.D.K.), the South African National Research Foundation (grant number 120735 to B.D.K., J.S.P., and N.I.), the Howard Hughes Medical Institute (to B.D.K.), the European and Developing Countries Clinical Trials Partnership (grant number TMA2018CDF-2374 to A.D.), and the Research Foundation Flanders (FWO Odysseus grant number G0F8316N) funded Tuberculosis Omics Research (TORCH) Consortium headed by Professor Annelies van Rie (to A.D., N.I., and R.M.W.). We thank Bhavna Gordhan and Christopher Ealand for critical evaluation of the manuscript.

LITERATURE CITED

1. Adams KN, Verma AK, Gopalaswamy R, Adikesavalu H, Singhal DK, et al. 2019. Diverse clinical isolates of *Mycobacterium tuberculosis* develop macrophage-induced rifampin tolerance. *J. Infect. Dis.* 219(10):1554–58
2. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, et al. 2019. Whole genome sequencing of *Mycobacterium tuberculosis* clinical isolates from India reveals genetic heterogeneity and region-specific variations that might affect drug susceptibility. *Front. Microbiol.* 10:309
3. Aguilar-López BA, Correa F, Moreno- Altamirano MMB, Espitia C, Hernández-Longoria R, et al. 2019. LprG and PE_PGRS33 *Mycobacterium tuberculosis* virulence factors induce differential mitochondrial dynamics in macrophages. *Scand. J. Immunol.* 89:e12728
4. Asare P, Asante-Poku A, Prah DA, Borrell S, Osei-Wusu S, et al. 2018. Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. *Int. J. Infect. Dis.* 73:30–42
5. Ates LS. 2020. New insights into the mycobacterial PE and PPE proteins provide a framework for future research. *Mol. Microbiol.* 113(1):4–21
6. Ates LS, Dippenaar A, Ummels R, Piersma SR, Van Der Woude AD, et al. 2018. Mutations in *ppe38* block PE-PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nat. Microbiol.* 3(2):181–88
7. Baena A, Cabarcas F, Alvarez-Eraso KLF, Isaza JP, Alzate JF, Barrera LF. 2019. Differential determinants of virulence in two *Mycobacterium tuberculosis* Colombian clinical isolates of the LAM09 family. *Virulence* 10(1):695–710
8. Baker L, Brown T, Maiden MC, Drobniewski F. 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerg. Infect. Dis.* 10(9):1568–77
9. Bettencourt P, Pires D, Anes E. 2016. Immunomodulating microRNAs of mycobacterial infections. *Tuberculosis* 97:1–7
10. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* 10(1):45–52
11. Birhanu AG, Yimer SA, Kalayou S, Riaz T, Zegeye ED, et al. 2019. Ample glycosylation in membrane and cell envelope proteins may explain the phenotypic diversity and virulence in the *Mycobacterium tuberculosis* complex. *Sci. Rep.* 9:2927
12. Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, et al. 2016. *pks5*-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. *Nat. Microbiol.* 1:15019
13. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, et al. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514(7253):494–97

14. Brennan MJ. 2017. The enigmatic PE/PPE multigene family of mycobacteria and tuberculosis vaccination. *Infect. Immun.* 85:e00969-16
15. Broset E, Martín C, Gonzalo-Asensio J. 2015. Evolutionary landscape of the *Mycobacterium tuberculosis* complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *mBio* 6(5):e01289-15
16. Brown T, Nikolayevskyy V, Velji P, Drobniewski F. 2010. Associations between *Mycobacterium tuberculosis* strains and phenotypes. *Emerg. Infect. Dis.* 16(2):272–80
17. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, et al. 2018. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci. Adv.* 4(10):eaat5869
18. Cambier CJ, Falkow S, Ramakrishnan L. 2014. Host evasion and exploitation schemes of *Mycobacterium tuberculosis*. *Cell* 159(7):1497–509
19. Cantrell SA, Pascopella L, Flood J, Crane CM, Kendall LV, et al. 2008. Community-wide transmission of a strain of *Mycobacterium tuberculosis* that causes reduced lung pathology in mice. *J. Med. Microbiol.* 57:21–27
20. Carey AF, Rock JM, Krieger IV, Chase MR, Fernandez-Suarez M, et al. 2018. TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. *PLOS Pathog*. 14(3):e1006939. Erratum. 2019. *PLOS Pathog*. 15(6):e1007846
21. Chandran A, Antony C, Jose L, Mundayoor S, Natarajan K, et al. 2015. *Mycobacterium tuberculosis* infection induces HDAC1-mediated suppression of *IL-12B* gene expression in macrophages. *Front. Cell. Infect. Microbiol.* 5:90
22. Chen YY, Chang JR, Huang WF, Hsu SC, Kuo SC, et al. 2014. The pattern of cytokine production *in vitro* induced by ancient and modern Beijing *Mycobacterium tuberculosis* strains. *PLOS ONE* 9(4):e94296
23. Chengalroyen MD, Beukes GM, Gordhan BG, Streicher EM, Churchyard G, et al. 2016. Detection and quantification of differentially culturable tubercle bacteria in sputum from tuberculosis patients. *Am. J. Respir. Crit. Care Med.* 194(12):1532–40
24. Chiner-Oms Á, Berney M, Boinett C, González-Candelas F, Young DB, et al. 2019. Genome-wide mutational biases fuel transcriptional diversity in the *Mycobacterium tuberculosis* complex. *Nat. Commun.* 10:3994
25. Chiner-Oms Á, González-Candelas F, Comas I. 2018. Gene expression models based on a reference laboratory strain are poor predictors of *Mycobacterium tuberculosis* complex transcriptional diversity. *Sci. Rep.* 8:3813
26. Chiner-Oms Á, Sánchez-Busó L, Corander J, Gagneux S, Harris SR, et al. 2019. Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex. *Sci. Adv.* 5:eaaw3307
27. Colangeli R, Jedrey H, Kim S, Al E. 2018. Bacterial factors that predict relapse after tuberculosis therapy. *N. Engl. J. Med.* 379:823–33
28. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393(6685):537–44
29. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, et al. 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42(6):498–503
30. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, et al. 2014. Out-of-Africa migration and Neolithic co-expansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45(10):1176–82
31. Coscolla M, Gagneux S. 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26(6):431–44
32. de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, et al. 2008. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in the Gambia. *J. Infect. Dis.* 198(7):1037–43
33. de Souza GA, Fortuin S, Aguilar D, Pando RH, McEvoy CRE, et al. 2010. Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo- and hypervirulent clinical *Mycobacterium tuberculosis* Beijing isolates. *Mol. Cell. Proteom.* 9(11):2414–23
34. Devasundaram S, Khan I, Kumar N, Das S, Raja A. 2015. The influence of reduced oxygen availability on gene expression in laboratory (H37Rv) and clinical strains (S7 and S10) of *Mycobacterium tuberculosis*. *J. Biotechnol.* 210:70–80

35. Devasundaram S, Raja A. 2016. Variable transcriptional adaptation between the laboratory (H37Rv) and clinical strains (S7 and S10) of *Mycobacterium tuberculosis* under hypoxia. *Infect. Genet. Evol.* 40:21–28
36. Domenech P, Kolly GS, Leon-Solis L, Fallow A, Reed MB. 2010. Massive gene duplication event among clinical isolates of the *Mycobacterium tuberculosis* W/Beijing family. *J. Bacteriol.* 192(18):4562–70
37. Drain PK, Bajema KL, Dowdy D, Dheda K, Naidoo K, Schumacher SG. 2018. Incipient and subclinical tuberculosis: a clinical review of early stages and progression of infection. *Clin. Microbiol. Rev.* 31(4):e00021-18
38. Ebrahimi-Rad M, Bifani P, Martin C, Kremer K, Samper S, et al. 2003. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* 9(7):838–45
39. Fenner L, Egger M, Bodmer T, Furrer H, Ballif M, et al. 2013. HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. *PLOS Genet.* 9(3):e1003318
40. Fishbein S, van Wyk N, Warren RM, Sampson SL. 2015. Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity. *Mol. Microbiol.* 96(5):901–16
41. Gagneux S. 2012. Host-pathogen coevolution in human tuberculosis. *Philos. Trans. R. Soc. B* 367(1590):850–59
42. Gagneux S, Narayanan S, Nicol M, Niemann S, Kremer K, et al. 2006. Variable host–pathogen compatibility in *Mycobacterium tuberculosis. PNAS* 103(8):2869–73
43. Gagneux S, Small PM. 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect. Dis.* 7(5):328–37
44. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, et al. 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 368:1575–80
45. Glynn JR. 1998. Resurgence of tuberculosis and the impact of HIV infection. *Br. Med. Bull.* 54(3):579–93
46. Godana Birhanu A, Abebe Yimer S, Holm-Hansen C, Norheim G, Aseffa A, et al. 2017. N ε-and O-acetylation in *Mycobacterium tuberculosis* lineage 7 and lineage 4 strains: Proteins involved in bioenergetics, virulence, and antimicrobial resistance are acetylated. *J. Proteome Res.* 16(11):4045–59
47. Gómez-Díaz E, Jordà M, Peinado MA, Riviero A. 2012. Epigenetics of host-pathogen interactions: the road ahead and the road behind. *PLOS Pathog.* 8(11):e1003007
48. Gomez-Gonzalez PJ, Andreu N, Phelan JE, Florez de Sessions P, Glynn JR, et al. 2019. An integrated whole genome analysis of *Mycobacterium tuberculosis* reveals insights into relationship between its genome, transcriptome and methylome. *Sci. Rep.* 9:5204
49. González-Escalante L, Peñuelas-Urquides K, Said-Fernández S, Silva-Ramírez B, de León MB. 2015. Differential expression of putative drug resistance genes in *Mycobacterium tuberculosis* clinical isolates. *FEMS Microbiol. Lett.* 362(23):fnv194
50. Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, et al. 2014. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *PNAS* 111(31):11491–96
51. Gonzalo-Asensio J, Pérez I, Aguiló N, Uranga S, Picó A, et al. 2018. New insights into the transposition mechanisms of IS*6110* and its dynamic distribution between *Mycobacterium tuberculosis* complex lineages. *PLOS Genet.* 14(4):e1007282
52. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, et al. 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J. Infect. Dis.* 193:121–28
53. Hanekom M, van Pittius NCG, McEvoy C, Victor TC, Van Helden PD, Warren RM. 2011. *Mycobacterium tuberculosis* Beijing genotype: a template for success. *Tuberculosis* 91(6):510–23
54. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CRE, et al. 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J. Clin. Microbiol.* 45(5):1483–90
55. Harapan H, Fitra F, Ichsan I, Mulyadi M, Miotto P, et al. 2013. The roles of microRNAs on tuberculosis infection: meaning or myth? *Tuberculosis* 93(6):596–605
56. Harboe M, Oettinger T, Wiker HG, Rosenkrands I, Andersen P. 1996. Evidence for occurrence of the ESAT-6 protein in *Mycobacterium tuberculosis* and virulent *Mycobacterium bovis* and for its absence in *Mycobacterium bovis* BCG. *Infect. Immun.* 64:16–22

57. Heemskerk D, Caws M, Marais B, Farrar J. 2015. Epidemiology. In *Tuberculosis in Adults and Children*, pp. 1–56. London: Springer
58. Homolka S, Niemann S, Russell DG, Rohde KH. 2010. Functional genetic diversity among *Mycobacterium tuberculosis* complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLOS Pathog*. 6(7):e1000988
59. Huet G, Constant P, Malaga W, Lanéele MA, Kremer K, et al. 2009. A lipid profile typifies the Beijing strains of *Mycobacterium tuberculosis*. Identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J. Biol. Chem.* 284(40):27101–13
60. Iantomasi R, Sali M, Cascioferro A, Palucci I, Zumbo A, et al. 2012. PE_PGRS30 is required for the full virulence of *Mycobacterium tuberculosis*. *Cell. Microbiol*. 14(3):356–67
61. Iwamoto T, Grandjean L, Arikawa K, Nakanishi N, Caviedes L, et al. 2012. Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. *PLOS ONE* 7(11):e49651
62. Jia X, Yang L, Dong M, Chen S, Lv L, et al. 2017. The bioinformatics analysis of comparative genomics of *Mycobacterium tuberculosis* complex (MTBC) provides insight into dissimilarities between intraspecific groups differing in host association, virulence, and epitope diversity. *Front. Cell. Infect. Microbiol.* 7:88
63. Kathirvel M, Mahadevan S. 2016. The role of epigenetics in tuberculosis infection. *Epigenomics* 8(4):547–49
64. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, et al. 2018. Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nat. Commun.* 9:4306
65. Khosla S, Sharma G, Yaseen I. 2016. Learning epigenetic regulation from mycobacteria. *Microb. Cell* 3(2):92–94
66. Konstantynovska O, Rekrotchuk M, Hrek I, Rohozhyn A, Rudova N, et al. 2019. Severe clinical outcomes of tuberculosis in Kharkiv Region, Ukraine, are associated with Beijing strains of *Mycobacterium tuberculosis*. *Pathogens* 8:75
67. Koo M-S, Subbian S, Kaplan G. 2012. Strain specific transcriptional response in *Mycobacterium tuberculosis* infected macrophages. *Cell Commun. Signal.* 10(1):2
68. Koster K, Largen A, Foster JT, Drees KP, Qian L, et al. 2018. Whole genome SNP analysis suggests unique virulence factor differences of the Beijing and Manila families of *Mycobacterium tuberculosis* found in Hawaii. *PLOS ONE* 13(7):e0201146
69. Krishnan N, Malaga W, Constant P, Caws M, Thi Hoang Chau T, et al. 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLOS ONE* 6(9):e23870
70. Lanzas F, Karakousis PC, Sacchettini JC, Ioerger TR. 2013. Multidrug-resistant tuberculosis in Panama is driven by clonal expansion of a multidrug-resistant *Mycobacterium tuberculosis* strain related to the KZN extensively drug-resistant *M. tuberculosis* strain from South Africa. *J. Clin. Microbiol.* 51(10):3277–85
71. Lari N, Rindi L, Bonanni D, Tortoli E, Garzelli C, et al. 2018. Mutations in *mutT* genes of *Mycobacterium tuberculosis* isolates of Beijing genotype. *J. Med. Microbiol.* 55(5):599–603
72. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, et al. 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. *J. Infect. Dis.* 187(1):117–23
73. Liang S, Song Z, Wu Y, Gao Y. 2018. MicroRNA-27b modulates inflammatory response and apoptosis during *Mycobacterium tuberculosis* infection. *J. Immunol.* 200(10):3506–18
74. Liu CF, Tonini L, Malaga W, Beau M, Stella A, et al. 2013. Bacterial protein-*O*-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*. *PNAS* 110(16):6560–65
75. Liu F, Chen J, Wang P, Li H, Zhou Y, et al. 2018. MicroRNA-27a controls the intracellular survival of *Mycobacterium tuberculosis* by regulating calcium-associated autophagy. *Nat. Commun.* 9:4295
76. Liu Q, Luo T, Dong X, Sun G, Liu Z, et al. 2016. Genetic features of *Mycobacterium tuberculosis* modern Beijing sublineage. *Emerg. Microbes Infect.* 5:e14
77. Liu Q, Wang D, Martinez L, Lu P, Zhu L, Lu W. 2019. *Mycobacterium tuberculosis* Beijing genotype strains and unfavorable treatment outcomes: a systematic review and meta-analysis. *Clin. Microbiol. Infect.* 26(2):180–88

78. López B, Aguilar D, Hernández-Pando R, Orozco H, Burger M, et al. 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* 133:30–37
79. Ma S, Minch KJ, Rustad TR, Hobbs S, Zhou SL, et al. 2015. Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. *PLOS Comput. Biol*. 11(11):e1004543
80. Mai TQ, Martinez E, Menon R, Van Anh NT, Hien NT, et al. 2018. *Mycobacterium tuberculosis* drug resistance and transmission among human immunodeficiency virus-infected patients in Ho Chi Minh City, Vietnam. *Am. J. Trop. Med. Hyg.* 99(6):1397–406
81. McEvoy CRE, Cloete R, Müller B, Schürch AC, van Helden PD, et al. 2012. Comparative analysis of *Mycobacterium tuberculosis pe* and *ppe* genes reveals high sequence variation and an apparent absence of selective constraints. *PLOS ONE* 7(4):e30593
82. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, et al. 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* 47(3):242–49
83. Minato Y, Gohl DM, Thiede JM, Maruyama F, Baughn AD, Harcombe WR. 2019. Genomewide assessment of *Mycobacterium tuberculosis* conditionally essential metabolic pathways. *Mol. Biol. Physiol.* 4:e00070-19
84. Naidoo P, Theron G, Rangaka MX, Chihota VN, Vaughan L, et al. 2017. The South African tuberculosis care cascade: estimated losses and methodological challenges. *J. Infect. Dis.* 216(April):S702–13
85. Nemes E, Geldenhuys H, Rozot V, Rutkowski KT, Ratangee F, et al. 2018. Prevention of *M. tuberculosis* infection with H4:IC31 vaccine or BCG revaccination. *N. Engl. J. Med.* 379(2):138–49
86. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, et al. 2020. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat. Commun.* 11:2917
87. O'Neill MB, Mortimer TD, Pepperell CS. 2015. Diversity of *Mycobacterium tuberculosis* across evolutionary scales. *PLOS Pathog*. 11:e1005257
88. Orgeur M, Brosch R. 2018. Evolution of virulence in the *Mycobacterium tuberculosis* complex. *Curr. Opin. Microbiol.* 41:68–75
89. Øyås O, Borrell S, Trauner A, Zimmermann M, Feldmann J. 2020. Model-based integration of genomics and metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. *PNAS* 117(15):8494–502
90. Perrone R, Lavezzo E, Riello E, Manganelli R, Palù G, et al. 2017. Mapping and characterization of G-quadruplexes in *Mycobacterium tuberculosis* gene promoter regions. *Sci. Rep.* 7:5743
91. Peterson EJR, Ma S, Sherman DR, Baliga NS. 2016. Network analysis identifies Rv0324 and Rv0880 as regulators of bedaquiline tolerance in *Mycobacterium tuberculosis*. *Nat. Microbiol*. 1:16078
92. Pheiffer C, Betts JC, Flynn HR, Lukey PT, van Helden P. 2005. Protein expression by a Beijing strain differs from that of another clinical isolate and *Mycobacterium tuberculosis* H37Rv. *Microbiology* 151(4):1139–50
93. Phelan J, de Sessions PF, Tientcheu L, Perdigao J, Machado D, et al. 2018. Methylation in *Mycobacterium tuberculosis* is lineage specific with associated mutations present globally. *Sci. Rep.* 8:160
94. Portevin D, Sukumar S, Coscolla M, Shui G, Li B. 2014. Lipidomics and genomics of *Mycobacterium tuberculosis* reveal lineage-specific trends in mycolic acid biosynthesis. *MicrobiologyOpen* 3(6):823–35
95. Rajwani R, Yam WC, Zhang Y, Kang Y, Wong BKC, et al. 2018. Comparative whole-genomic analysis of an ancient L2 lineage *Mycobacterium tuberculosis* reveals a novel phylogenetic clade and common genetic determinants of hypervirulent strains. *Front. Cell. Infect. Microbiol.* 7:539
96. Raman K, Rajagopalan P, Chandra N. 2005. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLOS Comput. Biol.* 1(5):e46
97. Ranjbar S, Boshoff HI, Mulder A, Siddiqi N, Rubin EJ, Goldfeld AE. 2009. HIV-1 replication is differentially regulated by distinct clinical strains of *Mycobacterium tuberculosis*. *PLOS ONE* 4(7):e6116
98. Reed MB, Domenech P, Manca C, Su H, Barczak AK, et al. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 431(7004):84–87
99. Reed MB, Gagneux S, DeRiemer K, Small PM, Barry CE. 2007. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J. Bacteriol.* 189(7):2583–89
100. Reva O, Korotetskiy I, Ilin A. 2015. Role of the horizontal gene exchange in evolution of pathogenic Mycobacteria. *BMC Evol. Biol.* 15:S2

101. Richard-Greenblatt M, Av-Gay Y. 2017. Epigenetic phosphorylation control of *Mycobacterium tuberculosis* infection and persistence. *Microbiol. Spectr.* 5(2):TBTB2-0005-2015
102. Rutaihwa LK, Menardo F, Stucki D, Gygli SM, Ley SD, et al. 2019. Multiple introductions of *Mycobacterium tuberculosis* lineage 2-Beijing into Africa over centuries. *Front. Ecol. Evol.* 7:112
103. Ryoo SW, Park YK, Park S-N, Shim YS, Liew H, et al. 2007. Comparative proteomic analysis of virulent Korean *Mycobacterium tuberculosis* K-strain with other mycobacteria strain following infection of U-937 macrophage. *J. Microbiol.* 45(3):268–71
104. Saini NK, Baena A, Ng TW, Venkataswamy MM, Kennedy SC, et al. 2016. Suppression of autophagy and antigen presentation by *Mycobacterium tuberculosis* PE-PGRS47. *Nat. Microbiol.* 1:16133
105. Sampson SL. 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clin. Dev. Immunol.* 2011:497203
106. Seif Y, Kavvas E, Lachance JC, Yurkovich JT, Nuccio SP, et al. 2018. Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat. Commun.* 9:3771
107. Sharma G, Upadhyay S, Srilalitha M, Nandicoori VK, Khosla S. 2015. The interaction of mycobacterial protein Rv2966c with host chromatin is mediated through non-CpG methylation and histone H3/H4 binding. *Nucleic Acids Res.* 43(8):3922–37
108. Shell SS, Prestwich EG, Baek S-H, Shah RR, Sassetti CM, et al. 2013. DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. *PLOS Pathog.* 9(7):e1003419
109. Shitikov E, Kolchenko S, Mokrousov I, Bespyatykh J, Ischenko D, et al. 2017. Evolutionary pathway analysis and unified classification of East Asian lineage of *Mycobacterium tuberculosis*. *Sci. Rep.* 7:9227
110. Sigal GB, Segal MR, Mathew A, Jarlsberg L, Wang M, et al. 2017. Biomarkers of tuberculosis severity and treatment effect: a directed screen of 70 host markers in a randomized clinical trial. *EBioMedicine* 25:112–21
111. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *PNAS* 94(18):9869–74
112. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, et al. 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nat. Genet.* 48(12):1535–43
113. Su H, Peng B, Zhang Z, Liu Z, Zhang Z. 2019. The *Mycobacterium tuberculosis* glycoprotein Rv1016c protein inhibits dendritic cell maturation, and impairs Th1 /Th17 responses during mycobacteria infection. *Mol. Immunol.* 109:58–70
114. Supply P, Brosch R. 2017. The biology and epidemiology of *Mycobacterium canettii*. *Adv. Exp. Med. Biol.* 1019:27–41
115. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, et al. 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat. Genet.* 45:172–79
116. Tait DR, Hatherill M, Van Der Meeren O, Ginsberg AM, Van Brakel E, et al. 2019. Final analysis of a trial of M72/AS01$_E$ vaccine to prevent tuberculosis. *N. Engl. J. Med.* 381(25):2429–39
117. Thwaites G, Caws M, Chau TTH, D'Sa A, Lan NTN, et al. 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J. Clin. Microbiol.* 46(4):1363–68
118. Thye T, Niemann S, Walter K, Homolka S, Intemann CD, et al. 2011. Variant G57E of Mannose Binding Lectin associated with protection against tuberculosis caused by *Mycobacterium africanum* but not by *M. tuberculosis*. *PLOS ONE* 6(6):e20908
119. Toungoussova OS, Sandven P, Mariandyshev AO, Nizovtseva NI, Bjune G, Caugant DA. 2002. Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. *J. Clin. Microbiol.* 40(6):1930–37
120. Tracevska T, Jansone I, Baumanis V, Marga O, Lillebaek T. 2003. Prevalence of Beijing genotype in Latvian multidrug-resistant *Mycobacterium tuberculosis* isolates. *Int. J. Tuberc. Lung Dis.* 7:1097–103
121. Trinh QM, Nguyen HL, Nguyen VN, Nguyen TVA, Sintchenko V, Marais BJ. 2015. Tuberculosis and HIV co-infection—focus on the Asia-Pacific region. *Int. J. Infect. Dis.* 32:170–78

122. Tsenova L, Ellison E, Harbacheuski R, Moreira AL, Kurepina N, et al. 2005. 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.* 192(1):98–106
123. Van den Bossche A, Varet H, Sury A, Sismeiro O, Legendre R, et al. 2019. Transcriptional profiling of a laboratory and clinical *Mycobacterium tuberculosis* strain suggests respiratory poisoning upon exposure to delamanid. *Tuberculosis* 117:18–23
124. Van Der Meeren O, Hatherill M, Nduba V, Wilkinson RJ, Muyoyeta M, et al. 2018. Phase 2b controlled trial of M72/AS01$_E$ vaccine to prevent tuberculosis. *N. Engl. J. Med.* 379(17):1621–34
125. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J. Clin. Microbiol.* 29(11):2578–86
126. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, et al. 2000. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. *J. Clin. Microbiol.* 33(12):3234–38
127. van Tong H, Velavan TP, Thye T, Meyer CG. 2017. Human genetic factors in tuberculosis: an update. *Trop. Med. Int. Heal.* 22(9):1063–71
128. Vander Beken S, Al Dulayymi JR, Naessens T, Koza G, Maza-Iglesias M, et al. 2011. Molecular structure of the *Mycobacterium tuberculosis* virulence factor, mycolic acid, determines the elicited inflammatory pattern. *Eur. J. Immunol.* 41(2):450–60
129. Veyrier FJ, Dufort A, Behr MA. 2011. The rise and fall of the *Mycobacterium tuberculosis* genome. *Trends Microbiol.* 19:156–61
130. Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, et al. 2013. *Mycobacterium tuberculosis* Beijing genotype is associated with HIV infection in Mozambique. *PLOS ONE* 8(8):e71999
131. Wampande EM, Mupere E, Jaganath D, Nsereko M, Mayanja HK, et al. 2015. Distribution and transmission of *Mycobacterium tuberculosis* complex lineages among children in peri-urban Kampala, Uganda. *BMC Pediatr.* 15:140
132. Wang J, Jia Z, Wei B, Zhou Y, Niu C, et al. 2017. MicroRNA-27a restrains the immune response to *Mycobacterium tuberculosis* infection by targeting IRAK4, a promoter of the NF-κB pathway. *Int. J. Clin. Exp. Pathol.* 10(9):9894–901
133. Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, et al. 2012. Independent large scale duplications in multiple *M. tuberculosis* lineages overlapping the same genomic region. *PLOS ONE* 7(2):e26038
134. WHO (World Health Organ.) 2012. *Global Tuberculosis Report 2012*. Geneva: WHO
135. Wiens KE, Woyczynski LP, Ledesma JR, Ross JM, Zenteno-Cuevas R, et al. 2018. Global variation in bacterial strains that cause tuberculosis disease: a systematic review and meta-analysis. *BMC Med.* 16:196
136. Yadav V, Dwivedi VP, Bhattacharya D, Mittal A, Das G. 2015. Understanding the host epigenetics in *Mycobacterium tuberculosis* infection. *J. Genet. Genome Res.* 2:016
137. Yang H, Sha W, Liu Z, Tang T, Liu H, et al. 2018. Lysine acetylation of DosR regulates the hypoxia response of *Mycobacterium tuberculosis*. *Emerg. Microbes Infect.* 7(1):34
138. Yang Y, Walker TM, Walker AS, Wilson DJ, Peto TEA, et al. 2019. DeepAMR for predicting co-occurrent resistance of *Mycobacterium tuberculosis*. *Bioinformatics* 35(18):3240–49
139. Yaseen I, Kaur P, Nandicoori VK, Khosla S. 2015. Mycobacteria modulate host epigenetic machinery by Rv1988 methylation of a non-tail arginine of histone H3. *Nat. Commun.* 6:8922
140. Yesilkaya H, Dale JW, Strachan NJC, Forbes KJ. 2005. Natural transposon mutagenesis of clinical isolates of *Mycobacterium tuberculosis*: How many genes does a pathogen need? *J. Bacteriol.* 187(19):6726–32
141. Yong YK, Tan HY, Saeidi A, Wong WF, Vignesh R, et al. 2019. Immune biomarkers for diagnosis and treatment monitoring of tuberculosis: current developments and future prospects. *Front. Microbiol.* 10:2789