Comparative Genomics of Mycobacteria: Some Answers, Yet More New Questions

Marcel A. Behr[1,2]

[1]McGill International Tuberculosis Centre, McGill University, Montreal, Quebec H3G 1A4, Canada
[2]Research Institute of the McGill University Health Centre, Montreal, Quebec H3G 1A4, Canada

Correspondence: marcel.behr@mcgill.ca

Comparative genomic studies permit a genus-level perspective on the distinction between environmental mycobacteria and *Mycobacterium tuberculosis,* as well as a species-level assessment of genetic variability within *M. tuberculosis.* Both of these strata of evolutionary analysis serve to generate hypotheses regarding the genomic basis of *M. tuberculosis* virulence. In contrasting lessons from macroevolutionary study and microevolutionary study, one can form predictions about which segments of the genome are likely to be essential for or dispensable for the pathogenesis of tuberculosis. Although some of these predictions have been experimentally verified, notable exceptions challenge the direct link between these virulence factors and the capacity of *M. tuberculosis* to successfully cause disease and propagate between human hosts. These unexpected findings serve as the stimulus for further studies, using genomic comparisons and other approaches, to better define the remarkable success of this recalcitrant pathogen.

Since the first genome sequence of *Mycobacterium tuberculosis* was published in 1998 (Cole et al. 1998), a number of investigators have analyzed and used these data to ask what makes this bacteria particularly effective at causing infection and disease in humans. One attractive approach involves comparing the genome of the reference strain *M. tuberculosis* H37Rv with other related bacteria; in theory, genes that are deleted among bacteria that successfully cause disease are dispensable for full virulence (Tsolaki et al. 2004), whereas genes that have been deleted in strains of low virulence present as candidates for virulence genes (Behr et al. 1999). Although the technologies applied have evolved considerably from late 1990s, the goal of this review is not to focus on the techniques (subtractive hybridization [Mahairas et al. 1996], DNA microarrays [Behr et al. 1999], GeneChips [Salamon et al. 2000], tiling arrays [Leung et al., 2008], whole-genome sequencing [Garnier et al. 2003; Brosch et al. 2007]) but instead to highlight certain key findings that have been confirmed in controlled, experimental studies. Additionally, an attempt will be made to introduce somewhat unexpected observations from comparative genomics that suggest that our understanding of the genomic basis of *M. tuberculosis* pathogenicity might benefit from further investigation.

In terms of strata of investigation, comparative genomics can address microevolution or macroevolution. Because *M. tuberculosis* is strictly a subspecies of the *M. tuberculosis* complex, one can consider genomic differences at the level of (1) variability within *M. tuberculosis* sensu stricto, such as strain–strain differences; (2) variability within the *M. tuberculosis* complex, such as *M. tuberculosis* versus *Mycobacterium bovis*; and (3) variability between *M. tuberculosis* and nontuberculous mycobacteria (NTM), such as *Mycobacterium kansasii* and *Mycobacterium marinum*. In this review, I will comment on each of these levels of variability, citing selectively rather than systemically from the published literature that has accrued in the past two decades of comparative genomic studies.

M. tuberculosis SENSU STRICTO VARIABILITY

The existence of *M. tuberculosis* variability was recognized for more than two decades for molecular epidemiologic studies that exploited DNA differences between isolates to track transmission patterns in space and time (discussed in more detail in Niemann and Supply 2014). Using a number of different genotyping modalities, it has now been robustly shown that there are genetically defined lineages of *M. tuberculosis* (Fig. 1) that differ in their geographic distribution, broadly following the history of human migrations out of Africa toward different re-

gions of the planet. A key pending question is whether these genomically distinct strains of *M. tuberculosis* differ in terms of their ability to infect, cause disease, and transmit across human populations.

A number of papers have experimentally documented *M. tuberculosis* strain differences, in vitro more so than in vivo (Manca et al. 1999, 2001, 2004; Reed et al. 2004, 2007; Barczak et al. 2005; Tsenova et al. 2005; Fallow et al. 2010; Homolka et al. 2010; Krishnan et al. 2011; Reiling et al. 2013). However, a key epidemiologic filter that should not be overlooked is that the bacteria chosen for such studies have all most recently caused clinical disease. Therefore, although we might potentially be exposed to mutants that are unable to cause infection, or we might be infected with mutants that are unable to cause disease, these bacteria will not available for genomic interrogation. When we do comparative genomics of *M. tuberculosis* isolates, we cannot ask about the absolute capacity to cause disease, but rather about relative differences in pathogenicity, *among successful strains*. A further consideration is that when experimentally testing whether certain strains are more or less virulent, many investigators use a murine model that best approximates progressive primary tuberculosis (TB) disease, without latency, reactivation, or chronic cavitary disease. In such a model, one might interpret an earlier time to death as evidence of a more "virulent" strain, which it clearly is, from the vantage of the in-
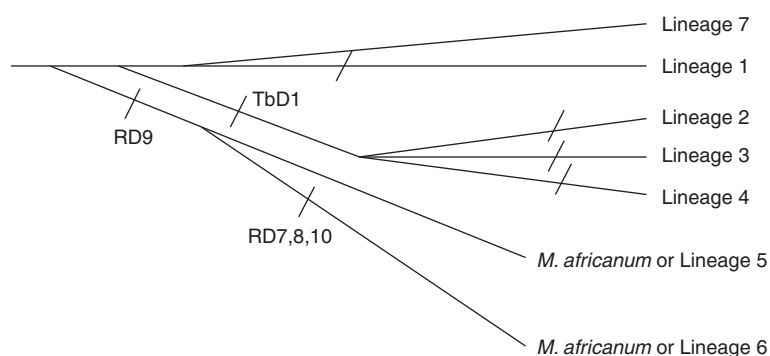

Figure 1. Phylogenetic relationship of *Mycobacterium tuberculosis* lineages. Lineages 5 and 6 were previously named *Mycobacterium africanum*, types 1 and 2, although some authors consider these to be two lineages of human-adapted *M. tuberculosis*. The short lines at right angles to lineages refer to specific genomic deletions.

fected mouse (Reed et al. 2004). However, this laboratory phenotype may be a proxy for lower transmissibility, because invasive, fatal TB is detrimental to the fitness of both the host and the pathogen. Conversely, a strain that is unable to kill the mouse after a prolonged infection might be deemed to have attenuated virulence (Manca et al. 1999), when an alternative interpretation is that this strain is better suited for generating a chronic pulmonary focus in a relatively well host. Thus, before asking whether we can document different between strains in controlled, experimental infections, we need to ask whether the laboratory models we use address the epidemiologic observations of "success" in the field. Whether different strains have a different propensity toward drug resistance and whether this could impart fitness advantages in the face of antibiotic pressures are completely separate matters that are addressed in Niemann and Supply (2014).

Given the preceding considerations, it can be hypothesized that genomic regions that are deleted in isolates from active TB cases might be "dispensable" regions that are not absolutely required for the capacity of the organism to cause transmissible disease. This latter proposition, however, challenges the importance of certain experimentally proven virulence factors. For example, Lineage 4 (or Euro-American) bacteria are incapable of producing phenolic glycolipids (PGLs), a property that has been variably associated with virulence in murine models of disease (Reed et al. 2004; Sinsimer et al. 2008). Yet, Lineage 4 bacilli successfully swept across the Americas, as documented in molecular epidemiologic papers tracking their spread in historical (Pepperell et al. 2011) and modern epidemics (Nguyen et al. 2003). Does this suggest something about the human hosts encountered by these bacteria, or does this imply that *M. tuberculosis* can successfully cause disease without PGLs by altering the expression and/or production of alternative virulence factors? Another example is the *DosR* regulon, so named because of the regulator that controls expression of a defined set of genes upon bacterial exposure to hypoxic conditions (Park et al. 2003). In laboratory investigations, it has been

shown that this regulatory response is critical for the capacity of *M. tuberculosis* to endure low oxygen tension (Gautam et al. 2014). Furthermore, the demonstration that some of the proteins coded by *DosR* genes elicit adaptive immune responses in mice and humans suggests that at least some of these proteins are involved in infection, persistence, or the generation of transmissible disease (Leyten et al. 2006; Roupie et al. 2007). Yet, sequence analysis of a Colombian virulent isolate found a 3.6-kB deletion that eliminates two *DosR* genes, *Rv1996* and *Rv1997* (Isaza et al. 2012). Furthermore, transcriptome studies indicate that some of the so-called Beijing strains have lost the fine-tuning ability of the *DosR* regulon and constitutively express these genes instead (Fallow et al. 2010). Again, does this suggest that certain components of the *DosR* regulon are not important in certain people or rather does the genome provide redundant mechanisms to cope with changes in oxygen tension during the infection cycle? If PGLs and *DosR* are not essential for *M. tuberculosis* to cause disease *in natura*, are there unknown virulence mechanisms that remain to be discovered?

The converse approach is to ask which genes are retained in clinical isolates, as there are two possible explanations: Either these genes are situated in regions of the genome that do not accommodate genomic disruption or deletions at these sites are structurally permissible but have been selected against during the thousands of years of *M. tuberculosis*−human coevolution. The list of conserved genes is not readily accessible, in part because most publications detail missing genes and in part because whole-genome sequencing of isolates from around the world is ongoing. Therefore, with one publication, a "conserved" gene today can become a "dispensable" gene tomorrow. However, among a smaller list of genes that appear to be consistently found in isolates to date, it is possible to ask about variability within genes, as has been performed in a study of genes coding for experimentally proven antigenic proteins. This provocative paper found that genes for antigens are less variable than genes defined as "essential genes" in transposon screens (Sassetti and Ru-

bin 2003; Sassetti et al. 2003), and that the motif coding for the epitopes was most conserved (Comas et al. 2010). This finding suggests that *M. tuberculosis* may depend on adaptive immune responses for its capacity to cause disease and/or transmit, such that mutations in these epitopes adversely affect the bacterial life cycle. The notion that *M. tuberculosis* requires the elaboration of a cell-mediate immune response to successfully transmit among humans has received independent support from epidemiologic studies, in which HIV-negative TB patients are more likely to transmit to their contacts than TB patients suffering from coinfection from HIV/AIDS (Corbett et al. 2006).

Beyond providing lists of genes that are variable or conserved, comparative genomics allows one to derive robust phylogenies of lineages that are found around the world, revealing geographic associations of strains over time suggestive of a stable host–pathogen relationship. Settings with more than one strain offer epidemiologists an opportunity to ask whether different major lineages of *M. tuberculosis* behave comparably *in natura*. Although there have been a number of reports in which one or another genomically defined lineage has been particularly successful, it is not always clear whether that particular strain was "good" or "lucky." For instance, an outbreak of TB in an AIDS hospice does not prove a more successful strain when a more likely explanation here would be a cluster of profoundly susceptible hosts (Daley et al. 1992). Furthermore, in many molecular epidemiology studies, the foreign/imported strains were less likely to transmit than the local strain, for a variety of social and behavioral reasons (Alland et al. 1994; Small et al. 1994). Thus, if a molecular epidemiology study investigates bacterial lineage, it can be predicted that the "local" strain may appear to be a risk factor for ongoing transmission, when in fact the apparent association may be confounded by the kinds of people who are infected with this strain. An alternative to looking at bacterial performance in a setting with high transmission is to observe the clinical and public health consequences of different strains that are introduced into a low-transmission setting. In Montreal,

Canada, *M. tuberculosis* is imported via immigration from more than 80 countries in the world, whereupon these bacteria encounter a city with very little ongoing transmission (Kulaga et al. 2002). Additionally, all TB patients are managed in a comparable manner through a publicly funded health-care system, with patient care evidently provided blinded to strain identity. By typing isolates according to the major global lineages (Reed et al. 2009), we have found remarkably few differences in clinical outcomes or rates of tuberculin conversions of household contacts (Albanna et al. 2011). Importantly, TB due to the East-Asian/Lineage 2/Beijing strain has not behaved in a discernibly different manner than other strains from around the world, an observation also described in Switzerland (Fenner et al. 2012). Moreover, although we have obtained no evidence whatsoever of a "more virulent" strain in Montreal, in-depth analysis of Lineage 3 (sometimes called East-African-Indian lineage) revealed that these organisms were *less likely* to cause severe disease and that close contacts of patients producing these organisms were *less likely* to have a tuberculin conversion (Albanna et al. 2011). How a strain can propagate despite lower virulence begs a largely theoretical question. One possibility is that strains that evolved in densely populated regions of the world might depend on a high-transmission environment for their propagation. If true, when such a strain is introduced into a low-transmission environment, the relatively reduced virulence is more easily observed. The notion of whether (and how) strains with lesser virulence are able to persist in their host population is reconsidered below in the section on *M. tuberculosis* complex variability.

M. tuberculosis COMPLEX VARIABILITY

The existence of phenotypic differences between the human and the bovine tubercle bacillus was documented in the 19th century. These organisms, subsequently named *M. tuberculosis* and *M. bovis*, were assayed in a comparative virulence assay; both organisms were able to cause disease in guinea pigs but only *M. bovis* was

virulent in rabbits (Smith 1898). Subsequent laboratory isolations revealed a collection of related bacteria, of which at least eight are recognized members of the *M. tuberculosis* complex (MTC), complete with a Latin name: *M. tuberculosis*, *Mycobacterium africanum*, *Mycobacterium microti*, *Mycobacterium mungi*, *Mycobacterium orygis, Mycobacterium pinnipedii*, *Mycobacterium caprae*, and *M. bovis*. The dassie bacillus has not been formally named but has a distinct genomic profile that will be detailed below (Mostowy et al. 2004). A single recent isolate from a wild chimpanzee has been subject to genomic study, potentially suggesting either a new lineage, but more such isolates will be required before its situation in the MTC can be affirmed (Coscolla et al. 2013). A group of bacteria called the smooth tubercle bacilli (STB), including *Mycobacterium canetti*, are considered by some to be an out-group and by others to be members of the MTC (Gutierrez et al. 2005; Brisse et al. 2006; Supply et al. 2013). With further sampling of mammalian hosts, in the wild and in zoos, it is possible that new variants of the MTC might be uncovered in coming years.

In terms of genetic differences, a number of platforms have been used to confirm distinct genomic identities of MTC members (Fig. 2) (Brosch et al. 2002; Mostowy et al. 2002). What is harder to establish is the phenotypic basis for these variable host–pathogen relationships. For some of these organisms, the bacteria are not easily obtained for laboratory investigation (Alexander et al. 2010). Even if these were to be obtained and grown to a standard inoculum, with minimal in vitro passage, the kinds of comparative virulence studies conducted by Theobald Smith in the 1890s may not explain how each MTC member causes disease in its own particular host. Therefore, to date, genomic comparisons among the MTC have helped formulate a MTC phylogeny, but linking genomic deletions to a phenotypic effect suffers from many of the same issues raised for *M. tuberculosis* sensu stricto, with the additional complicating factor of the many different hosts involved. Because it is not currently feasible to experimentally complete the in vivo 8×8 host–pathogen matrix (i.e., by putting *M. pinnipedii* into dassies, etc.), the explanation for the long-standing associations might be an intimate host–pathogen interaction determined by a specific molecular fit or more trivially by the possibility that dassies have never been exposed to *M. pinnipedii*.
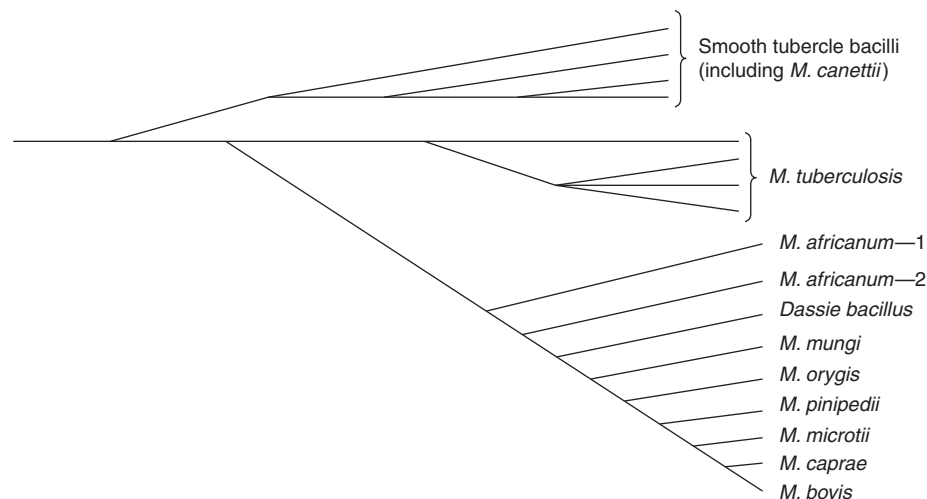

Figure 2. Phylogenetic relationship of *Mycobacterium tuberculosis* complex (MTC) lineages. Here, *Mycobacterium africanum* is clustered with "exotic" members of the MTC based on shared genomic deletions. Only virulent *Mycobacterium bovis* is shown, with *M. bovis* BCG representing a laboratory-adapted strain of *M. bovis*.

As noted before, if a MTC variant lacks a genomic region and is able to successfully transmit within its host species, it can be argued that genes in this deleted region are dispensable for full virulence. Yet, the RD1 locus, now known to code for proteins of the ESX-1 secretion system, has been independently deleted three times in naturally occurring members of the MTC. RD1 was first described as deleted from BCG (Bacillus Calmette–Guérin) vaccines (Mahairas et al. 1996), and this deletion is now accepted to contribute to the safety profile of these live, attenuated strains (Pym et al. 2002; Lewis et al. 2003). Yet, genomic studies have documented that *M. microtii* (Pym et al. 2002) and the dassie bacillus (Mostowy et al. 2004) have distinct deletions in the RD1 locus, such that both bacteria lack the genes *Rv3874-Rv3876*, including *esxB* (coding for CFP-10) and *esxA* (coding for ESAT-6). More recently, a genomic study of the tubercle bacillus associated with disease in mongooses reported a new RD1 deletion, in the organism now named *M. mungi* (Alexander et al. 2010). How can three different MTC members successfully cause disease and transmit among members of their respective mammalian hosts in the absence of the ESX-1 secretion system? Is ESX-1 overrated, or is this virulence system not needed for disease in these hosts? If the latter, what is the common feature of voles, dassies, and mongooses that could explain the convergent deletion of RD1 in their respective tubercle bacilli? In the absence of experimental data, the answer to this question is perforce speculative. One hypothesis notes that voles, dassies, and mongooses are burrowing animals that inhabit spaces with limited ventilation, unlike oryxes, seals, goats, and cattle. Because the life cycle of *M. tuberculosis* involves (1) achieving productive infection in a number of contacts and (2) having a subset of these progress to active, transmissible disease, a bacterium with a reduced attack rate could achieve a transmission index of unity or above as long as it is being spread in a high-transmission environment. If true, this may explain why *M. microtii*, *M. mungi*, and the dassie bacillus are not isolated from human TB cases, unlike *M. pinnipedii* (Thompson et al. 1993), *M. caprae* (Kubica et al.

2003), and *M. bovis* (Evans et al. 2007). Given that TB in humans may be considered a "crowd" disease, the increasing density of human communities over time may also permit a gradual attenuation of virulence of the human-adapted form, potentially explaining why we see reduced transmission of at least one *M. tuberculosis* lineage among immigrants to Montreal.

Because there are limited phenotypic data for the "exotic" members of the MTC, the most phenotypic data on MTC variability derive from the classical comparison of *M. tuberculosis* versus *M. bovis*. In many animal models, *M. bovis* is more "virulent" (Smith 1898; Medina et al. 2006; Nedeltchev et al. 2009), although as mentioned above, the laboratory definition of virulence again may not serve as the most appropriate proxy for bacterial transmissibility. In natural cross-host studies, *M. bovis* does not successfully transmit among human populations; conversely, *M. tuberculosis* has been used as a naturally attenuated vaccine in cattle. Therefore, the notion of greater or lesser virulence has to be framed in a host-specific context. Nonetheless, the observation that *M. bovis* has managed to generate sustained infections in a broader range of hosts (cattle, deer, bison, brush-tailed possums, badgers, East-Asian water buffaloes) suggests some sort of adaptive gain of the veterinary pathogen, as compared with the restricted host distribution of the human variant. As *M. bovis* has a smaller genome, with no unique DNA, it has been inferred that any gain-of-function phenotypes in *M. bovis* must be caused by altered regulation of extant genes (Garnier et al. 2003).

In an attempt to explain the basis for the different behaviors of *M. tuberculosis* and *M. bovis*, our laboratory and others have performed comparative transcriptomics of these organisms, to catalog differences in gene expression (Said-Salim et al. 2006; Rehren et al. 2007). Perhaps the most obvious difference involves a small set of genes under control of the extracytosolic regulator, Sigma factor K (SigK). Previously it had been shown that virulent *M. bovis* and some (but not all) *M. bovis* BCG strains produce high quantities of the antigenic proteins MPT70 and MPT83, with MPT70 produc-

tion accounting for >10% of the mass of culture filtrate proteins (Harboe and Nagai 1984). In contrast, *M. tuberculosis* does not produce these proteins in vitro, but expression of the genes is seen rapidly after infection (Schnappinger et al. 2003). This offers three "states": *M. tuberculosis* (LOW-INDUCIBLE), *M. bovis* including BCG Russia (ON), and selected BCG including BCG Pasteur (OFF). By transcriptome analysis, we showed that the differential expression of MPT70 and MPT83 across BCG strains was caused by a start mutation in *sigK* that occurred between 1927 and 1931 at the Institut Pasteur (Charlet et al. 2005). Given that the SigK mutation explains BCG differences, we next turned to the *M. tuberculosis*−*M. bovis* difference, showing that the cognate anti-Sigma K, called the Regulator of SigK (RskA), is mutated in *M. bovis* (Said-Salim et al. 2006). Curiously, *rskA* is also mutated in *M. orygis*, with a distinct, independent mutation also resulting in overexpression of the SigK regulon. The explanation for how two different MTC members have independently achieved constitutive expression of these antigenic proteins remains elusive. In unpublished work (MA Behr and FJ Veyrier, unpubl.), it has been shown that the mutant forms of *rskA* are not simply defective for inhibiting SigK activity; but rather that introduction of $rskA_{bovis}$ or $rskA_{orygis}$ into *M. tuberculosis* activates the SigK regulon, resulting in constitutive production of MPT70 and MPT83, even in the presence of a wild-type RskA. Ongoing work seeks to define the phenotypic consequence of altering the expression of the SigK-controlled genes, with a view to understanding how such an expression change could have been twice selected among MTC species circulating in different mammalian hosts.

Mycobacterium-LEVEL VARIABILITY

The preceding sections focused first on variability between *M. tuberculosis* strains and then variability between MTC species. By "zooming up" one more level, one can ask about the similarities and differences between *M. tuberculosis* and nontuberculosis mycobacteria (NTM). Through sequence analysis of a number of NTM species, it is currently appreciated that the NTM that are genetically most similar to *M. tuberculosis* are *M. marinum* and *M. kansasii* (Fig. 3) (Stinear et al. 2008; Veyrier et al. 2009; Wang and Behr 2014). *M. marinum* has already been used with great success as a model for TB pathogenesis, given the conservation of ESX-1 secretion across organisms (Gao et al. 2004) and the optical advantages of live videomicroscopy for in vivo infection studies (Cosma et al. 2003; Ramakrishnan 2012). The *M. marinum* genome has served as the first out-group to ask what genes are present in *M. tuberculosis* but lacking in NTM species, suggesting the acquisition of new *M. tuberculosis*−specific genes by the ancestor of the MTC, after their common ancestor (Stinear et al. 2008). These findings were supported by an independent compositional analysis of the *M. tuberculosis* genome that also sug-
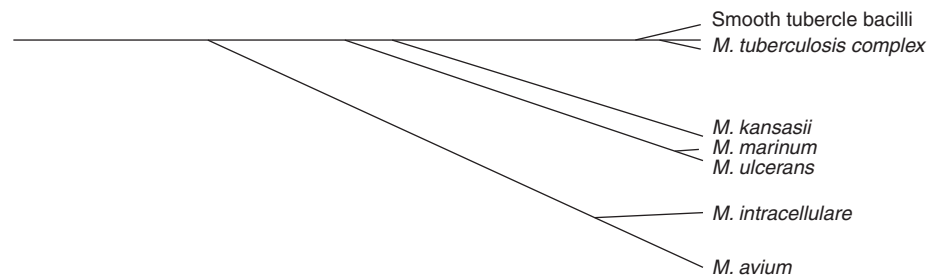

Figure 3. Phylogenetic relationship of slow-growing mycobacterial species, based on multilocus sequence analysis of 20 conserved housekeeping genes (as described in Veyrier et al. 2009). Outside of the *Mycobacterium tuberculosis* complex, the closely nontuberculous mycobacteria are *Mycobacterium kansasii* and the *Mycobacterium marinum-ulcerans* complex, which are organisms that share a number of virulence determinants with *Mycobacterium tuberculosis*.

gested the existence of horizontal gene transfer (HGT) as the certain *M. tuberculosis* genes (Becq et al. 2007).

M. kansasii offers a further perspective on the emergence of *M. tuberculosis* for a number of reasons. Epidemiologically, *M. kansasii* is recognized to be an environmental mycobacterium that can be readily isolated from water sources, but does not spread from person to person. Thus the genome of *M. kansasii* affords a perspective of common mycobacterial survival strategies that are not necessarily selected for the capacity to infect or cause disease. Yet, clinically, *M. kansasii* lung disease can present as right upper lobe cavitary disease that is indistinguishable from TB. This suggests that certain features of the pathogenesis of TB, including inhalation of bacteria into the alveoli and intrapulmonary spread to the lung apices, are not directed by *M. tuberculosis* in a specific manner. Genetically, analysis of 20 housekeeping genes conserved across mycobacterial genomes suggests that *M. kansasii* may be a closer relative of *M. tuberculosis* than *M. marinum* (Veyrier et al. 2009). Based on these considerations, and the shotgun sequence reads for the *M. kansasii* genome, we asked whether the *M. tuberculosis* genome contains genes that lack orthologs in *M. kansasii*, *M. marinum*, and other NTM. Our analysis suggested the stepwise accumulation of genes coincident with the divergent of slow-growing mycobacteria from rapid-growing mycobacteria, and then after the divergence with *M. avium*, the divergence with *M. marinum*, and the divergence with *M. kansasii* (Veyrier et al. 2011). *M. tuberculosis* contains at least 55 coding genes that lack orthologs in all other mycobacteria, with bioinformatic evidence that they were acquired via HGT including the presence of vectors of HGT and differences in the GC content of these putative HGT islands (Veyrier et al. 2009). As there are considerable differences between conserved sequences of *M. tuberculosis* and *M. kansasii*, it cannot be inferred that these are closely related organisms that recently diverged because of HGT elements, as is the case with other pathogens, like *Yersinia pestis*. Rather, the emergence of *M. tuberculosis* has apparently occurred over a protracted peri-

od of time, with a number of stepwise additions of HGT islands, which together enabled an environmental mycobacteria to alter its lifestyle to become a committed professional pathogen (Veyrier et al. 2011). At present there are no obvious intermediate hosts that can be assessed to test this model, with the exception of STB. However, the number of STB isolates is small, arguing against an ongoing human epidemic, and their definitive reservoir is presently unknown. Should STB, including *M. canetii*, be isolated from environmental sources, such as water, then common features of the *M. canetii* and *M. kansasii* genomes might point to environmental survival strategies, whereas differences between these organisms and *M. tuberculosis* would point to key pathogen-specific elements. Alternatively, if STB are host-adapted organisms, without an environmental reservoir, then the study of the STB genomes might provide a glimpse into genomic remodeling of an apparently moribund lineage of the tubercle bacillus. More information about the epidemiologic profile and potential sources of exposure of humans with TB due to STB isolates might shed some light on these possibilities.

Beyond looking at differences between *M. tuberculosis* and NTM, the availability of NTM genomes also provides an opportunity to catalog conserved genes. Just as the genes conserved across *M. tuberculosis* isolates suggested selection for these genes to cause disease, the conservation of genes across a broad range of mycobacterial species suggests a selection for these genes to survive and replicate in the various environments these bacteria encounter. From the vantage of TB research, it can be argued that *M. tuberculosis* must survive the first contact with human cells, persist within granulomatous lesions, induce immunopathology, and spread via the airways in aerosols to the next host. Thus, it follows that *M. tuberculosis* must regulate gene expression, protein production, and protein degradation to appropriately adjust to the various conditions encountered during the infectious cycle. From the vantage of NTM research, it can be argued that *M. avium* in a pond must endure much more severe fluctuations in temperature and pH, while with-

standing the antimicrobial pressures of a polymicrobial community of pond-dwelling organisms. Thus, it is perhaps not surprising that NTM have larger genomes and more Sigma factors than *M. tuberculosis*. By extension, a number of *M. tuberculosis* "virulence factors" are not specific to the pathogen; to a varying extent, NTM genomes code for the ESX-1 secretion system, PhoP and LprG, among other examples. Although these observations by no means detract from the importance of these systems for the full virulence of *M. tuberculosis*, from the vantage of comparative genomics, the existence of these elements in NTM genomes suggests that these virulence factors cannot on their own explain the remarkable success of *M. tuberculosis* as a human pathogen. Recent findings that a gene island specific to *M. tuberculosis* (Rv3377-78), two genes previously implicated in the modulation of phagosomal acidification (Pethe et al. 2004), produces a novel lipid, called 1-tuberculosinyladenosine (Layre et al. 2014). This finding provides an example of the kinds of investigations that may shed further light on the genesis of *M. tuberculosis*. Further studies of *M. tuberculosis*–specific genes, many of which have currently have unknown or hypothetical functions, might ultimately provide fundamental insights on *M. tuberculosis* as a pathogen and also uncover novel, TB-specific biomarkers for clinical applications.

CONCLUDING REMARKS

After nearly two decades of study, it can be affirmed that comparative genomics helped uncover the ESX-1 secretion system but then challenged whether it is essential for virulence in all members of the MTC. On a microevolutionary scale, comparative genomics has shown that strains of *M. tuberculosis* lacking in PGLs and *DosR* regulon genes are able to cause disease while on a macroevolutionary scale, comparative genomics has shown that a number of virulence factors critical for TB disease are conserved in low-virulence environmental mycobacteria. Thus, there is no single virulence factor yet that is both specific to *M. tuberculosis* and invariant across TB-causing bacteria.

Together these observations beg the question of whether comparative genomics can detect the elusive nature of what makes *M. tuberculosis* such a recalcitrant foe. Just as the human genome has not laid bare what makes people different than chimpanzees, perhaps the *M. tuberculosis* genome is the cornerstone for a different kind of investigation, asking not simply what is present and absent, but how the ensemble is orchestrated, at the levels of transcription, translation, and molecular recycling pathways. Comparative genomics has clearly provided guidance of cataloguing the right bacteria for investigation; misnomers can confound investigation and lead to false hypotheses that can never be confirmed despite the efforts of our trainees. Thus, we now know that *M. avium* is a complex of both pathogenic and nonpathogenic bacteria (Turenne et al. 2007), that *M. marinum* and *Mycobacterium ulcerans* diverged quite recently (Veyrier et al. 2009), and that *M. kansasii* is closely related to *M. tuberculosis* but epidemiologically quite different. We now know that within the MTC there are many variants of the tubercle bacillus that are apparently persisting in symbiotic relationships with their respective hosts. And we now know that within the *M. tuberculosis* sensu stricto, there is also evidence of symbiosis, defined as cohabitation of divergent organisms (Gagneux et al. 2006). Perhaps what is needed next is a global perturbogen, to disrupt the apparently comfortable residence that *M. tuberculosis* has established in one-third of humanity. Whether this can be achieved with existing tools, applied in a scaled-up manner, or achieved with novel approaches, to be derived from other research described in this book, is itself an ongoing question.

REFERENCES

*Reference is also in this collection.

Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, Menzies D. 2011. Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. *PLoS ONE* **6:** e25075.

Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC. 2010. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* **16:** 1296–1299.

Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, Drucker E, Bloom BR. 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* **330:** 1710–1716.

Barczak AK, Domenech P, Boshoff HI, Reed MB, Manca C, Kaplan G, Barry CE III. 2005. In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J Infect Dis* **192:** 600–606.

Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P. 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* **24:** 1861–1871.

Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284:** 1520–1523.

Brisse S, Supply P, Brosch R, Vincent V, Gutierrez MC. 2006. "A re-evaluation of M. prototuberculosis": Continuing the debate. *PLoS Pathog* **2:** e95.

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, et al. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci* **99:** 3684–3689.

Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos SS, Duthoy S, Lacroix C, Garcia-Pelayo C, et al. 2007. Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci* **104:** 5596–5601.

Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG, Behr MA. 2005. Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in sigK. *Mol Microbiol* **56:** 1302–1313.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537–544.

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S. 2010. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* **42:** 498–503.

Corbett EL, Marston B, Churchyard GJ, De Cock KM. 2006. Tuberculosis in sub-Saharan Africa: Opportunities, challenges, and change in the era of antiretroviral treatment. *Lancet* **367:** 926–937.

Coscolla M, Lewin A, Metzger S, Maetz-Rennsing K, Calvignac-Spencer S, Nitsche A, Dabrowski PW, Radonic A, Niemann S, Parkhill J, et al. 2013. Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. *Emerg Infect Dis* **19:** 969–976.

Cosma CL, Sherman DR, Ramakrishnan L. 2003. The secret lives of the pathogenic mycobacteria. *Annu Rev Microbiol* **57:** 641–676.

Daley CL, Small PM, Schecter GF, Schoolnik GK, McAdam RA, Jacobs WR Jr, Hopewell PC. 1992. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms. *N Engl J Med* **326:** 231–235.

Evans JT, Smith EG, Banerjee A, Smith RM, Dale J, Innes JA, Hunt D, Tweddell A, Wood A, Anderson C, et al. 2007. Cluster of human tuberculosis caused by *Mycobacterium bovis*: Evidence for person-to-person transmission in the UK. *Lancet* **369:** 1270–1276.

Fallow A, Domenech P, Reed MB. 2010. Strains of the East Asian (W/Beijing) lineage of *Mycobacterium tuberculosis* are DosS/DosT-DosR two-component regulatory system natural mutants. *J Bacteriol* **192:** 2228–2238.

Fenner L, Gagneux S, Helbling P, Battegay M, Rieder HL, Pfyffer GE, Zwahlen M, Furrer H, Siegrist HH, Fehr J, et al. 2012. *Mycobacterium tuberculosis* transmission in a country with low tuberculosis incidence: Role of immigration and HIV infection. *J Clin Microbiol* **50:** 388–395.

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, et al. 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci* **103:** 2869–2873.

Gao LY, Guo S, McLaughlin B, Morisaki H, Engel JN, Brown EJ. 2004. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* **53:** 1677–1693.

Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, et al. 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci* **100:** 7877–7882.

Gautam US, Sikri K, Vashist A, Singh V, Tyagi JS. 2014. Essentiality of DevR/DosR interaction with SigA for the dormancy survival program in *Mycobacterium tuberculosis*. *J Bacteriol* **196:** 790–799.

Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V. 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* **1:** e5.

Harboe M, Nagai S. 1984. MPB70, a unique antigen of *Mycobacterium bovis* BCG. *Am Rev Respir Dis* **129:** 444–452.

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:** e1000988.

Isaza JP, Duque C, Gomez V, Robledo J, Barrera LF, Alzate JF. 2012. Whole genome shotgun sequencing of one Colombian clinical isolate of *Mycobacterium tuberculosis* reveals DosR regulon gene deletions. *FEMS Microbiol Lett* **330:** 113–120.

Krishnan N, Malaga W, Constant P, Caws M, Tran TH, Salmons J, Nguyen TN, Nguyen DB, Daffe M, Young DB, et al. 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLoS ONE* **6:** e23870.

Kubica T, Rusch-Gerdes S, Niemann S. 2003. *Mycobacterium bovis* subsp. *caprae* caused one-third of human *M. bovis*-associated tuberculosis cases reported in Germany between 1999 and 2001. *J Clin Microbiol* **41:** 3070–3077.

Kulaga S, Behr M, Musana K, Brinkman J, Menzies D, Brassard P, Kunimoto D, Tannenbaum TN, Thibert L, Joseph L, et al. 2002. Molecular epidemiology of tuberculosis in Montreal. *CMAJ* **167:** 353–354.

Layre E, Lee HJ, Young DC, Jezek MA, Buter J, Minnaard AJ, Annand JW, Fortune SM, Snider BB, Matsunaga I, et al. 2014. Molecular profiling of *Mycobacterium tuberculosis* identifies tuberculosinyl nucleoside products of the virulence-associated enzyme Rv3378c. *Proc Natl Acad Sci* **111:** 2978–2983.

Leung AS, Tran V, Wu Z, Yu X, Alexander DC, Gao GF, Zhu B, Liu J. 2008. Novel genome polymorphisms in BCG vaccine strains and impact on efficacy. *BMC Genomics* **9:** 413.

Lewis KN, Liao RL, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR. 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette–Guerin attenuation. *J Infect Dis* **187:** 117–123.

Leyten EM, Lin MY, Franken KL, Friggen AH, Prins C, van Meijgaarden KE, Voskuil MI, Weldingh K, Andersen P, Schoolnik GK, et al. 2006. Human T-cell responses to 25 novel antigens encoded by genes of the dormancy regulon of *Mycobacterium tuberculosis*. *Microbes Infect* **8:** 2052–2060.

Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178:** 1274–1282.

Manca C, Tsenova L, Barry CE III, Bergtold A, Freeman S, Haslett PA, Musser JM, Freedman VH, Kaplan G. 1999. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J Immunol* **162:** 6740–6746.

Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, Barry CE III, Freedman VH, Kaplan G. 2001. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-α/β. *Proc Natl Acad Sci* **98:** 5752–5757.

Manca C, Reed MB, Freeman S, Mathema B, Kreiswirth B, Barry CE III, Kaplan G. 2004. Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. *Infect Immun* **72:** 5511–5514.

Medina E, Ryan L, LaCourse R, North RJ. 2006. Superior virulence of Mycobacterium bovis over *Mycobacterium tuberculosis* (*Mtb*) for *Mtb*-resistant and *Mtb*-susceptible mice is manifest as an ability to cause extrapulmonary disease. *Tuberculosis (Edinb)* **86:** 20–27.

Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr MA. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J Infect Dis* **186:** 74–80.

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

Nedeltchev GG, Raghunand TR, Jassal MS, Lun S, Cheng QJ, Bishai WR. 2009. Extrapulmonary dissemination of *Mycobacterium bovis* but not *Mycobacterium tuberculosis* in a bronchoscopic rabbit model of cavitary tuberculosis. *Infect Immun* **77:** 598–603.

Nguyen D, Proulx JF, Westley J, Thibert L, Dery S, Behr MA. 2003. Tuberculosis in the Inuit community of Quebec, Canada. *Am J Respir Crit Care Med* **168:** 1353–1357.

* Niemann S, Supply P. 2014. Diversity and evolution of *Mycobacterium tuberculosis*: Moving to whole-genome-based approaches. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a021188.

Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR. 2003. Rv3133c/*dosR* is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol Microbiol* **48:** 833–843.

Pepperell CS, Granka JM, Alexander DC, Behr MA, Chui L, Gordon J, Guthrie JL, Jamieson FB, Langlois-Klassen D, Long R, et al. 2011. Dispersal of *Mycobacterium tuberculosis* via the Canadian fur trade. *Proc Natl Acad Sci* **108:** 6526–6531.

Pethe K, Swenson DL, Alonso S, Anderson J, Wang C, Russell DG. 2004. Isolation of *Mycobacterium tuberculosis* mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci* **101:** 13642–13647.

Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* **46:** 709–717.

Ramakrishnan L. 2012. Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol* **12:** 352–366.

Reed MB, Domenech P, Manca C, Su H, Barczak AK, Kreiswirth BN, Kaplan G, Barry CE III. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431:** 84–87.

Reed MB, Gagneux S, DeRiemer K, Small PM, Barry CE III. 2007. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J Bacteriol* **189:** 2583–2589.

Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, Domenech P, Zwerling A, Thibert L, Menzies D, et al. 2009. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J Clin Microbiol* **47:** 1119–1128.

Rehren G, Walters S, Fontan P, Smith I, Zarraga AM. 2007. Differential gene expression between *Mycobacterium bovis* and *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **87:** 347–359.

Reiling N, Homolka S, Walter K, Brandenburg J, Niwinski L, Ernst M, Herzmann C, Lange C, Diel R, Ehlers S, et al. 2013. Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *MBio* **4:** e00250-13.

Roupie V, Romano M, Zhang L, Korf H, Lin MY, Franken KL, Ottenhoff TH, Klein MR, Huygen K. 2007. Immunogenicity of eight dormancy regulon-encoded proteins of *Mycobacterium tuberculosis* in DNA-vaccinated and tuberculosis-infected mice. *Infect Immun* **75:** 941–949.

Said-Salim B, Mostowy S, Kristof AS, Behr MA. 2006. Mutations in *Mycobacterium tuberculosis* Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in *Mycobacterium bovis*. *Mol Microbiol* **62:** 1251–1263.

Salamon H, Kato-Maeda M, Small PM, Drenkow J, Gingeras TR. 2000. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. *Genome Res* **10:** 2044–2054.

Sassetti CM, Rubin EJ. 2003. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci* **100:** 12989–12994.

Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48:** 77–84.

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, et al. 2003. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: Insights into the phagosomal environment. *J Exp Med* **198:** 693–704.

Sinsimer D, Huet G, Manca C, Tsenova L, Koo MS, Kurepina N, Kana B, Mathema B, Marras SA, Kreiswirth BN, et al. 2008. The phenolic glycolipid of *Mycobacterium tuberculosis* differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect Immun* **76:** 3027–3036.

Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* **330:** 1703–1709.

Smith T, 1898. A comparative study of bovine tubercle bacilli and of human bacilli from sputum. *J Exp Med* **3:** 451–511.

Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, Johnson PD, Abdellah Z, Arrowsmith C, Chillingworth T, Churcher C, et al. 2008. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res* **18:** 729–741.

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, et al. 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat Genet* **45:** 172–179.

Thompson PJ, Cousins DV, Gow BL, Collins DM, Williamson BH, Dagnia HT. 1993. Seals, seal trainers, and mycobacterial infection. *Am Rev Respir Dis* **147:** 164–167.

Tsenova L, Ellison E, Harbacheuski R, Moreira AL, Kurepina N, Reed MB, Mathema B, Barry CE III, Kaplan G. 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:** 98–106.

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere YO, Aman K, Kato-Maeda M, Small PM. 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: Insights from genomic deletions in 100 strains. *Proc Natl Acad Sci* **101:** 4865–4870.

Turenne CY, Wallace R Jr, Behr MA. 2007. *Mycobacterium avium* in the postgenomic era. *Clin Microbiol Rev* **20:** 205–229.

Veyrier F, Pletzer D, Turenne C, Behr MA. 2009. Phylogenetic detection of horizontal gene transfer during the step-wise genesis of *Mycobacterium tuberculosis*. *BMC Evol Biol* **9:** 196.

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

Wang J, Behr MA. 2014. Building a better bacillus: The emergence of *Mycobacterium tuberculosis*. *Front Microbiol* **5:** 139.