

## The Origin and Evolution of *Mycobacterium tuberculosis*

Serge Mostowy<sup>a</sup>, Marcel A. Behr, MD<sup>a,b,\*</sup>

<sup>a</sup>McGill University Health Centre, 1650 Cedar Avenue, Montreal, QC H3G 1A4, Canada

<sup>b</sup>Division of Infectious Diseases and Medical Microbiology, A5-156, Montreal General Hospital, 1650 Cedar Avenue, Montreal, QC H3G 1A4, Canada

With tuberculosis (TB) having plagued mankind for centuries, there can be no doubt that *Mycobacterium tuberculosis*, the causative agent of human TB, has been successful in adapting for human infection. *M. tuberculosis* belongs to the *Mycobacterium tuberculosis* complex (MTC), itself comprised of bacterial agents responsible for TB or TB-like disease. Members of the MTC are known to infect mammalian hosts, and the extent and consequence of this infection is gaining greater recognition in part because of the availability of diagnostic tools to classify specific isolates appropriately. This article introduces the tools and terminology used for this classification and illustrates their utility by discussing work from independent laboratories that have established a genome-based phylogeny for the MTC [1–5]. Next, it considers the use of these markers to distinguish atypical isolates not conforming to attributes of traditional MTC members [6,7]. Finally, it discusses the current genomic evidence regarding the origin and evolution of *M. tuberculosis* in the context of its relevance for TB control in humans and other mammalian hosts.

### Characteristics of the *Mycobacterium tuberculosis* complex

The MTC consists of bacteria that genetically share identical 16S rRNA sequence and greater than 99.9% nucleotide identity. *M. tuberculosis*, *M. africanum*, *M. microti*, and *M. bovis* have been regarded as the four traditional species of the MTC, although the extent of MTC speciation is not yet resolved. In this article, MTC organisms are referred to as members, and the nomenclature provided in the most recent literature is used.

Members characteristically differ in their host range, epidemiology, clinical presentation in humans, and laboratory phenotype, although little is known about these differences or why these differences have evolved. The human form (*M. tuberculosis sensu stricto*) and the bovine form (*M. bovis*) have been nominally distinct for more than a century; other members have been identified more recently (Table 1). The members classically were described by their biochemical properties or by targeting their specific genetic regions. Genomic insights now show a new approach to MTC speciation outside the scope of these more traditional tools [8].

### Genetic resources to study the *Mycobacterium tuberculosis* complex

With the availability of complete sequence information, several methodologies have developed to understand the MTC genetically. These methods can

---

\* Corresponding author. Division of Infectious Diseases and Medical Microbiology, A5-156, Montreal General Hospital, 1650 Cedar Avenue, Montreal, QC H3G 1A4, Canada.

E-mail address: marcel.behr@mcgill.ca (M.A. Behr).

Table 1  
*Mycobacterium tuberculosis* complex members

| MTC member                                     | Natural host | Virulence |            |        | Unique attribute   |
|--|--------------|-----------|------------|--------|--|
|  |              | Mouse     | Guinea pig | Rabbit |  |
| <i>M. canettii</i>                             | Human?       | +         | +          | –      | Most ancestral recognized MTC member, anecdotal isolation                    |
| <i>M. tuberculosis</i>                         | Human        | +         | +          | –      | Predominant cause of human TB  |
| <sup>a</sup> <i>M. africanum</i> subtype II    | Human        | +         | +          | –      | Reclassified as atypical <i>M. tuberculosis</i>                              |
| <sup>b</sup> <i>M. africanum</i> subtype I (a) | Human?       | +         | +          | –      | Rarely isolated  |
| <sup>c</sup> <i>M. africanum</i> subtype I (b) | Human?       | +         | +          | –      | Phenotypically heterogeneous   |
| <i>M. pinnipedii</i>                           | Pinnipeds    | +         | +          | +      | Closely related to <i>M. microti</i>   |
| <i>M. microti</i>                              | Vole         | –         | –          | –      | Attenuated, used as live vaccine in humans                                   |
| Dassie bacillus                                | Dassie       | –         | –          | –      | More attenuated than <i>M. microti</i>                                       |
| <i>M. caprae</i>                               | Goat         | +         | +          | +      | Only described in Europe?  |
| <i>M. bovis</i>                                | Cow          | +         | +          | +      | Dynamic pathogen with wildlife reservoirs?                                   |
| <i>M. bovis</i> BCG                            | None         | –         | –          | –      | Family of laboratory adapted strains of <i>M. bovis</i> used as live vaccine |

Abbreviations: +, animal typically succumbs to infection; –, animal typically survives infection.

<sup>a</sup> Although previously suggested as a unique member of the MTC, *M. africanum* subtype II isolates cannot be genomically distinguished from *M. tuberculosis* [7]; and throughout this article, *M. africanum* subtype II is included within the *M. tuberculosis* lineage.

<sup>b</sup> Refers to the genotype '(a)' of *M. africanum* subtype I having deleted RD9, but not RD7, RD8, and RD10.

<sup>c</sup> Refers to the genotype '(b)' of *M. africanum* subtype I having deleted RD9, RD7, RD8, and RD10.

Data from Refs. [38,39,42,43,62,63].

be categorized as genetically fast or slow and as having phenotypic consequences or not. Each methodology has advantages and disadvantages. Whereas each methodology has proven useful, a tool is only as informative as the question toward which it is applied. Responsible contributions ideally should draw information from all available typing methods to conclude with the most parsimonious scenario.

#### Fingerprinting patterns

The use of DNA fingerprinting patterns, in which samples are genotyped by restriction-fragment-length polymorphisms using genetic attributes specific to the MTC as markers, has proven valuable for tracking MTC disease [9]. Molecular epidemiologic markers used include the MTC-specific insertion sequence IS6110 [10], polymorphic glycine- and cytosine-rich sequences [11], the direct-repeat region [12], spacer-oligonucleotide typing (spoligotyping) [13,14], and variable-number tandem repeats of genetic elements termed mycobacterial interspersed repetitive units [15,16]. Although these genetic markers are known to mutate at rates suitable for tracing a chain of disease transmission, their patterns of change are potentially too common to act as reliable markers over longer periods of evolutionary time. Therefore, they do not seem to be reliable for phylogenetic studies and speciation of clinical isolates.

#### Sequenced genomes

A wealth of genomic insight for the Mycobacterium genus is available through whole-genome sequence information for several species (Table 2), including six entire MTC genomic sequences completed or in progress. These are *M. tuberculosis* H37Rv [17], *M. tuberculosis* CDC1551 [18], *M. tuberculosis* 210 [18a], *M. microti* OV254 [19], *M. bovis* 2122 [20], and *M. bovis* bacille Calmette-Guerin (BCG) Pasteur [20a]. Mycobacteria sequenced or being sequenced outside the MTC include *M. leprae* [21], *M. ulcerans* [22], *M. avium* 104 [51], *M. paratuberculosis* K10 [22a], *M. marinum* [22b], and the relatively fast-growing *M. smegmatis* MC2 155 [18a]. Even the most distant of these sequenced mycobacterial genomes are minimally related by 60% DNA/DNA homology, and comparative genomic analysis has shown that gene loss is a significant part of the ongoing evolution of the slow-growing mycobacterial pathogens [23].

#### Single-nucleotide polymorphisms

Single-nucleotide polymorphisms (SNPs) can result in a silent amino acid substitution in which the protein coding sequence remains unchanged (synonymous) or can alter the protein-coding sequence (nonsynonymous) and hence act as a substrate for

Table 2  
Overview of mycobacterial genome sequencing projects

| Species                              | First author/date [reference]            | Genome size (base pairs) | No. of protein-coding genes | G + C nucleotide content (%) | Insight from sequencing project  |
|--------------------------------------|--|--------------------------|-----------------------------|------------------------------|--|
| <i>M. tuberculosis</i> H37Rv         | Cole, 1998 [17]                          | 4,411,532                | 3995                        | 65.6                         | First sequenced mycobacterial genome   |
| <i>M. tuberculosis</i> CDC1551       | Fleischmann, 2002 [18]                   | 4,403,836                | 4249                        | 65.6                         | Polymorphisms among <i>M. tuberculosis</i> strains more extensive than initially anticipated |
| <i>M. tuberculosis</i> 210           | The Institute for Genomic Research [18a] | 4,400,000 <sup>a</sup>   | NA                          | NA                           | Describes hyper-virulence of 'Beijing' strain family?  |
| <i>M. bovis</i> 2122                 | Garnier, 2003 [20]                       | 4,345,492                | 3951                        | 65.6                         | <i>M. bovis</i> is derivative compared to <i>M. tuberculosis</i>                             |
| <i>M. microti</i>                    | Brodin, 2002 [19]                        | 4,400,000 <sup>a</sup>   | NA                          | 64.0 <sup>a</sup>            | Loss of RD1 contributed to attenuation of <i>M. microti</i>                                  |
| <i>M. bovis</i> BCG Pasteur          | Sanger Institute [20a]                   | 4,083,000 <sup>a</sup>   | NA                          | NA                           | Describes live vaccine administered to humans  |
| <i>M. marinum</i>                    | Sanger Institute [22b]                   | 6,636,827                | NA                          | 65.73%                       | Describes causative agent of TB-like disease in fish and 'fish tank granuloma' of humans     |
| <i>M. ulcerans</i>                   | Stinear, 2004 [22]                       | 6,032,000 <sup>a</sup>   | NA                          | 65.0 <sup>a</sup>            | Plasmid-encoded toxin responsible for Buruli ulcer   |
| <i>M. leprae</i>                     | Cole, 2001 [21]                          | 3,268,203                | 1604                        | 57.8                         | Massive gene decay in the leprosy bacillus   |
| <i>M. avium avium</i> 104            | Semret, 2004 [51]                        | 5,475,491                | 4480                        | 69                           | Extensive genomic polymorphism among <i>M. avium</i> sub-species                             |
| <i>M. avium paratuberculosis</i> K10 | GenBank [22a]                            | 4,829,781                | 4350                        | 69.3                         | Describes causative agent of Johne's disease in cattle                                       |
| <i>M. smegmatis</i> MC2 155          | The Institute for Genomic Research [18a] | 7,000,000 <sup>a</sup>   | NA                          | NA                           | Describes fast growing, model organism for mycobacteria                                      |

Mycobacteria are listed in the order of 16S rRNA sequence relatedness to *M. tuberculosis* [64].

Abbreviations: G + C, guanine plus cytosine; NA, not available.

<sup>a</sup> Parameter estimates.

evolutionary selection. Both types of mutations have been applied toward differentiation and diagnostics of MTC members [24–26]. In a landmark study, a first sequence analysis of MTC isolates revealed that allelic polymorphism is impressively rare, occurring on the order of 1 in 10,000 base pairs (bp), suggesting that the complex could be dated to about 15,000 to 20,000 years of age [24].

Genomic comparison of multiple sequenced MTC strains has made possible the identification of SNP markers for studies of evolution, pathogenesis, and epidemiology in clinical *M. tuberculosis* [27] and *M. bovis* [20], supporting a clonal evolution of the MTC without detectable lateral gene exchange. The ratio of SNP types within a genome can act as a molecular clock [28] in which the high ratio of non-synonymous to synonymous mutations across coding sequences within MTC genomes suggests a recent divergence of *M. bovis* and *M. tuberculosis* [20].

#### *Large-sequence polymorphisms*

Unlike other mycobacterial species in which horizontal gene transfer has been demonstrated [22], this mode of generating genomic diversity has not been observed for the obligately intracellular MTC. Genomic comparisons for the MTC reveal a prominent role of genomic deletions relative to the sequenced strains of *M. tuberculosis*. For example, the complete genome sequence of *M. bovis* 2122 contains 66,037 bp less than *M. tuberculosis* H37Rv, and no genomic region exclusive to *M. bovis* but consistently absent from *M. tuberculosis* has been detected [20].

To uncover deletions in nonsequenced strains efficiently, one can hybridize whole genomic DNA of a MTC member against a spotted array [29] or an Affymetrix GeneChip (Santa Clara, California) [30] representative of the entire *M. tuberculosis* H37Rv genome [17]. Regions of the prototype strain that seem to be absent from the test strain are then confirmed by performing polymerase chain reaction (PCR) with primers approximating the deleted region, to amplify across the deletion. This amplicon is then sequenced to define the deletion point precisely. Isolates are said to share a genomic deletion when sequencing shows the deletion occurs in different isolates at exactly the same cut point [2]. Because independently arisen chromosomal rearrangements sometimes involve the same strategically located elements, only upon exact description of the specific genomic event (ie, genomic location within a reference strain) can one determine with confidence whether genomic deletions behave as unidirectional

event polymorphisms (UEPs). These UEPs, which represent one-time events in the evolution of the organism, can serve as robust markers of clonal organisms, useful for determining phylogenetic classification.

A valuable use of these genomic deletions pertains to their application in defining specific MTC members and accurately assessing their prevalence in clinical specimens [8,26]. Unlike biochemical testing, which for individual results had imperfect sensitivity and specificity, the use of genomic events in these studies provided unambiguous classification, thereby simplifying the process considerably. To explore the basis for the previously observed biochemical attributes used for MTC speciation, an association was sought between deleted sequences and phenotypic results for isolates assigned as *M. africanum*. Results indicate that convergent biochemical profiles can be independently obtained in different MTC members. For instance, organisms presenting the distinct deletion profile of *M. africanum* and *M. bovis* can manifest the same biochemically based profile [7]. These results confirm the limitations of biochemically derived speciation and, by extension, challenge the taxonomic divisions currently in place for classifying members of the MTC.

Beyond diagnostics, different studies have all supported the potential value of most MTC genomic deletions (with the exception of mycobacteriophage DNA) as evolutionary markers. In separate genomic studies of *M. bovis* BCG vaccine strains, it has been documented that BCG-specific deletions superimpose perfectly on the historical record [3,29]. In studies of genomic deletions within clinical isolates of *M. tuberculosis* [31,32], mycobacterial clones shared the same genomic deletions, again suggesting that deletions can be used to reconstruct phylogenetic trees. Finally, a recent analysis of 100 *M. tuberculosis* clones from San Francisco has again confirmed that these deletions are UEPs [5], and therefore genomic deletions can effectively brand a particular clone [33]. A practical use of this approach will be to provide a secure genomic definition for prominent strains, such as the Beijing [34] and Manila [35] strains of *M. tuberculosis*, and to assess their prevalence through space and time.

#### **Genomic deletions and the origin and evolution of the *Mycobacterium tuberculosis* complex**

The long-recognized presence of a human TB bacillus and a closely related bovine form has given rise to speculation that TB originally came to humans as a zoonotic infection from cattle [36]. In

retrospect, this view was probably influenced by the types of *M. tuberculosis* isolates available for study, biased toward hosts (namely cattle and humans) for which a diagnosis of TB would lead to microbiologic investigation. To explore the evolutionary relationship of members of the MTC, the presence or absence of deletions was tested within complex isolates derived from different hosts and from isolates in various geographic locales [1,2]. Analysis revealed a stepwise accumulation of genomic deletions among isolates interrogated, but the distribution of genomic deletions argued against present-day *M. bovis* as the evolutionary precursor of *M. tuberculosis*, making it improbable that human TB originated with the domestication of cattle. Instead, a number of MTC organisms, both long established and more recently described, present genomic profiles that seem to be intermediate between the ancestor of modern *M. tuberculosis* and that of present-day *M. bovis*.

The availability of improved laboratory tools has facilitated the description of a number of novel variants of the MTC, including *M. canettii* [37,38], *M. caprae* [39,40], *M. pinnipedii* [41,42], and the dassie bacillus [2,43] (Table 1).

Before these tools were available, MTC members had presented a well-established host range, presumably biased by expectations: *M. tuberculosis* (and sometimes *M. africanum*) is classically isolated from humans, *M. microti* from voles, and *M. bovis* from a broad range of hosts including (but not limited to) cows. More careful study, however, has revealed a wider range of host animals. A practical issue arising from these studies involves the generally held belief that *M. bovis* infects an extensive range of animal species, including the badger, opossum, elk, cougar, and buffalo. Until recently, *M. caprae* and *M. pinnipedii* were considered to be forms of *M. bovis* [40,42]. Although *M. bovis* might be versatile enough to accommodate such a dynamic host range, the inclusion of such organisms probably overestimates the true host range of *M. bovis*. Detailed genomic analysis of isolates from unusual hosts is underway, with the expectation that results will continue to challenge accepted notions of MTC speciation and taxonomy [2].

Just as genomic deletions have proven unique to isolates of *M. tuberculosis* affecting only human hosts [5,30], deletions unique to these other MTC members permit resolution of their phylogenetic situation (Fig. 1) [1,2]. A first observation from this distribution of organisms is that the MTC affects a number of undomesticated and domesticated mammals, both terrestrial and aquatic. *M. marinum* can be

considered a piscine/amphibian mycobacterium [44] and *M. avium* an avian mycobacterium [45]. The MTC is a relatively broad-ranging mammalian mycobacterium. Organisms of the four most ancestral lineages (*M. canettii*, *M. tuberculosis*, and both genotypes of *M. africanum* subtype I) have been cultured predominantly from humans. Because isolation of *M. canettii* has been extremely rare [37,38], an unrecognized nonhuman reservoir might exist, with humans representing an accidental or circumstantial host. The next three MTC lineages (*M. microti*, *M. pinnipedii*, and the dassie bacillus) affect undomesticated mammals irrespective of their geographic location. *M. microti*, first identified in Europe, infects the field vole [19], the dassie bacillus infects the dassie and the surikat from Africa [6,43], and *M. pinnipedii* globally infects a variety of seals and sea lions from Oceania to South America [4,42]. Finally, more derivative forms of the MTC are seen in goats (*M. caprae*) and subsequently cattle (classic *M. bovis*), suggesting that the organism was introduced into livestock in the order of their domestication. More recently, spillover of *M. bovis* from farms has been seen in the case of badgers in the United Kingdom [46] and the brushtail opossum in New Zealand [46a]. Far from suggesting that human TB originated with livestock, the genomic record suggests that, directly or indirectly, humans were responsible for bringing MTC to the farm, with secondary foci of spread now observed in animals associated with this setting.

### Geographic, chronologic, and ecologic origins of the *Mycobacterium tuberculosis* complex

An absolute chronology of the TB epidemic is difficult to discern by genomic deletions, because they do not evolve on a predictable time scale. The genetic record can potentially point to the geographic origins, however, because the ancestral form *M. canettii* [1,4,25] has been isolated only in persons living in Africa [37,38]. If the origins of human TB are situated in the same the continent as the origins of man, it is conceivable that the organism spread with humans during the paleomigration, explaining the presence of MTC DNA in 5000-year-old samples from Egypt [47] and pre-Columbian mummies from Ecuador [48]. Because more derivative organisms are found in hosts domesticated 10,000 to 12,000 years ago, these clues suggest that the organism accessed humans before that era and subsequently spread to other hosts, either from man directly or through an unrecognized vector.

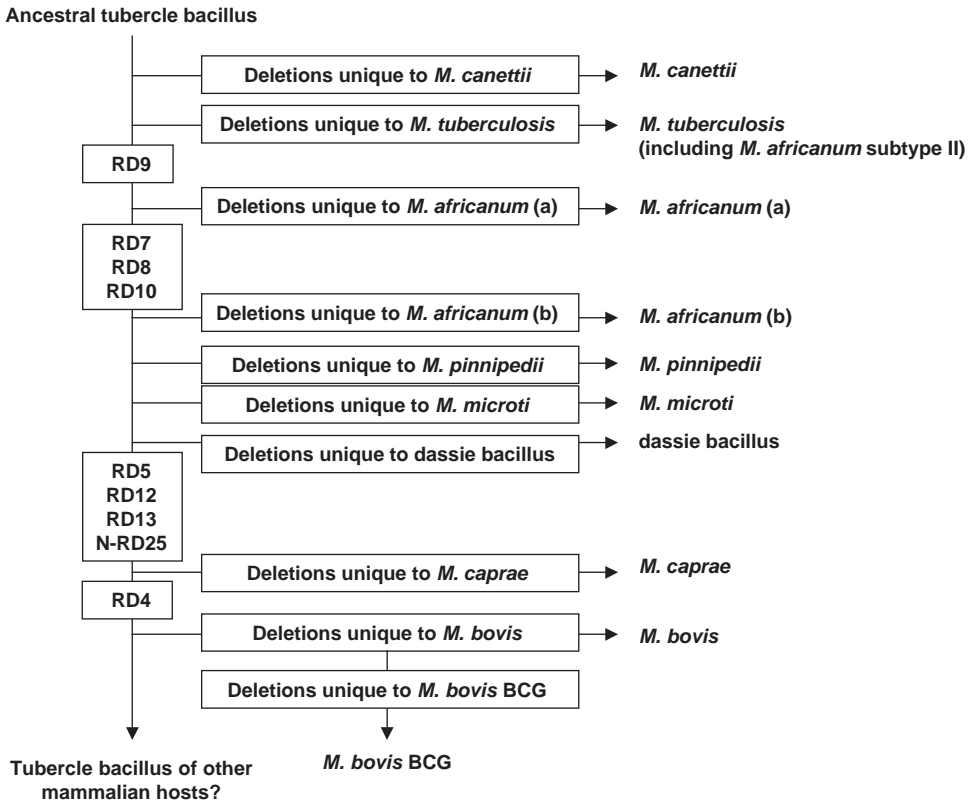


Fig. 1. Deletion-based phylogeny of the MTC based on deleted regions demonstrated through genomic analysis. The vertical axis presents the stepwise accumulation of unidirectional evolutionary polymorphisms (RDs and N-RD) previously characterized among members of the MTC [1,2]. Clustered along each horizontal axis are organisms for which one or more genomic deletions specific to this evolutionary branch have been revealed in supporting citations [4,6,7,19,32,58] and unpublished observations. N-RD, new deletions. (Serge Mostowy, Marcel Behr, MD, unpublished data, 2005.)

Using deletions to assign directionality to the MTC phylogeny, one can employ sequence-based analysis to estimate the chronology of this scenario and refine the previous nucleotide-based analysis that suggested a 20,000-year divergence between *M. tuberculosis* and *M. bovis* [24]. Another approach to date these events uses testing for genomic regions directly on paleo-DNA samples [49] (Mostowy et al, unpublished data). Because these samples can be carbon dated independently, it is possible to provide genomic signatures for samples of human or non-human provenance and to derive minimal estimates for the ages of genomic events portrayed in Fig. 1.

Turning to the ecologic origins of the MTC, a livestock source seems to be unlikely, because human forms diverged before the modern caprine and bovine forms. Although it is attractive to consider another mammalian host as the ancestral niche, observations for other mycobacteria suggest that nonmammalian

reservoirs such as plant or insects deserve consideration [50]. With the ability to test rapidly for genomic deletions by PCR, one can test putative wildlife reservoirs for variants of the MTC to find the natural host of relatively ancestral forms such as *M. canettii*.

#### What is being deleted from the *Mycobacterium tuberculosis* complex?

When compared with other bacterial species, members of the MTC present relatively little genomic diversity. Estimates of large-sequence polymorphism diversity among MTC members [32], in agreement with similar conclusions drawn from estimates of SNPs [25], have been consistently described as low in comparison with other microbes. Nonetheless,



genomic flexibility does seem to exist within the MTC for specific host adaptation, and a similar potential is beginning to reveal itself among other mycobacterial complexes. Although the amount of diversity revealed within the *Mycobacterium avium* complex is 10-fold more than that of the MTC [51], host-specific genomic contents are being observed there as well (*M. Semret* and *M. Behr*, unpublished data). Taken together, these data highlight a comparative genomics approach to understanding an evolutionary potential of mycobacterium pathogenesis, in which genomic content can suggest DNA features for host-specific adaptation.

#### *Genomic deletions and virulence*

With host-specific MTC extending beyond a domesticated setting, how TB spreads from host to host is difficult to ascertain. From what is known about humans and cows with TB, it is reasonable to expect that transmission would occur through aerosols from a diseased animal to a contact animal. If so, a requisite of host adaptation is a certain degree of virulence in that host. Although greater virulence might facilitate transmission, too much virulence could be detrimental if host mortality is excessive or if the organism causes an invasive form of TB that is generally nontransmissible (such as TB meningitis). Thus, optimal transmissibility requires some degree of virulence (ie, pulmonary pathology) but a sufficiently contained disease process to generate the agents required for spread (ie, aerosols).

Support for this notion comes from studying the content of the genomic regions that have been deleted in different MTC members. A general observation is that although each deletion noted in Fig. 1 is unique to the bp, both genomic regions and the predicted function of implicated genes are nonrandom. Several regions of difference (RD) seem to be prone to genomic deletion, with different specific deletions having occurred near the same locus [6,7]. Most prominent among these is RD1, a series of nine genes implicated in the attenuation of *M. bovis* BCG strains, that has suffered three distinct genomic deletions. Although this confluence of deletions might point to genetic instability at this locus, a study of 100 circulating *M. tuberculosis* clones documenting 176 deletions failed to detect a single deletion in this region, arguing against an inherently elevated mutation rate [32]. The absence of RD1 was first observed for BCG vaccines [51a]; subsequently, targeted disruption of RD1 from *M. tuberculosis* was shown to decrease bacterial replication and educe pulmonary pathology in a mouse model [52,53]. More recently,

*M. microti* and the dassie bacillus also were shown to have deletions in the RD1 region; notably, both have been characterized as having low virulence in animal models [19,43]. More detailed analysis of the RD1 region revealed it contains genes encoding a novel secretion system of two important secreted antigens (CFP-10, ESAT-6) [53,54]. Presumably a metabolically expensive process, the loss of this region in BCG was probably advantageous with no selective pressure in favor of synthesizing and secreting antigenic proteins in vitro. Although the independent loss of CFP-10 and ESAT-6 in *M. microti* and the dassie bacillus explains their attenuated phenotype, the selective pressures for their deletion in vivo are more speculative.

Given the documented impact of the RD1 region on virulence, the observation of its deletion in both the vole and dassie hosts is provocative. Nothing evident points to why the genetically distant vole (a rodent) and dassie (closely related to the elephant) would share some unique immunologic susceptibility. A more likely explanation might involve social conditions and transmissibility [55], given that voles and dassies congregate in high-density underground communities, unlike other MTC hosts that predominantly live aboveground in open-air conditions. Such congregate living settings would be extremely favorable for TB transmission, and an organism of lesser virulence might be successful in such burrowing hosts so long as host populations remain sufficiently abundant [56]. Conversely, conditions for transmission aboveground between goats and seals are less ideal and would probably require an organism of relatively high virulence to optimize transmissibility.

#### *Summary of catalogued deletions*

From Fig. 1, deletions represented along the vertical line of the phylogeny preceded spread of the bacillus into new hosts; those along the horizontal axes arose during coevolution of the organism with new hosts. Evidence supporting this scenario is that organisms lacking RD7, RD8, RD9, and RD10 have been recovered from the entire MTC host range, whereas the precise deletions seen along the horizontal lineages are observed in only restricted, one-host settings. To derive a scenario for the loss of genomic regions in vivo, genes lost on the vertical axis and those lost along horizontal lineages can be directly compared, pointing to nonrandom distinction between the functional classification of these two sets of deleted genes. Although such studies generate hypotheses regarding the evolution of MTC members in different hosts, these studies are naturally biased

to successful pathogenic strains, as opposed to those that caused infection but not disease. In this light, an examination of BCG vaccines provides telling insights into MTC evolution when the selective pressures for virulence were absent in the host and limited to the culture media employed. Remarkably, the sheer volume of genomic disturbance incurred by BCG vaccines during a half-century of laboratory evolution is on the same order as observed among virulent *M. tuberculosis* isolates that have been circulating through the human hosts through millennia [3]. Here, the preponderant message is that BCG evolution has favored the elimination of regulatory elements and antigens. These results reinforce the notion that the capacity to engage the host immune system is selected for during *in vivo* conditions, consistent with MTC members being professional pathogens [57]. More practically, the absence of numerous antigens from BCG vaccines may in part explain its limitations as an immunizing agent [58].

#### Summary and concluding thoughts: lessons for tuberculosis control

*M. tuberculosis* has probably been with humans for millennia and thus probably became adapted to humans during times of low population density and predominantly outdoor living. Unlike diseases such as HIV that rapidly spread in epidemic form soon after introduction into humans, the TB epidemic peaked in Western Europe during the nineteenth century and has arguably yet to peak in certain parts of the world. Although it is possible that the organism has evolved toward greater virulence in recent centuries, it seems more probable that that social changes brought about by industrialization were paramount in altering the transmission dynamics. This argument would also apply to *M. bovis*, in which an organism that evolved to persist in free-ranging cattle would predictably wreak havoc in the environment provided by modern-day factory farming. Finally, whereas tuberculous animals in the wild might normally succumb to predation, the increasing protection of these hosts in wildlife refuges and zoos should provide a greater chance for progression to TB, as attested to by reports of TB in farmed deer [59], zoo tigers [60], and seal-trainers [61]. Although these cases are generally rare, these anecdotes do serve notice that MTC has the capacity to adapt to the immunologic environment it engages and suggest that nonhuman reservoirs may become pertinent to human TB control.

#### References

- [1] Brosch R, Gordon SV, Marmiesse M, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 2002;99:3684–9.
- [2] Mostowy S, Cousins D, Brinkman J, et al. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J Infect Dis* 2002;186:74–80.
- [3] Mostowy S, Tsolaki AG, Small PM, et al. The *in vitro* evolution of BCG vaccines. *Vaccine* 2003;21:4270–4.
- [4] Marmiesse M, Brodin P, Buchrieser C, et al. Macroarray and bioinformatic analyses reveal mycobacterial ‘core’ genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiol* 2004;150:483–96.
- [5] Hirsh AE, Tsolaki AG, DeRiemer K, et al. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A* 2004;101:4871–6.
- [6] Mostowy S, Cousins D, Behr MA. Genomic interrogation of the *dassie* bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. *J Bacteriol* 2004;186:104–9.
- [7] Mostowy S, Onipede A, Gagneux S, et al. Genomic analysis distinguishes *Mycobacterium africanum*. *J Clin Microbiol* 2004;42:3594–9.
- [8] Parsons LM, Brosch R, Cole ST, et al. Rapid and simple approach for identification of *Mycobacterium tuberculosis* complex isolates by PCR-based genomic deletion analysis. *J Clin Microbiol* 2002;40:2339–45.
- [9] Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med* 2003;349:1149–56.
- [10] van Embden JD, Cave MD, Crawford JT, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406–9.
- [11] Ross BC, Raios K, Jackson K, et al. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J Clin Microbiol* 1992;30:942–6.
- [12] Hermans PW, Van Soolingen D, Bik EM, et al. Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* 1991;59:2695–705.
- [13] Aranaz A, Liebana E, Mateos A, et al. Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. *J Clin Microbiol* 1996;34:2734–40.
- [14] Kamerbeek J, Schouls L, Kolk A, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;35:907–14.
- [15] Frothingham R, Meeker-O’Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 1998;144:1189–96.



- [16] Mazars E, Lesjean S, Banuls AL, et al. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci U S A* 2001;98:1901–6.
- [17] Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;393:537–44.
- [18] Fleischmann RD, Alland D, Eisen JA, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 2002;184:5479–90.
- [18a] The Institute for Genomic Research. (Preliminary sequence data.) Available at: <http://www.tigr.org>. Accessed March 1, 2005.
- [19] Brodin P, Eiglmeier K, Marmiesse M, et al. Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun* 2002;70:5568–78.
- [20] Garnier T, Eiglmeier K, Camus JC, et al. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* 2003;100:7877–82.
- [20a] The Sanger Institute. Available at: <ftp://ftp.sanger.ac.uk/pub/pathogens/mb/>. Accessed March 1, 2005.
- [21] Cole ST, Eiglmeier K, Parkhill J, et al. Massive gene decay in the leprosy bacillus. *Nature* 2001;409:1007–11.
- [22] Stinear TP, Mve-Obiang A, Small PL, et al. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A* 2004;101:1345–9.
- [22a] GenBank. Accession no NC\_002944. Accessed March 1, 2005.
- [22b] The Sanger Institute. Available at: <ftp://ftp.sanger.ac.uk/pub/pathogens/mm/>. Accessed March 1, 2005.
- [23] Brosch R, Pym AS, Gordon SV, et al. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol* 2001;9:452–8.
- [24] Sreevatsan S, Pan X, Stockbauer KE, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 1997;94:9869–74.
- [25] Gutacker MM, Smoot JC, Migliaccio CA, et al. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 2002;162:1533–43.
- [26] Huard RC, Oliveira Lazzarini LC, Butler WR, et al. PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *J Clin Microbiol* 2003;41:1637–50.
- [27] Alland D, Whittam TS, Murray MB, et al. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J Bacteriol* 2003;185:3392–9.
- [28] Kimura M. *The neutral theory of molecular evolution*. Cambridge (UK): Cambridge University Press; 1983.
- [29] Behr MA, Wilson MA, Gill WP, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999;284:1520–3.
- [30] Salamon H, Kato-Maeda M, Small PM, et al. Detection of deleted genomic DNA using a semi-automated computational analysis of GeneChip data. *Genome Res* 2000;10:2044–54.
- [31] Kato-Maeda M, Rhee JT, Gingeras TR, et al. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* 2001;11:547–54.
- [32] Tsolaki AG, Hirsh AE, DeRiemer K, et al. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci U S A* 2004;101:4865–70.
- [33] Nguyen D, Brassard P, Menzies D, et al. Genomic characterization of an endemic *Mycobacterium tuberculosis* strain: evolutionary and epidemiologic implications. *J Clin Microbiol* 2004;42:2573–80.
- [34] Bifani PJ, Mathema B, Kurepina NE, et al. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol* 2002;10:45–52.
- [35] Douglas JT, Qian L, Montoya JC, et al. Characterization of the Manila family of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2003;41:2723–6.
- [36] Stead WW, Eisenach KD, Cave MD, et al. When did *Mycobacterium tuberculosis* infection first occur in the New World? An important question with public health implications. *Am J Respir Crit Care Med* 1995;151:1267–8.
- [37] Pfyffer GE, Auckenthaler R, van Embden JD, et al. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. *Emerg Infect Dis* 1998;4:631–4.
- [38] Van Soolingen D, Hoogenboezem T, de Haas PE, et al. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, *Canetti*: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol* 1997;47:1236–45.
- [39] Aranaz A, Liebana E, Gomez-Mampaso E, et al. *Mycobacterium tuberculosis* subsp. *caprae* subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. *Int J Syst Bacteriol* 1999;49(Pt 3):1263–73.
- [40] Aranaz A, Cousins D, Mateos A, et al. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int J Syst Evol Microbiol* 2003;53:1785–9.
- [41] Cousins DV, Williams SN, Reuter R, et al. Tuberculosis in wild seals and characterisation of the seal bacillus. *Aust Vet J* 1993;70:92–7.
- [42] Cousins DV, Bastida R, Cataldi A, et al. Tuberculosis

- in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol* 2003; 53:1305–14.
- [43] Cousins DV, Peet RL, Gaynor WT, et al. Tuberculosis in imported hyrax (*Procavia capensis*) caused by an unusual variant belonging to the *Mycobacterium tuberculosis* complex. *Vet Microbiol* 1994;42:135–45.
- [44] Decostere A, Hermans K, Haesebrouck F. Piscine mycobacteriosis: a literature review covering the agent and the disease it causes in fish and humans. *Vet Microbiol* 2004;99:159–66.
- [45] Thoen CO, Karlson AG, Himes EM. Mycobacterial infections in animals. *Rev Infect Dis* 1981;3:960–72.
- [46] Donnelly CA, Woodroffe R, Cox DR, et al. Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature* 2003;426:834–7.
- [46a] Morris RS, Pfeiffer DU. Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. *N Z Vet J* 1995;43(7):256–65.
- [47] Zink AR, Sola C, Reischl U, et al. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* 2003;41:359–67.
- [48] Konomi N, Lebwohl E, Mowbray K, et al. Detection of mycobacterial DNA in Andean mummies. *J Clin Microbiol* 2002;40:4738–40.
- [49] Donoghue HD, Spigelman M, Greenblatt CL, et al. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. *Lancet Infect Dis* 2004;4: 584–92.
- [50] Marsollier L, Stinear T, Aubry J, et al. Aquatic plants stimulate the growth of and biofilm formation by *Mycobacterium ulcerans* in axenic culture and harbor these bacteria in the environment. *Appl Environ Microbiol* 2004;70:1097–103.
- [51] Semret M, Zhai G, Mostowy S, et al. Extensive genomic polymorphism within *Mycobacterium avium*. *J Bacteriol* 2004;186:6332–4.
- [51a] Mahairas GG, Sabo PJ, Hickey MJ, et al. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 1996;178:1274–82.
- [52] Lewis KN, Liao R, Guinn KM, et al. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. *J Infect Dis* 2003;187:117–23.
- [53] Hsu T, Hingley-Wilson SM, Chen B, et al. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc Natl Acad Sci U S A* 2003;100:12420–5.
- [54] Pym AS, Brodin P, Majlessi L, et al. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat Med* 2003;9:533–9.
- [55] Boots M, Hudson PJ, Sasaki A. Large shifts in pathogen virulence relate to host population structure. *Science* 2004;303:842–4.
- [56] Davis S, Begon M, De Bruyn L, et al. Predictive thresholds for plague in Kazakhstan. *Science* 2004; 304:736–8.
- [57] Monack DM, Mueller A, Falkow S. persistent bacterial infections: the interface of the pathogen and the host immune system. *Nat Rev Microbiol* 2004;2:747–65.
- [58] Behr MA. BCG—different strains, different vaccines? *Lancet Infect Dis* 2002;2:86–92.
- [59] Griffin JFT, Chinn DN, Rodgers CR. Diagnostic strategies and outcomes on three New Zealand deer farms with severe outbreaks of bovine tuberculosis. *Tuberculosis* 2004;84:293–302.
- [60] Lantos A, Niemann S, Mezosi L, et al. Pulmonary tuberculosis due to *Mycobacterium bovis* subsp. *caprae* in captive Siberian tiger. *Emerg Infect Dis* 2003;9:1462–4.
- [61] Thompson PJ, Cousins DV, Gow BL, et al. Seals, seal trainers, and mycobacterial infection. *Am Rev Respir Dis* 1993;147:164–7.
- [62] Wagner JC, Buchanan G, Bokkenheuser V, et al. An acid fast bacillus isolated from the lungs of a cape hyrax. *Procavia capensis* (Pallus). *Nature* 1958;181: 284–5.
- [63] Dannenberg Jr AM. Pathogenesis of pulmonary *Mycobacterium bovis* infection: basic principles established by the rabbit model. *Tuberculosis (Edinb)* 2001;81:87–96.
- [64] Springer B, Stockman L, Teschner K, et al. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. *J Clin Microbiol* 1996;34:296–303.