Genomics and the evolution, pathogenesis, and diagnosis of tuberculosis

Joel D. Ernst,[1,2,3] Giraldina Trevejo-Nuñez,[1] and Niaz Banaiee[1]

[1]Department of Medicine, Division of Infectious Diseases, [2]Department of Pathology, and
[3]Department of Microbiology, New York University School of Medicine, New York, New York, USA.

Tuberculosis kills nearly 2 million people annually, and current approaches to tuberculosis control are expensive, have limited efficacy, and are vulnerable to being overcome by extensively drug-resistant strains of *Mycobacterium tuberculosis*. Determination of the genome sequence of *M. tuberculosis* has revolutionized tuberculosis research, contributed to major advances in the understanding of the evolution and pathogenesis of *M. tuberculosis*, and facilitated development of new diagnostic tests with increased specificity for tuberculosis. In this review, we describe some of the major progress in tuberculosis research that has resulted from knowledge of the genome sequence and note some of the problems that remain unsolved.

Evolution of the *Mycobacterium tuberculosis* complex

The *Mycobacterium tuberculosis* complex presently consists of seven species and subspecies that include *M. tuberculosis*, *M. canettii*, *M. africanum*, *M. bovis*, *M. microti*, *M. pinnipedii*, and *M. caprae*. This number is likely to increase as new genetic differences between strains of the existing members are identified. Although each species of the *M. tuberculosis* complex behaves as an ecotype, with preferred mammalian hosts (1), they are remarkable for their paucity of interspecies genetic variation. Members of the *M. tuberculosis* complex are characterized by more than 99.9% identity at the nucleotide level and an identical 16S rRNA sequence (2–6). Furthermore, there is little evidence for ongoing horizontal gene transfer between contemporary members of the *M. tuberculosis* complex (7–10). Taken together, these findings suggest that members of the *M. tuberculosis* complex are relatively young pathogens that are the clonal progeny of an ancestral strain that underwent an evolutionary bottleneck 20,000–35,000 years ago (3, 5, 6, 11). More recently, sequence analysis of six housekeeping genes revealed that human isolates of *M. canettii* from East Africa represent extant progenies of an ancestral species (termed *M. prototuberculosis*) from which the *M. tuberculosis* complex evolved (12). *M. canettii* housekeeping genes are mosaic, and their individual segments can be found in the other members of the *M. tuberculosis* complex, thus providing evidence for horizontal gene transfer among the ancestral strains. By analysis of synonymous nucleotide variation, it has now been estimated that ancestral tubercle bacilli may have existed as far back as 3 million years ago (13, 14). The same studies also suggest that the geographic restriction of *M. canettii* strains to East Africa indicates that tubercle bacilli originated in Africa and thus afflicted human ancestors much earlier than previously believed.

Such long-lasting association of tubercle bacilli with the hominids has undoubtedly shaped the evolution of *M. tuberculosis*. The outcome is a pathogen tha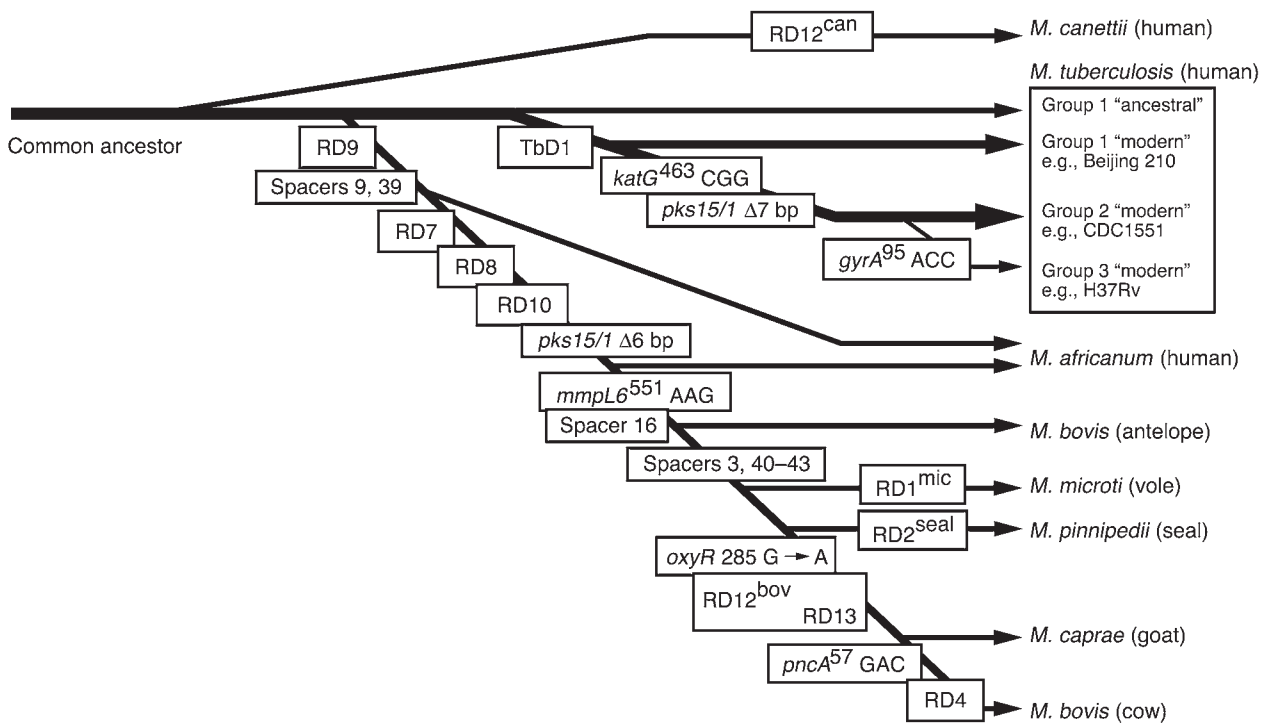t is transmitted efficiently and is capable of causing disease, yet it causes a latent infection in the majority of the individuals it infects. We speculate that the population structure of early humans provided selection pressure for these phenotypes. It was likely essential for *M. tuberculosis* to transmit efficiently when human tribes came in transient contact with each other. However, only those strains that caused chronic disease survived and continue to afflict us, as strains with a high rate of reactivation or acute virulence likely eliminated susceptible hosts within the population and caused their own extinction. While the oldest genetic lineage of *M. tuberculosis*, principal genetic group 1 (6), has retained a WT genotype at the polyketide synthase 15/1 (*pks* 15/1) locus, more recent lineages have undergone a 7-bp deletion in this locus and appear to be less virulent, as indicated by population studies and experiments in animal models (15–18). *M. tuberculosis* continues to adapt to selective pressures, as indicated by the ease of development of drug resistance.

Several studies have taken advantage of irreversible genetic events such as chromosomal deletions (also known as large sequence polymorphisms [LSPs]), together with analysis of single nucleotide polymorphisms (SNPs) and direct repeat content ("spoligotype") patterns, to decipher the phylogeny of the *M. tuberculosis* complex (8, 11, 19, 20). In this approach, the successive and unidirectional loss of DNA in representative strains reveals the order in which members of the complex descended from their ancient common ancestor (Figure 1) (11). The extant representatives of the common ancestor of the *M. tuberculosis* complex include the ancestral and modern strains of *M. tuberculosis*. The latter are characterized by the deletion of a 2.1-kb fragment termed TbD1, account for most current tuberculosis (TB) cases, and include the Beijing family of strains that are prevalent in Asia and have caused outbreaks worldwide (21). A separate lineage, marked by deletion of region of difference 9 (RD9), diverged from an ancestral *M. tuberculosis* and sequentially gave rise to *M. africanum*, *M. microti*, *M. pinnipedii*, *M. caprae*, and *M. bovis*. It is thought that *M. bovis*, which has the greatest number of deletions, is the most recent member of this lineage (10). These results clearly refute the popular belief that *M. tuberculosis* evolved from *M. bovis* through the adaptation of a bovine strain to the human host (22). Since *M. tuberculosis* contains chromosomal segments that have been deleted from *M. bovis*, *M. tuberculosis* could not have evolved from *M. bovis*. In fact, these data

---

Figure 1
Evolutionary scheme for the *M. tuberculosis* complex. The phylogenetic scheme is based on informative markers present or absent in the progeny of each lineage. Markers in boxes include lost RD regions, SNPs, lost spoligotype spacers, and deletions in the *pks15/1* and TbD1 loci. The seven species of the *M. tuberculosis* complex and their respective natural hosts are shown, as well as the three principal genetic groups of *M. tuberculosis* and their representative strains. Adapted with permission from the *Journal of Theoretical Biology* (1), *Proceedings of the National Academy of Sciences of the United States of America* (11), and *Microbiology* (16).

raise the converse possibility that humans transmitted tubercle bacilli to animals and those bacilli subsequently evolved into *M. bovis*. Although this is possible, the isolation of *M. tuberculosis* complex DNA from a 17,000-year-old bison skeleton in North America suggests that wild animals were infected prior to domestication of livestock 8,000–10,000 years ago (23, 24).

Modern evolution of *M. tuberculosis*
Several recent studies have used synonymous SNPs and/or LSPs to construct deep phylogenetic lineages for *M. tuberculosis* isolates from around the globe (3, 6, 7, 25–27). These studies uniformly show that distinct geographic regions of the world have variable distributions of lineages, with certain regions having a major lineage that is a minor contributor elsewhere. Moreover, these studies have identified a clonal population structure for *M. tuberculosis* and confirmed the absence of ongoing horizontal gene transfer between strains. It is hypothesized that the genetic population structure of *M. tuberculosis* identified by these studies will yield an association between different mycobacterial lineages and important traits such as clinical phenotypes, transmissibility, likelihood of drug resistance, and adaptation to distinct human populations. However, the relative importance of SNPs versus LSPs as the molecular *determinants* of trait variability is debated. Studies indicate that each strain has roughly 200 intergenic SNPs that can potentially alter gene expression, 500 genes that are affected by at least one nonsynonymous SNP, and 100 genes that are lost or inactivated due to LSPs (2, 28, 29). Given that the

effect of SNPs is usually smaller than that of the loss of an entire gene, it is suggested that LSPs are the main source of phenotypic variability between strains (29). However, while the majority of drug-resistant mutations are due to SNPs (30–33), the best-characterized example of variability in virulence is due to an LSP in the pks 15/1 locus (15, 17). Thus both SNPs and LSPs contribute to strain variability, and it is possible that specific selective pressures favor one over the other.

Several studies have found associations between mycobacterial genotypes and phenotypes in humans. First, in an urban cosmopolitan setting, the transmission of phylogeographic lineages of *M. tuberculosis* is nonrandomly associated with various ethnic populations (7, 26, 34). For example, an individual of Chinese ethnicity in San Francisco is more likely to acquire and develop TB due to a strain of *M. tuberculosis* of the East Asian lineage than any other phylogeographic lineage. These results suggest that over human evolutionary times, specific phylogeographic lineages have adapted to (or coevolved with) distinct human populations and that such host-pathogen compatibility operates even when transmission takes place outside the region of origin (26). The extent to which these host-pathogen associations are driven by sociologic and epidemiologic factors versus biological determinants is not yet known. However, if biological determinants make a major contribution, then the greatest diversity among *M. tuberculosis* lineages should be observed in Africa, where the genetic diversity among human populations is greater than anywhere else in the world (35).
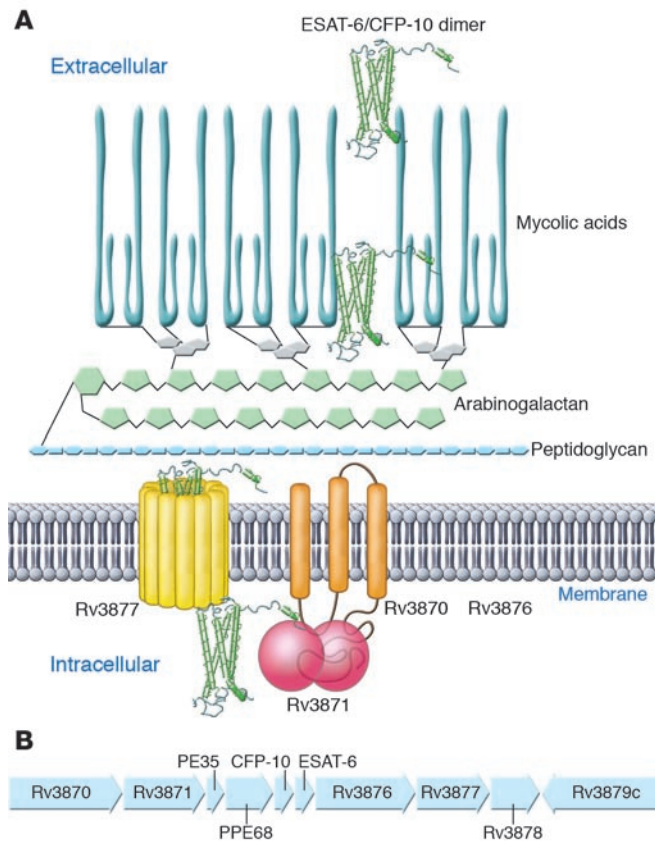
Figure 2
The RD1 locus and components of the ESX-1 secretion system. ESAT-6 and CFP-10 dimer secretion (**A**) by *M. tuberculosis* depends on other genes in the RD1 locus (**B**). The product of *Rv3870* is a membrane protein that interacts directly with Rv3871, a predicted cytoplasmic protein. Based on homology to the SpoIIIE/FtsK family, Rv3870 and Rv3871 are thought to form a membrane-bound ATPase that provides energy for export of secretion substrates; the carboxyl terminus of CFP-10 interacts directly with the carboxyl terminus of Rv3871. *Rv3877* encodes a protein with 12 membrane-spanning domains and is likely to form a secretion pore. *Rv3876* is essential for secretion of CFP-10 and ESAT-6 dimers, but its role has not been determined. Secretion substrates encoded by genes outside the BCG RD1 locus include the product of *Rv3616c* (57), and additional genes essential for ESAT-6 secretion that are outside of the BCG RD1 include *Rv3614c*, *Rv3615c*, and *Rv3616c* (57, 60). The products of *Rv3614c* and *Rv3882c* have been found to directly interact (60).

Similar host-pathogen associations have been documented between the phylogeographic lineage of *M. tuberculosis* and the type of mutations (i.e., KatG[S315T] or a C-to-T substitution in *inhA* promoter–15) that renders strains resistant to the first-line antituberculosis drug isoniazid (INH) (34, 36). These findings suggest that the lineage of *M. tuberculosis* influences the mechanism of INH resistance, which in turn affects bacterial fitness and the transmission of specific INH-resistant strains (36).

RD1/ESX-1 and the pathogenesis of TB
Determination of the genome sequence of *M. tuberculosis* (37) has also facilitated discovery of bacterial determinants of virulence. One particularly productive line of investigation has revealed the molecular basis of attenuation of the bacille Calmette-Guérin (BCG) vaccine strain(s) of *M. bovis*. *M. bovis* BCG, the most widely used human vaccine in history (more than 3.5 billion doses have been administered), was derived by in vitro passage of an isolate of *M. bovis*. A spontaneous change in colony morphology was followed by discovery that the bacteria with altered colony morphology had lost their virulence in experimental animals (38). Since its first use as a human vaccine against TB in 1921, BCG has shown a high degree of safety, indicating that it clearly lost elements that are essential to cause disease in immunocompetent humans.

Initial insight into the mode of attenuation of BCG was obtained when subtractive hybridization identified three RDs that were present in *M. tuberculosis* and *M. bovis* but absent from BCG (39). Subsequent analysis using whole-genome microarrays or bacterial artificial chromosome arrays revealed additional RDs, but RD1, a 9.5-kb deletion, is the only region of difference

that is absent from all strains of BCG (*n* = 13) and present in all of the hundreds of strains of virulent *M. tuberculosis* analyzed to date (38, 40, 41). Compared with *M. tuberculosis*, BCG RD1 encompasses seven genes and truncations of two additional genes (in the original annotated genome sequence, these genes are termed *Rv3870–Rv3879c*; ref. 37 and Figure 2). *M. microti*, which has also been used as a human vaccine, lacks an overlapping set of genes: *Rv3864–Rv3876* (42). Therefore, comparative genomic analysis of mycobacteria that have been proven to be attenuated in humans has pointed to a specific chromosomal region implicated in virulence. It is interesting to note that another human mycobacterial pathogen, *M. ulcerans*, lacks an intact RD1 (43) and causes disease as an extracellular, not an intracellular, pathogen that expresses a characteristic mycolactone toxin (44, 45).

There is now strong experimental evidence that genes present in RD1 are essential for virulence of members of the *M. tuberculosis* complex. BCG complemented with a cosmid that spans RD1 increased growth of the bacteria in the lungs and spleen of mice compared with the parental BCG, although RD1-complemented BCG remained less virulent than virulent *M. tuberculosis* (H37Rv) (46). Complementation with RD1 increased virulence of BCG more markedly in SCID mice than in WT mice, implying that the effects of RD1 genes in vivo include roles in evasion of innate immunity. In addition, deletion of RD1 from virulent *M. tuberculosis* H37Rv resulted in attenuation of the ability to grow intracellularly in human macrophages, to cause macrophage cytotoxicity, and to grow and cause pathology in the lungs and disseminate to the spleen of immunocompetent mice after aerosol infection; deletion of RD1 from *M. tuberculosis* yielded a strain whose virulence was similar to that of BCG Russia (47). Deletion of RD1 also attenuated other strains of virulent mycobacteria, including *M. tuberculosis* H37Rv, Erdman, and CDC1551, and *M. bovis* Ravenel, and virulence of the mutants was attenuated in SCID mice, as well as in immunocompetent mice (48). Together, these results indicate that genes within RD1 are essential for complete virulence of *M. tuberculosis* and that deletion of RD1 was likely the original mutation attenuating virulence that occurred during derivation of BCG.
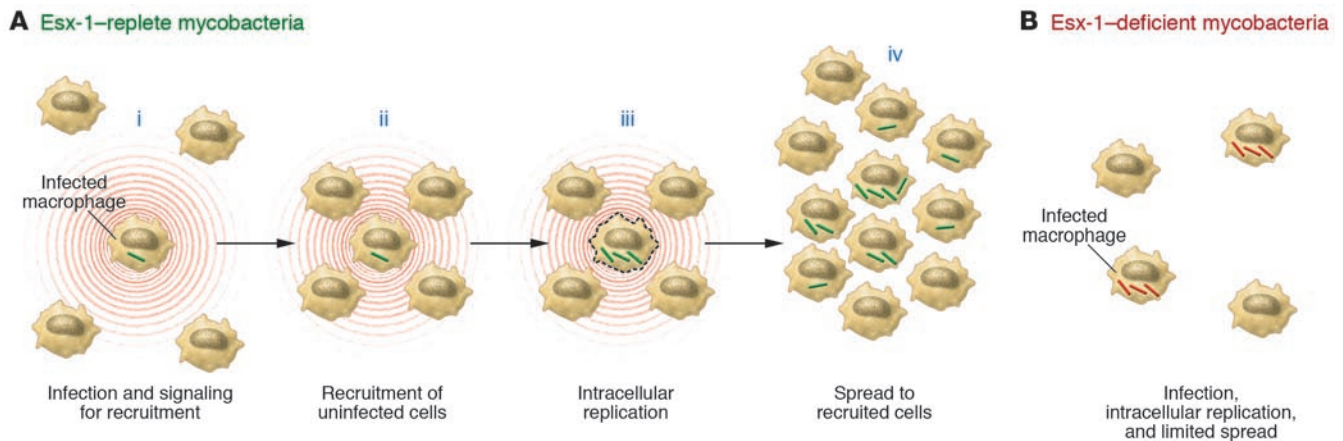
RD1 genes contribute to the virulence of *M. tuberculosis* by encoding a secretory apparatus and its substrates (48–50) (together termed early secreted antigen 6 kDa [ESAT-6] secretion system 1 [ESX-1]) or secretion in mycobacteria (SNM). Two of the genes, *Rv3874* and *Rv3875*, encode two previously identified secreted proteins, ESAT-6 (51) and culture filtrate protein 10 kDa (CFP-10)

Figure 3
Role of ESX-1 in cell-to-cell transmission of *M. tuberculosis* in vivo. (**A**) Macrophages infected with ESX-1–replete mycobacteria signal to recruit uninfected macrophages to the surrounding area (i), including in close proximity to the initially infected macrophage (ii). After intracellular replication, ESX-1–replete bacteria spread to the closely apposed, newly recruited macrophages, with or without lysis of the initially infected macrophage (iii). After spread to the newly recruited macrophages (and dendritic cells), the mycobacteria replicate further (iv), to sustain the cycle of cell-to-cell spread until the onset of adaptive immunity. (**B**) ESX-1–deficient mycobacteria are defective in the ability to recruit uninfected macrophages and to form aggregates with the initially infected cell and in the ability to spread to adjacent cells. As a consequence, ESX-1–deficient mycobacteria are less efficient in expanding the size of the pool of infected cells but replicate efficiently in the initially infected cells.

(52), which are targets of cellular immune responses in mice and humans infected with *M. tuberculosis*. While these proteins are detected in supernatants of *M. tuberculosis* cultures, it was unclear how they were secreted, since they both lack recognizable signal sequences. The observation that other RD1 genes encode predicted membrane proteins suggested that they encode a novel system for protein secretion, and this hypothesis has been sustained by strong evidence. In *M. tuberculosis*, mutation of *Rv3870*, *Rv3871*, *Rv3876*, and *Rv3877* results in lack of secretion, but not lack of synthesis, of ESAT-6 and CFP-10 (48, 50, 53). ESAT-6 (Rv3875) and CFP-10 (Rv3874) interact with each other with high affinity (≤1 × 10⁻⁸ M) to form 1:1 heterodimers (54, 55), and their secretion appears to be mutually dependent, at least in *M. tuberculosis* and *M. bovis* (48, 49, 53, 56, 57). The ESAT-6/CFP-10 complex interacts directly with the C-terminal domain of Rv3871 (50) through the disordered carboxyterminal tail of CFP-10 (58). Addition of the carboxyterminal domain of CFP-10 to yeast ubiquitin caused the heterologous protein to be secreted from *M. tuberculosis*, implying that this domain serves as a signal sequence for the ESX-1 system (Figure 2). In contrast to ESAT-6 and CFP-10, Rv3873 (also known as PPE68) is neither secreted itself nor essential for secretion of ESAT-6/CFP-10 (56, 59). Rv3872 (also known as PE35) has been detected as a secreted protein (57), but its role in the secretion of ESAT-6 is unknown, although it may be needed for synthesis of ESAT-6 (56).

Secretion of ESAT-6 requires the genes in RD1 and at least 3 additional genes at a different locus. A signature-tagged mutagenesis screen identified *Rv3614c* and *Rv3615c* (60), while a proteomic screen identified the product of *Rv3616c* (termed EspA) in supernatants of WT, but not RD1-deficient, *M. tuberculosis* (57). Secretion of EspA required both ESAT-6 and CFP-10, and secretion of ESAT-6 and CFP-10 required EspA. Yeast two-hybrid analysis revealed that the product of *Rv3615c* interacted stably with itself but not other identified components of the secretion system and the product of *Rv3614c* interacted with the product of *Rv3882c* (a gene near

the RD1 locus), while EspA exhibited no detectable interactions with other screened proteins (60). It is likely that other proteins, in addition to ESAT-6, CFP-10, PE35, and EspA, are secreted by the RD1/ESX-1 system, and the identity of the secreted effector(s) of virulence remains to be established.

M. marinum, which causes TB in fish and amphibians, contains an RD1 locus with an organization identical to that of *M. tuberculosis*, and the RD1 locus is required for *M. marinum* virulence (61, 62). Moreover, mutation of the *M. marinum* homologs of the *M. tuberculosis* RD1 genes *Rv3876*, *Rv3878*, and *Rv3879c* resulted in loss of virulence and loss of secretion of ESAT-6. Secretion and virulence were also lost with mutations in the homologs of genes flanking RD1, including *Rv3866*, *Rv3867*, *Rv3868*, and *Rv3881c*, indicating that the corresponding *M. tuberculosis* genes should be examined for roles in secretion and virulence. While the existing evidence in *M. tuberculosis*, *M. bovis*, and *M. microti* indicates that secretion of ESAT-6 and CFP-10 may be absolutely interdependent, mutations in the *M. marinum* homologs of Rv3876, Rv3878, and Rv3879c allowed secretion of CFP-10 without secretion of ESAT-6 (61). The possibility that dissociation of ESAT-6 and CFP-10 secretion may occur in human TB is suggested by the recent observation that immune responses to ESAT-6, but not to CFP-10, are attenuated in people infected with *M. africanum*, in which the homolog of *Rv3879c* is a pseudogene (63). Analysis of protein secretion by *M. africanum* will provide information as to whether defective secretion of ESAT-6 accounts for the attenuated immune responses to ESAT-6 in people infected with *M. africanum*, and complementation with a functional copy of *Rv3879c* will determine whether the inactivation of this gene accounts for the observation.

What are the roles of the RD1/ESX-1 system in virulence? ESX-1–replete mycobacteria exhibit higher cytotoxicity for cultured human and mouse macrophages and for a human alveolar epithelial cell line than do ESX-1–deficient mycobacteria (47, 48, 53, 61). Since RD1-containing pathogenic mycobacteria are facultative intracellular pathogens, the fact they cause death of their resident

cells might seem to be self-destructive, but strong evidence indicates that ESX-1–dependent cytotoxicity is coupled to enhanced cell-to-cell spread of mycobacteria. This was initially noted in cultured epithelial cells, in which WT *M. tuberculosis* and RD1-deleted *M. tuberculosis* entered cells with indistinguishable frequencies, but after 72 hours (2–3 generations for *M. tuberculosis*), the frequency of cells infected with RD1-replete bacteria was approximately twice that of cells infected with RD1-deleted bacteria, and there were 10-fold more RD1-deleted bacteria than RD1-replete bacteria per infected cell (48). This observation was extended in experiments that demonstrated that in cultured epithelial cells or murine bone marrow–derived macrophages, the spreading of ESX-1–deficient bacteria to adjacent cells is impaired, as revealed by serial analysis of foci containing GFP-expressing bacteria (61). In those studies, there was complete concordance between loss of cytolytic activity and loss of spreading activity in all of the single-gene mutants examined. In vivo studies also indicate that ESX-1 contributes to cell-to-cell spread. In adult zebrafish inoculated with *M. marinum* in the peritoneum, dissemination to the liver was markedly decreased in fish inoculated with ESX-1–deficient bacteria (61). In a real-time assay using GFP-expressing *M. marinum* and transparent zebrafish embryos, RD1-deficient bacteria were defective in causing aggregation of macrophages to initiate granuloma formation, and subsequent to macrophage aggregation, RD1-deficient bacteria were defective in their ability to spread from initially infected macrophages to their newly recruited neighbors (62). Studies on *M. marinum* expressing red or green fluorescent protein to differentiate between RD1-deficient and WT bacteria revealed that macrophages infected with RD1-deficient bacteria could traffic to aggregates of macrophages infected with WT bacteria, but macrophages infected with RD1-deficient bacteria did not cause recruitment of macrophages infected with WT bacteria (62). These results indicate that one or more components of ESX-1 are required for a signal for recruitment and aggregation of macrophages in order to bring additional target cells into close apposition to promote spread of mycobacteria within the host (Figure 3). That this mechanism is likely to occur in lungs of animals infected with *M. tuberculosis* is supported by the observation that infection with ESX-1–deficient bacteria results in aggregates of cells immediately adjacent to the airways but not to more distal portions of the lung parenchyma (47, 53). Additional activities attributed to the ESX-1 secretion system include suppression of proinflammatory cytokine production (50, 60), arrest of phagosome maturation (64), and induction of IFN-β (65).

The mechanisms whereby proteins exported by the ESX-1 system exert their effects on host cells have not yet been fully elucidated. The observation that ESAT-6 alone or complexes of ESAT-6 and CFP-10 can cause membrane permeabilization (48) has been challenged (54), and the solution structure of the complex does not reveal hydrophobic domains or patches of charged residues that would provide a structural basis for perturbation of a phospholipid bilayer. In the solution structure of ESAT-6/CFP-10 dimers, the complexes are composed of a 4-helix core (2 helices contributed by each protein), with the amino and carboxyl termini of both proteins unstructured (54). The dimers bind selectively to mononuclear phagocytes, and deletion of the carboxyl terminus of CFP-10, but not ESAT-6, markedly diminished binding, indicating that this unstructured domain may specifically interact with a surface receptor to modulate macrophage activities. While this may be a lead to understanding the mechanisms of ESX-1–dependent

activities of mycobacteria, it has not been established that ESAT-6 and/or CFP-10 are the effector molecules that exert the effects observed with ESX-1–replete mycobacteria. It is also worth noting that the *M. tuberculosis* genome contains 5 paralogs of the ESX-1 secretion system that have functions distinct from those of ESX-1; the functions of these remain to be elucidated.

In summary, the RD1/ESX-1 system is clearly an essential contributor to the pathogenesis of TB. Its best established mode of action is to facilitate cell-to-cell spread during the early stages of infection, but the molecular mediators and mechanisms of these activities remain to be characterized. As described in the following section, the essential role of ESX-1 in virulence makes it an attractive target for diagnosis of TB infection, since it is present in all strains of *M. tuberculosis* that are capable of causing disease in humans.

Diagnosis of TB

Improved control of TB depends on development and application of better means of diagnosing it. While nucleic acid amplification would seem an obvious choice for the detection of *M. tuberculosis*, this approach has so far been successfully applied only to confirm the identity of an acid-fast smear–positive or culture-positive sample as a member of the *M. tuberculosis* complex and for sequence-based detection of drug-resistant mutations.

One significant advance that has benefited from *M. tuberculosis* genomics is the development of new in vitro diagnostic assays that distinguish immune responses due to infection with *M. tuberculosis* from responses due to BCG vaccination. For more than 4 decades, tuberculin skin testing with purified protein derivative (PPD; which is hardly pure by modern standards, but is purer than its predecessors) has been the major means of diagnosing latent infection with *M. tuberculosis*. However, since BCG vaccination also causes tuberculin skin test reactivity, the diagnostic value of tuberculin skin testing is largely abrogated in individuals that have received BCG. The discoveries that ESAT-6 and CFP-10 are present in all strains of virulent *M. tuberculosis*, and absent from all strains of BCG, together with the knowledge that ESAT-6 and CFP-10 are immunogenic in humans (reviewed in ref. 66), has led to development of diagnostic assays that use these proteins as antigens for detection of *M. tuberculosis*–specific immune responses.

The two new diagnostic assays, both termed IFN-γ release assays (IGRAs), differ mainly in the technique for detecting responses to *M. tuberculosis*–derived antigens. QuantiFERON–TB Gold (Cellestis) uses ELISA and T-SPOT.*TB* (Oxford Immunotec) uses ELISPOT to detect secretion of IFN-γ released from ESAT-6– and/or CFP-10–specific T lymphocytes. Both are in vitro assays that use fresh blood samples, so they offer the advantage over tuberculin skin testing of not requiring two visits for every subject (first to have PPD injected and then to have a skilled reader measure and record the result). While the absence of a gold standard prohibits accurate assessment of the sensitivity and specificity of the IGRAs for diagnosis of latent TB infection, a prospective study that assayed immune responses before and after BCG vaccination confirmed that BCG vaccination does not induce immune responses to ESAT-6 (67). A major limitation that the IGRAs share with tuberculin skin testing is their inability to distinguish between latent TB infection and active TB disease. Moreover, recent studies indicate that the sensitivity of IGRAs in culture-positive TB may be as low as 60%, indicating that they cannot be used to exclude a diagnosis of TB (68). While further investigation is necessary to define the optimal roles of ESAT-6– and CFP-10–specific IGRAs, exploiting

the information from comparative genomics of *M. tuberculosis* and BCG has led to the development of these new diagnostic assays, developed based on the information from comparative genomics of *M. tuberculosis* and BCG, are likely to replace tuberculin skin testing for detection of latent TB infection.

Unfortunately, diagnosis and treatment of latent TB infection is a luxury restricted to areas of low TB prevalence. The greater urgent need is the development and application of improved diagnostic techniques for active TB, to identify individuals who need treatment, and to keep them from transmitting TB. Current TB control programs in nonindustrialized countries depend on microscopic examination of stained sputum smears, a procedure with less than 50% sensitivity for active pulmonary TB. While recent initiatives have stimulated pursuit of a wide variety of approaches to diagnose active TB and detect drug resistance, the most promising new assay available so far is an adaptation of long-established broth culture techniques. The microscopic-observation drug susceptibility (MODS) assay, developed in Peru, combines broth culture in 24-well tissue culture plates with early detection of growth by microscopic examination of cultures to detect characteristic cord formation of *M. tuberculosis* (69). A recent study in Lima, Peru, found that the sensitivity of MODS exceeded that of automated mycobacterial culture or culture on Löwenstein-Jensen solid medium (70). Moreover, MODS provided positive results in a median of 7 days, compared with 13 days for automated mycobacterial culture, and 26 days for culture on Löwenstein-Jensen medium, and MODS provided similar advantages in the time to obtain drug susceptibility results. While MODS shows clear promise for higher-sensitivity detection of cases of TB, further innovations will be required to provide even more rapid diagnosis of active TB in resource-poor settings.

Unsolved questions and problems

TB is an enormous health problem worldwide, and the spread of extensively drug-resistant *M. tuberculosis* threatens to overcome the available tools of TB control. Despite this, TB has not yet attracted celebrity advocates in the manner of HIV and malaria. Thanks in part to the availability of the genome sequences of multiple strains of *M. tuberculosis*, the pathogenesis of TB is increasingly amenable to study, but many important unsolved biological problems remain.

Among the unsolved biological problems is the lack of understanding of the mechanisms of the characteristic slow growth of *M. tuberculosis*. The doubling time of *M. tuberculosis* in rich broth media or during logarithmic growth in vivo is approximately 24 hours, but the rate-limiting determinants of this growth rate are not known. In addition, the advantages to the bacteria of such slow replication are unclear. Does slow growth specifically contribute to stealth mechanisms that minimize the efficacy of innate and adaptive immune responses, or is slow growth an effect of evolution of the mechanisms that allow *M. tuberculosis* to remain latent for decades within a healthy host? Perhaps the greatest benefit to the bacteria is the maximization of transmission by slow growth during reactivation. By promoting an insidious onset and slow progression without immobilizing or causing rapid death of the host, the bacteria optimize the chances of being spread to additional susceptible individuals. While comparative genomic studies have revealed clear divisions between slowly growing and rapidly growing mycobacteria, further studies are needed to elucidate the mechanisms that determine the rate of growth and to identify the specific aspects of *M. tuberculosis* biology that benefit from slow growth.

Another unsolved biological problem is the identification of the breadth and thorough characterization of the ecological niches of *M. tuberculosis* in vivo. *M. tuberculosis* is a facultative intracellular bacterium, but the diversity of the cells in which *M. tuberculosis* resides, replicates, and traffics through the host during the course of infection is incompletely defined. For example, it is widely believed that macrophages are the predominant resident cells of *M. tuberculosis* in vivo, though dendritic cells have also been observed to be infected. What is the quantitative distribution of *M. tuberculosis* in cells of distinct phenotypes? Do distinct cell types differ as sanctuaries from innate or adaptive immune responses? It is also unclear whether, or at what stages of infection, *M. tuberculosis* exists extracellularly, other than at the terminal stages of a pulmonary cavity. It is similarly unclear how the bacteria are specifically adapted to the lungs in preference to other tissues and how the bacteria survive and grow within phagosomes of macrophages and other cells. Even when *M. tuberculosis* is injected intravenously into mice, subsequent replication in the lungs far exceeds that in other tissues (71, 72). Is this solely due to factors that are enriched in the lungs (such as oxygen) that benefit the bacteria, or are the lungs immunologically disadvantaged in specific ways that allow *M. tuberculosis* to thrive? There is no question that *M. tuberculosis* transmission is enhanced by its adaptation to the lungs, but the mechanisms that are essential for the bacteria to selectively thrive there still need to be defined, and the success of drug and vaccine development may depend on understanding and overcoming those mechanisms. The availability of the genome sequences of several strains of *M. tuberculosis* has already facilitated discovery of determinants of certain aspects of virulence, and application of that knowledge in novel genetic screens offers considerable promise for the discovery of unique aspects of *M. tuberculosis* biology.

One of the most important unsolved biological problems is that *M. tuberculosis* can persist and reactivate, despite development of adaptive immunity. The adaptive immune response to *M. tuberculosis* requires weeks to months to develop, but the rate-limiting step(s) are not well defined, and it is not known whether *M. tuberculosis* actively intervenes to delay initiation of immune responses. It is similarly poorly understood how *M. tuberculosis* persists and progresses despite development of adaptive immune responses that generate appropriately differentiated effector T lymphocytes. Does *M. tuberculosis* modulate its antigen expression to evade adaptive immunity? As noted in the preceding section, humans infected with *M. tuberculosis* have memory and/or effector T lymphocytes that recognize peptides from ESAT-6, but there is evidence that expression of the gene that encodes ESAT-6 decreases during the chronic stage of infection in mice (73). If ESAT-6 protein expression is significantly decreased, then ESAT-6–specific T cells may have little protective effect after the initial stage of infection. If this is a general phenomenon for genes that encode immunogenic proteins of *M. tuberculosis*, then vaccine development efforts should involve identification of antigens expressed during later stages of infection. The finding that *M. tuberculosis* inhibits macrophage responses to exogenous IFN-γ in vitro suggests another potential immune evasion strategy (74–76), but whether or how this contributes to the success of *M. tuberculosis* in vivo has not been established. Much more knowledge is needed about how *M. tuberculosis* evades adaptive immunity in order to facilitate development and optimal testing of efficacious TB vaccines.

In the decade since the genome sequence of a strain of *M. tuberculosis* was first determined, there have been major advances in understanding the biology of TB. Researchers have made particularly large strides in understanding the biology of *M. tuberculosis* itself. In ongoing and future studies, similar large steps are required to gain a comprehensive understanding of the biology of host responses and how the bacteria exploit and evade them. With rapid advances in basic immunology and human genetics and their application to the problems of TB, we will learn a great deal more, and the new knowledge should provide a source of better diagnostic tests, novel therapies, and efficacious vaccines. Nearly 20 million humans have died of TB in the past decade, but the future holds great promise for making TB a disease of the past.

Address correspondence to: Joel D. Ernst, New York University School of Medicine, 522 First Avenue, Smilow 901, New York, New York 10016, USA. Phone: (212) 263-5165; Fax: (212) 263-9230; E-mail: joel.ernst@med.nyu.edu.

N. Banaiee's present address is: Department of Pathology, Stanford University School of Medicine, Palo Alto, California, USA.

1. Smith, N.H., et al. 2006. Ecotypes of the Mycobacterium tuberculosis complex. *J. Theor. Biol.* **239**:220–225.
2. Fleischmann, R.D., et al. 2002. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *J. Bacteriol.* **184**:5479–5490.
3. Gutacker, M.M., et al. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics.* **162**:1533–1543.
4. Huard, R.C., et al. 2006. Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. *J. Bacteriol.* **188**:4271–4287.
5. Hughes, A.L., Friedman, R., and Murray, M. 2002. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of Mycobacterium tuberculosis. *Emerg. Infect. Dis.* **8**:1342–1346.
6. Sreevatsan, S., et al. 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. U. S. A.* **94**:9869–9874.
7. Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., and Small, P.M. 2004. Stable association between strains of Mycobacterium tuberculosis and their human host populations. *Proc. Natl. Acad. Sci. U. S. A.* **101**:4871–4876.
8. Smith, N.H., et al. 2003. The population structure of Mycobacterium bovis in Great Britain: clonal expansion. *Proc. Natl. Acad. Sci. U. S. A.* **100**:15271–15275.
9. Supply, P., et al. 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area. *Mol. Microbiol.* **47**:529–538.
10. Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., and Hewinson, R.G. 2006. Bottlenecks and broomsticks: the molecular evolution of Mycobacterium bovis. *Nat. Rev. Microbiol.* **4**:670–681.
11. Brosch, R., et al. 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc. Natl. Acad. Sci. U. S. A.* **99**:3684–3689.
12. Gutierrez, M.C., et al. 2005. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog.* **1**:e5.
13. Smith, N.H. 2006. A re-evaluation of M. prototuberculosis. *PLoS Pathog.* **2**:e98.
14. Brisse, S., Supply, P., Brosch, R., Vincent, V., and Gutierrez, M.C. 2006. "A re-evaluation of M. prototuberculosis": continuing the debate. *PLoS Pathog.* **2**:e95.
15. Constant, P., et al. 2002. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methly esters and that strains devoid of

phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. *J. Biol. Chem.* **277**:38148–38158.
16. Marmiesse, M., et al. 2004. Macro-array and bioinformatic analyses reveal mycobacterial "core" genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. *Microbiology.* **150**:483–496.
17. Reed, M.B., et al. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature.* **431**:84–87.
18. Tsolaki, A.G., et al. 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. *J. Clin. Microbiol.* **43**:3185–3191.
19. Kamerbeek, J., et al. 1997. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
20. Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A., and Behr, M.A. 2002. Genomic deletions suggest a phylogeny for the Mycobacterium tuberculosis complex. *J. Infect. Dis.* **186**:74–80.
21. Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K., and van Soolingen, D. 2002. Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review. *Emerg. Infect. Dis.* **8**:843–849.
22. Stead, W.W., et al. 1995. When did Mycobacterium tuberculosis infection first occur in the New World? An important question with public health implications. *Am. J. Respir. Crit. Care Med.* **151**:1267–1268.
23. Bruford, M.W., Bradley, D.G., and Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* **4**:900–910.
24. Rothschild, B.M., et al. 2001. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin. Infect. Dis.* **33**:305–311.
25. Gutacker, M.M., et al. 2006. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J. Infect. Dis.* **193**:121–128.
26. Gagneux, S., et al. 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**:2869–2873.
27. Filliol, I., et al. 2006. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J. Bacteriol.* **188**:759–772.
28. Musser, J.M., Amin, A., and Ramaswamy, S. 2000. Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure. *Genetics.* **155**:7–16.
29. Tsolaki, A.G., et al. 2004. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci. U. S. A.* **101**:4865–4870.

30. Cavusoglu, C., Hilmioglu, S., Guneri, S., and Bilgic, A. 2002. Characterization of rpoB mutations in rifampin-resistant clinical isolates of Mycobacterium tuberculosis from Turkey by DNA sequencing and line probe assay. *J. Clin. Microbiol.* **40**:4435–4438.
31. Finken, M., Kirschner, P., Meier, A., Wrede, A., and Bottger, E.C. 1993. Molecular basis of streptomycin resistance in Mycobacterium tuberculosis: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol. Microbiol.* **9**:1239–1246.
32. Gagneux, S., et al. 2006. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. *Science.* **312**:1944–1946.
33. Ramaswamy, S.V., et al. 2000. Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of Mycobacterium tuberculosis. *Antimicrob. Agents Chemother.* **44**:326–336.
34. Baker, L., Brown, T., Maiden, M.C., and Drobniewski, F. 2004. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. *Emerg. Infect. Dis.* **10**:1568–1577.
35. Ramachandran, S., et al. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U. S. A.* **102**:15942–15947.
36. Gagneux, S., et al. 2006. Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. *PLoS Pathog.* **2**:e61.
37. Cole, S.T., et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature.* **393**:537–544.
38. Behr, M.A., and Small, P.M. 1999. A historical and molecular phylogeny of BCG strains. *Vaccine.* **17**:915–922.
39. Mahairas, G.G., et al. 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J. Bacteriol.* **178**:1274–1282.
40. Behr, M.A., et al. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science.* **284**:1520–1523.
41. Gordon, S.V., et al. 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol. Microbiol.* **32**:643–655.
42. Brodin, P., et al. 2002. Bacterial artificial chromosome-based comparative genomic analysis identifies Mycobacterium microti as a natural ESAT-6 deletion mutant. *Infect. Immun.* **70**:5568–5578.
43. Stinear, T.P., et al. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. *Genome Res.* **17**:192–200.
44. Adusumilli, S., et al. 2005. Mycobacterium ulcerans toxic macrolide, mycolactone modulates the host immune response and cellular location of M. ulcerans

in vitro and in vivo. *Cell. Microbiol.* **7**:1295–1304.

45. Coutanceau, E., et al. 2005. Modulation of the host immune response by a transient intracellular stage of Mycobacterium ulcerans: the contribution of endogenous mycolactone toxin. *Cell. Microbiol.* **7**:1187–1196.

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

47. Lewis, K.N., et al. 2003. Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation. *J. Infect. Dis.* **187**:117–123.

48. Hsu, T., et al. 2003. 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.* **100**:12420–12425.

49. Pym, A.S., et al. 2003. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. *Nat. Med.* **9**:533–539.

50. Stanley, S.A., Raghavan, S., Hwang, W.W., and Cox, J.S. 2003. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proc. Natl. Acad. Sci. U. S. A.* **100**:13001–13006.

51. Sorensen, A.L., Nagai, S., Houen, G., Andersen, P., and Andersen, A.B. 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. *Infect. Immun.* **63**:1710–1717.

52. Berthet, F.X., Rasmussen, P.B., Rosenkrands, I., Andersen, P., and Gicquel, B. 1998. A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology.* **144**:3195–3203.

53. Guinn, K.M., et al. 2004. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. *Mol. Microbiol.* **51**:359–370.

54. Renshaw, P.S., et al. 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J.* **24**:2491–2498.

55. Renshaw, P.S., et al. 2002. Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence. *J. Biol. Chem.* **277**:21598–21603.

56. Brodin, P., et al. 2006. Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on immunogenicity and virulence. *Infect. Immun.* **74**:88–98.

57. Fortune, S.M., et al. 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc. Natl. Acad. Sci. U. S. A.* **102**:10676–10681.

58. Champion, P.A., Stanley, S.A., Champion, M.M., Brown, E.J., and Cox, J.S. 2006. C-terminal signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis. *Science.* **313**:1632–1636.

59. Demangel, C., et al. 2004. Cell envelope protein PPE68 contributes to Mycobacterium tuberculosis RD1 immunogenicity independently of a 10-kilodalton culture filtrate protein and ESAT-6. *Infect. Immun.* **72**:2170–2176.

60. MacGurn, J.A., Raghavan, S., Stanley, S.A., and Cox, J.S. 2005. A non-RD1 gene cluster is required for Snm secretion in Mycobacterium tuberculosis. *Mol. Microbiol.* **57**:1653–1663.

61. Gao, L.Y., et al. 2004. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol. Microbiol.* **53**:1677–1693.

62. Volkman, H.E., et al. 2004. Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. *PLoS Biol.* **2**:e367.

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

64. Tan, T., Lee, W.L., Alexander, D.C., Grinstein, S., and Liu, J. 2006. The ESAT-6/CFP-10 secretion system of Mycobacterium marinum modulates phagosome maturation. *Cell. Microbiol.* **8**:1417–1429.

65. Stanley, S.A., Johndrow, J.E., Manzanillo, P., and Cox, J.S. 2007. The type I IFN response to infection with Mycobacterium tuberculosis requires ESX-1-mediated secretion and contributes to pathogenesis. *J. Immunol.* **178**:3143–3152.

66. Andersen, P., Munk, M.E., Pollock, J.M., and Doherty, T.M. 2000. Specific immune-based diagnosis of tuberculosis. *Lancet.* **356**:1099–1104.

67. Johnson, P.D., et al. 1999. Tuberculin-purified protein derivative-, MPT-64-, and ESAT-6-stimulated gamma interferon responses in medical students before and after Mycobacterium bovis BCG vaccination and in patients with tuberculosis. *Clin. Diagn. Lab. Immunol.* **6**:934–937.

68. Dewan, P.K., Grinsdale, J., and Kawamura, L.M. 2007. Low sensitivity of a whole-blood interferon-gamma release assay for detection of active tuberculosis. *Clin. Infect. Dis.* **44**:69–73.

69. Moore, D.A., et al. 2004. Microscopic observation drug susceptibility assay, a rapid, reliable diagnostic test for multidrug-resistant tuberculosis suitable for use in resource-poor settings. *J. Clin. Microbiol.* **42**:4432–4437.

70. Moore, D.A., et al. 2006. Microscopic-observation drug-susceptibility assay for the diagnosis of TB. *N. Engl. J. Med.* **355**:1539–1550.

71. Cox, J.S., Chen, B., McNeil, M., and Jacobs, W.R., Jr. 1999. Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature.* **402**:79–83.

72. Peters, W., et al. 2001. Chemokine receptor 2 serves an early and essential role in resistance to Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **98**:7958–7963.

73. Shi, L., North, R., and Gennaro, M.L. 2004. Effect of growth state on transcription levels of genes encoding major secreted antigens of Mycobacterium tuberculosis in the mouse lung. *Infect. Immun.* **72**:2420–2424.

74. Banaiee, N., Kincaid, E.Z., Buchwald, U., Jacobs, W.R., Jr., and Ernst, J.D. 2006. Potent inhibition of macrophage responses to IFN-gamma by live virulent Mycobacterium tuberculosis is independent of mature mycobacterial lipoproteins but dependent on TLR2. *J. Immunol.* **176**:3019–3027.

75. Pai, R.K., Convery, M., Hamilton, T.A., Boom, W.H., and Harding, C.V. 2003. Inhibition of IFN-gamma-induced class II transactivator expression by a 19-kDa lipoprotein from Mycobacterium tuberculosis: a potential mechanism for immune evasion. *J. Immunol.* **171**:175–184.

76. Ting, L.M., Kim, A.C., Cattamanchi, A., and Ernst, J.D. 1999. Mycobacterium tuberculosis inhibits IFN-gamma transcriptional responses without inhibiting activation of STAT1. *J. Immunol.* **163**:3898–3906.