ABSTRACT Most mycobacterial species are harmless saprophytes, often found in aquatic environments. A few species seem to have evolved from this pool of environmental mycobacteria into major human pathogens, such as Mycobacterium tuberculosis, the agent of tuberculosis, Mycobacterium leprae, the leprosy bacillus, and Mycobacterium ulcerans, the agent of Buruli ulcer. While the pathogenicity of M. ulcerans relates to the acquisition of a large plasmid encoding a polyketide-derived toxin, the molecular mechanisms by which M. leprae or M. tuberculosis have evolved to cause disease are complex and involve the interaction between the pathogen and the host.

Here we focus on M. tuberculosis and closely related mycobacteria and discuss insights gained from recent genomic and functional studies. Comparison of M. tuberculosis genome data with sequences from nontuberculous mycobacteria, such as Mycobacterium marinum or Mycobacterium kansasii, provides a perception of the more distant evolution of M. tuberculosis, while the recently accomplished genome sequences of multiple tubercle bacilli with smooth colony morphology, named Mycobacterium canettii, have allowed the ancestral gene pool of tubercle bacilli to be estimated. The resulting findings are instrumental for our understanding of the pathogenic evolution of tuberculosis-causing mycobacteria. Comparison of virulent and attenuated members of the M. tuberculosis complex has further contributed to identification of a specific secretion pathway, named ESX or Type VII secretion. The molecular machines involved are key elements for mycobacterial pathogenicity, strongly influencing the ability of M. tuberculosis to cope with the immune defense mounted by the host.

INTRODUCTION

Mycobacteria are widespread microorganisms characterized by the high G+C content of their genomes and a lipid-rich cell envelope. The genus Mycobacterium represents the only entity within the family Mycobacteriaceae, which belongs to the order Actinomycetales and the phylum Actinobacteria (1). Whereas the great majority of the ∼130 described species in the genus are harmless environmental saprophytes, some mycobacteria have evolved to be major pathogens. With the exception of Mycobacterium abscessus, which is recognized as an emerging human pathogen in cystic fibrosis patients (2–4), the pathogenic species mainly belong to the slowly growing mycobacteria and comprise well-known human pathogens such as Mycobacterium tuberculosis, Mycobacterium leprae, and Mycobacterium ulcerans as well as confirmed animal pathogens such as Mycobacterium bovis, Mycobacterium avium paratuberculosis, and Mycobacterium marinum (5).

Their slow axenic growth, their pathogenicity, and their particular physiology make these bacteria quite
difficult to work with. However, in order to prevent the diseases caused by these pathogens, a detailed understanding is required of their genetic and physiological resources and the mechanisms that have contributed to their evolutionary success. The development of mycobacterial genomics and related research disciplines that are building upon “omic” data now provide the scientific basis for the prediction and identification of factors that determine pathogenicity and differentiate mycobacterial pathogens from nonpathogenic strains. In-depth knowledge of these factors is a key outcome of the research that tries to understand the underlying biological mechanisms employed by the bacteria to circumvent host defense strategies and propagate in hostile environments. Here, we focus on insights gained from recent mycobacterial genome and functional analyses and provide an overview of the evolution and the infection strategies employed by selected mycobacterial pathogens, with the main emphasis on *M. tuberculosis*.

**MYCOBACTERIAL GENOMICS**

Mycobacterial genome research started well before the availability of complete genome sequences. Integrated genome maps that combined restriction digest–based physical maps (6) of the chromosome with sets of ordered cosmids or bacterial artificial chromosomes were of great use for building the scaffold of whole-genome sequencing projects (2) and allowed an independent estimate of genome organization and genome size of the studied strain. Furthermore, the ordered clone libraries used for constructing the integrated genome maps represent an archived source of mycobacterial DNA and serve as a valuable resource for genetic manipulation and gene complementation studies (8).

The first mycobacterial genome that was completely sequenced and made available to the scientific community was that of *M. tuberculosis* H37Rv. This strain, which was originally isolated from a pulmonary tuberculosis patient in 1905, is a widely studied *M. tuberculosis* reference strain that has retained its full virulence over the years (9). The *M. tuberculosis* H37Rv genome sequencing project was the first in a series of mycobacterial genome projects that were based on a fruitful collaboration involving the Institut Pasteur in Paris, France, and the Sanger Institute in Hinxton, United Kingdom, as core partners, allowing the genome sequences of several important mycobacterial species and strains to be obtained (10–13). The analysis of the genome of *M. tuberculosis* H37Rv showed that this bacterium encodes in its 4.4 megabase (Mb) genome around 4,000 predicted proteins and 50 RNA molecules. The G+C content of almost 66% confirmed the phylogenetic position as part of the G+C rich, Gram-positive bacteria. However, it should be noted that in contrast to true Gram-positive bacteria regrouped within the phylum *Firmicutes*, mycobacteria have a different cell envelope architecture that has been shown to contain an inner and an outer membrane (myco-membrane) (16–18), thereby organizationally resembling more closely the envelope of Gram-negative bacteria than Gram-positive bacteria. In this respect, it is noteworthy that about 8% of the *M. tuberculosis* H37Rv genome encodes proteins involved in lipid metabolism. This finding highlights the importance of this class of molecules for the lifestyle of *M. tuberculosis* as an intracellular pathogen that protects itself with a particular cell envelope that is rich in lipids, glycolipids, lipoglycans, and polyketides (18, 19). The presence of enzymes that are predicted to show lipolytic functions further suggests that *M. tuberculosis* might use host-cell lipids and sterols as energy sources via the β-oxidation cycle, a pathway that is required for lipid catabolism. The *M. tuberculosis* genome contains genes for more than 100 enzymes that might be involved in various lipid oxidation pathways used for metabolizing putative degradation products of host cell membranes (10, 20, 21).

While the analysis of the *M. tuberculosis* H37Rv genome sequence showed that around 3.4% of the genome is composed of mobile elements such as prophages (phiRv1, phiRv2) and insertion sequences (IS) belonging to various families (e.g., IS3, IS5, IS21, IS30, IS110, IS256, and ISL3) (22), another key finding was the identification of novel gene and protein families, some of which also contained highly repetitive motifs. Probably, the most surprising of these were the PE and PPE families, which were named according to their characteristic N-terminal motifs ProGlu (PE) or ProProGlu (PPE). There are 169 PE and PPE proteins, representing 7.1% of the genome’s coding capacity, in *M. tuberculosis* H37Rv. Whereas PE and PPE proteins have a highly conserved N-terminal domain of ∼110-amino-acid residues, followed by a C-terminal segment that varies in size and may contain a polymorphic G+C-rich sequence (subfamily PGRS), PPE proteins are characterized by a conserved N-terminal segment of ∼180 residues and a variable C-terminal domain that may contain major polymorphic tandem repeats (subfamily MPTR) (10, 23). It seems clear from recent studies that many of the PE and PPE proteins are surface exposed (24–28). Their transport depends on the functionality of a dedicated
secretion system named ESX-5, which belongs to the type VII secretion systems of mycobacteria (29–31). The exact biological role of these proteins remains obscure, but it appears that their encoding genes have undergone a dramatic expansion during mycobacterial evolution toward pathogenicity. Indeed, the subgroup of slowly growing, pathogenic mycobacteria harbor a much larger number of PE- and PPE-encoding genes than the fast-growing saprophytic species. Similarly, the slow growers also have genes that encode more complex PE and PPE proteins, which contain the PGRS and MPTR regions, respectively (32). Given the large number of genes that encode PE and PPE proteins in the genomes of M. tuberculosis or M. marinum, it seems likely that they fulfill important functions, which are, however, difficult to study due to the apparent redundancy of sequence motifs within these proteins.

One of the key questions in this respect is whether some PE/PPE members play a unique role in pathogenesis. The protein PE_PGRS33 (Rv1818c), for example, was described as exerting a cytotoxic effect on host cells (33, 34). Interestingly, the gene encoding PE_PGRS33 in M. tuberculosis strains is not present in the genomes of Mycobacterium canettii strains, representing early branching tubercle bacilli with smooth colony morphology and showing lower virulence and persistence in animal infection models relative to M. tuberculosis (35).

This gene is also absent from the more distantly related nontuberculous mycobacteria, such as M. marinum (15), suggesting that the gene rv1818c was specifically acquired by a recent ancestor of the M. tuberculosis complex from an unknown source. It remains to be determined whether the acquisition of this gene may have supplied some selective advantage to M. tuberculosis for survival inside the host.

Another large gene family identified in the M. tuberculosis H37Rv genome concerns, for example, the genes encoding the mammalian cell entry proteins Mce, originally named after the observation that one of the mce genes cloned into Escherichia coli conferred to the recombinant strain the ability to enter HeLa cells (36). Four mce clusters with a total of 24 genes were identified at four positions in the genome of M. tuberculosis H37Rv (10, 37), although the number of mce clusters is even higher in environmental mycobacteria (38). One Mce operon (Mce3) is deleted in the closely related M. bovis lineage due to the deletion of the region of difference (RD) 7 (39, 40). It is noteworthy that most recently, a fifth Mce operon was identified in M. canettii strain STB-J of the early branching tubercle bacilli (35). Certain mce operons of M. tuberculosis have been associated with a role in pathogenicity (41, 42); in particular, the Mce4 operon seems to be required for cholesterol uptake and enhanced persistence of M. tuberculosis (43, 44).

Another highlight of the M. tuberculosis H37Rv genome project was the identification of numerous ESAT-6 loci, which were named after the paradigm member of this class of proteins defined as Early Secreted Antigenic Target due to its presence in culture filtrate of M. tuberculosis and its relatively small size of 6 kDa (45). ESAT-6 is encoded together with its protein partner CFP-10, the 10-kDa culture filtrate protein of M. tuberculosis (46) by a so-called esx operon in the proximity of the origin of replication. The genome project revealed that apart from these two genes, there were 21 other esx genes in the genome, organized mainly in gene couples, including five loci, where the esx-gene couples were found within large operon structures (10, 37, 47) together with genes encoding components of dedicated secretion systems involved in their export. As a common sequence motif in this ESX protein family, a specific Trp-Xaa-Gly (W-X-G) motif forming a characteristic hairpin bend between the two helical parts of the proteins was determined (48). Insights into the effects of Esx proteins and their ESX systems on the pathogenicity of mycobacteria are described below in the section on mycobacterial systems discovered by genomics that are involved in pathogenicity.

In the 15 years since the first mycobacterial genome description of M. tuberculosis H37Rv, many other mycobacterial genome sequences have been determined, originating from more distantly related mycobacterial species as well as from different tubercle bacilli and M. tuberculosis clinical isolates from various parts of the world. The data are thus available for global analyses via systems-biology-based methods. The in-depth analysis of this enormous dataset will allow specific traits linked to pathogenicity, transmissibility, and/or other strain- or strain-lineage-specific particularities to be identified that have likely been selected during mycobacterial evolution.

INSIGHTS INTO THE MACROEVOLUTION OF M. TUBERCULOSIS BASED ON COMPARISON WITH M. MARINUM AND MYCOBACTERIUM KANSASII

Genome comparisons with more distantly related mycobacterial species have been powerful ways to gain deeper insights into the broader genetic changes that led to the emergence of tuberculosis-causing mycobacteria.
Whole-genome comparisons of *M. tuberculosis* with five selected mycobacterial species such as *M. avium* paratuberculosis (50), *M. leprae* (11), *M. marinum* (15), *Mycobacterium smegmatis* (GenBank accession-no. NC_008596), and *M. ulcerans* (13), for example, revealed 1,072 orthologous genes that are conserved across the six species tested, thus potentially defining the minimum mycobacterial genome (15, 49). Sequence comparisons of the core genes among these strains showed that *M. tuberculosis* is most closely related to *M. marinum* and least closely related to the environmental saprophyte *M. smegmatis* (13), which is in agreement with previous results from partial sequence analysis of the 16S rDNA (5). Phylogenetic reconstructions using these genome sequences strongly suggest a scenario in which *M. tuberculosis*, *M. marinum*, and *M. ulcerans* evolved from a common environmental ancestor, with the tuberculosis-causing mycobacteria undergoing extensive gene loss in parallel to gain of at least 600 new genes via horizontal gene transfer to become a specialized pathogen of humans and certain other mammals (15) (Fig. 1).

In contrast, *M. marinum* has maintained a genome that is 2.2 Mb larger than that of *M. tuberculosis*, which apparently is well adapted to the probably more fluctuating environmental challenges encountered by an aquatic microorganism and facultative pathogen. As an example, *M. marinum* has retained the faculty to produce photochromogenic pigment to protect itself from light, which is linked to the *crtE-crtY* carotenoid biosynthesis gene cluster that does not have a counterpart in *M. tuberculosis* or other tubercle bacilli (15, 51).
However, transposon mutagenesis of *M. marinum* also identified an associated gene named *crtP* to be involved in the regulation of pigmentation, which has homologs in the *M. tuberculosis* genomic region *rv2606c-rv2603c*, which is known to be involved in stress responses (52).

In parallel, a comparison between the 4.4-Mb genome of *M. tuberculosis* and the 6.4-Mb genome of *M. kansasii*, another closely related nontuberculous mycobacterium that can cause pulmonary infections in humans, was also instructive for insights into the macroevolution of tuberculosis-causing mycobacteria. The genomic comparison between the two species predicted that at least 137 genes had been acquired by *M. tuberculosis* mainly via horizontal gene transfer since the phylogenetic divergence of the two species, including mainly genes coding for metabolic functions and modification of mycobacterial lipids (53, 54). It is noteworthy that based on the analysis of shared mycobacterial genes, *M. kansasii* appears more closely related to *M. tuberculosis* than *M. marinum* and *M. ulcerans*, which were considered the closest relatives of the tubercle bacilli according to comparisons of the 16S rRNA encoding genes (5, 55). This indicates that depending on the genes considered for the analysis, the relative phylogenetic position of an organism may vary due to the different evolutionary histories of the genes. In any case, it is clear from these analyses that the *M. marinum* (including *M. ulcerans*) and *M. kansasii* phylogenetic lineages represent the most closely related nontuberculous mycobacterial species of *M. tuberculosis* for which genome information is presently available. Although both species can cause opportunistic infections in humans (15, 54), they are far from having acquired the outstanding capacity of *M. tuberculosis* to act as a professional pathogen that causes lung tissue necrosis and cavity formation in order to ensure efficient transmission via aerosol to new hosts.

**MICROEVOLUTIONARY GENOMICS OF THE TUBERCLE BACILLI**

A better understanding of the more recent evolution of *M. tuberculosis* has come from genome comparisons within the *M. tuberculosis* complex, which comprises a variety of ecotypes that seem to be specifically found in a variety of mammalian hosts (56, 57). Despite this host range, these different tubercle bacilli show an extraordinary genetic homogeneity, with only 0.01 to 0.03% synonymous nucleotide variations between their genomes. Nevertheless, they can be differentiated into some main phylogenetic lineages by the presence or absence of selected genomic regions (39, 57, 58) or, more recently, by comparison of single nucleotide polymorphism (SNP) datasets (59, 60). Based on the distribution of the region of difference 9 (RD9) and the *M. tuberculosis* deleted region TbD1, three main lineages were defined (57). The lineage deleted for RD9 comprises *M. africanum*, prevalent in humans in West Africa, as well as a range of subspecies with animal reservoirs, such as *Mycobacterium microti*, a pathogen originally isolated from voles that also infects cats (61) and sporadically humans (62), *Mycobacterium pinnipedi*, which was isolated from fur seals and sea lions in different continents (63), *Mycobacterium caprae*, isolated from goats and deer (64, 65), and *M. bovis*, the bovine tubercle bacillus, which can infect a wide range of wild and domestic animals (66). The spectrum of RD9-deleted tubercle bacilli was further enriched by the identification of *Mycobacterium mungi* (67), the dassie bacillus (68, 69), a chimpanzee isolate (70), and *Mycobacterium orygis*, which is found in antelopes but was also repeatedly isolated from human tuberculosis cases (40). In contrast to previous hypotheses, the absence of RD9 from these strains suggests that the agent of bovine tuberculosis, *M. bovis*, is not the ancestor of the human tuberculosis agent *M. tuberculosis*, as was thought for a long time (71). This perspective is further supported by results from paleomicrobiological investigations that indicate the presence of *M. tuberculosis* rather than *M. bovis* in ancient human remains (72, 73).

The second major phylogenetic marker identified is the TbD1 region, which is absent from a large cluster of *M. tuberculosis* lineages named “modern” *M. tuberculosis* strains but present in all other members of the *M. tuberculosis* complex (57). This deletion truncates a representative member of the *mmpL* gene family encoding mycobacterial membrane proteins dedicated to the transport of cell wall components (74), whose function is thus likely missing from all modern *M. tuberculosis* strains. When compared to the SNP-based phylogenetic scheme proposed by the team of Gagneux and colleagues, the TbD1-deleted (and RD9-intact) strains correspond to lineages 2, 3, and 4, while the above-mentioned RD9-deleted strains are regrouped in lineages 5 and 6 and the animal strain lineages (56, 59, 60). Finally, lineage 1 in the SNP-based phylogeny comprises *M. tuberculosis* strains that have both regions, RD9 and TbD1, intact and thus resemble the closest last common ancestor of the *M. tuberculosis* complex members with respect to the structure of these genomic regions (Fig. 2). These strains have thus been
named “ancestral” or TbD1-intact *M. tuberculosis* strains (56, 57, 59). They are most prevalent in countries bordering the Indian Ocean (75, 76) and the Philippines and have been reported to differ from strains of other lineages by the type of inflammatory response they induce (77). Strain lineages of *M. tuberculosis* isolates were reported to show close association with their host populations over time, so that a patient’s region of birth can even be used as a predictor of the strain type a patient might carry (78, 79). As such, TbD1-intact strains are predominantly found in patients of South Asian origin. As an exception, *M. tuberculosis* strains with a preserved TbD1 region belonging to a previously unknown, deep branching lineage, named lineage 7, were recently isolated from patients from around the Horn of Africa (80, 81). According to the wide distribution of the lineage 1 strains in South Asian countries with high tuberculosis prevalence, the total number of infections caused by TbD1-intact strains might be larger than that of cases caused by TbD1-deleted strains (82). Overall, RD and SNP analyses confirmed that *M. tuberculosis* strain lineages and the human host populations are geographically linked. Paleomicrobiological investigations using amplification of ancient *M. tuberculosis* DNA from human remains have further suggested that *M. tuberculosis* strains present in the UK and Hungary in the past centuries were TbD1-deleted strains, which are still the most prevalent genotypes in Europe today (72, 73, 83). This marked phylogeographic structure is consistent with the highly clonal population structure of *M. tuberculosis* strains, where genotypes are expected to be stable in space and time (84). Initial genome analyses indicated the lack of any significant recent horizontal gene transfer in classical members of the *M. tuberculosis* complex (12, 60, 85), a feature that was also pointed out by Supply and colleagues (84) based on analyses of genetic linkage between loci containing mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR). In addition, recent genome sequencing studies of *M. tuberculosis* strains involved in tuberculosis epidemics made it possible to follow the genetic events during the outbreaks and to confirm that *M. tuberculosis* strains evolve by clonal descent (86, 87).

In contrast, a recent study has reported homoplastic SNP regions with at least two nucleotides concerned, present in different strains of the *M. tuberculosis* complex, suggesting that potential recombination tracts of small sizes might exist within their genomes (88). However, three of the four examples of suggested recombination tracts presented (in Fig. S4 of reference 88), correspond to adjacent base pairs changes, i.e., putative

**FIGURE 2** Network phylogeny inferred among the five *M. canettii* strains subjected to complete genome sequence analysis and 39 selected genome sequences from members of the classical *M. tuberculosis* complex by NeighborNet analysis, based on pairwise alignments of whole-genome SNP data, which in part are also listed in the lower right portion of the figure. The color code and the naming of different phylogenetic lineages within the *M. tuberculosis* complex refer to the nomenclature used in reference 59. doi:10.1128/microbiolspec.MGM2-0025-2013.f2

ASMscience.org/MicrobiolSpectrum
tandem-base mutations, shared by strains within particular subclusters of *M. tuberculosis* lineages (88). While these tandem-base mutations seem perfectly compatible with a strictly clonal evolution scenario, the remaining example of a SNP signature shared among *M. africanum* CPHL_A and the *M. tuberculosis* CDC1551- and C-strains might more likely be the result of a recombination event (88). Because the question of potential recombination episodes among different *M. tuberculosis* strains is also important for evaluating the risk of potential transfer of drug resistance mutations between *M. tuberculosis* strains, further functional studies are needed to unambiguously determine if and how *M. tuberculosis* strains might transfer DNA among each other. However, the previously reported findings did not concern larger genomic segments, such as the RD regions, which thus remain stable markers for differentiating phylogenetic lineages of the *M. tuberculosis* complex (57, 79, 89).

In contrast to *M. tuberculosis* and the classical members of the *M. tuberculosis* complex, the situation with horizontal gene transfer and interstrain-recombination is clearly different for *M. canettii* strains, whose main phenotypic characteristics are their—for tubercle bacilli—unusual, smooth colony morphology and their somewhat faster growth (35, 90, 91). Such smooth tubercle bacilli were first observed in the 1970s by Georges Canetti from a patient with pulmonary tuberculosis and were later named *M. canettii* (49, 92, 93). However, these strains are very rare and geographically restricted. Patients infected with *M. canettii* mostly originate from or have contact with the region of the Horn of Africa. The majority of the less than 100 worldwide available *M. canettii* isolates come from patients from Djibouti, East Africa (35, 55, 91, 94), which corresponds to the same geographical region where the above-mentioned, early branching *M. tuberculosis* strains of lineage 7 were isolated (80). From the 16S rDNA sequence and genome sequencing data, it is clear that the *M. canettii* strains share more than 97 to 99% DNA similarity with *M. tuberculosis* strains and thus can be considered as belonging to the same bacterial species (35, 95). However, when considering the SNP levels reported for *M. tuberculosis*, *M. africanum*, and *M. bovis* genomes in the maximal range of around 2,200 to 2,350 SNPs (12, 60, 85), the observed 16,000 to 65,000 SNPs among *M. tuberculosis* and *M. canettii* strains clearly demonstrate the much greater genetic diversity and phylogenetic distance of the smooth tubercle bacilli compared to the classical members of the *M. tuberculosis* complex (35) (Fig. 2).

The smooth tubercle bacilli share almost the same core genome as *M. tuberculosis* complex members but individually contain many additional genes contributing to an ~20% larger pan-genome of tubercle bacilli than was previously known (35). These findings suggest that the genomes of *M. canettii* strains might mirror the ancestral gene pool of those tubercle bacilli from which the members of the classical *M. tuberculosis* complex evolved by clonal expansion. This model is in agreement with the finding that several previously identified interrupted coding sequences of *M. tuberculosis* (56, 96) were found intact in the *M. canettii* strains as well as in the *M. marinum* and/or *M. kansasii* outgroups, suggesting that the *M. tuberculosis* lineages diverged from the smooth strain lineages before these frameshift mutations had occurred in the most recent common ancestor of the classical *M. tuberculosis* complex (35). Similarly, the RD9 and Tbd1 regions are also preserved in smooth strains. Another striking feature of the *M. canettii* genome sequences was the identification of prophages and clustered, regularly interspaced short palindromic repeat (CRISPR) systems that were completely different from those present in the classical members of the *M. tuberculosis* complex strains (35, 97), suggesting the continued genetic changes after the branching of the lineage leading to the classical *M. tuberculosis* complex from the smooth tubercle bacilli.

The large genetic diversity found in smooth strains, all originating from a restricted geographical location, i.e., Djibouti, and the highly reduced diversity within the globally spread *M. tuberculosis* strains and *M. tuberculosis* complex members suggest that the close ancestors of tubercle bacilli emerged in East Africa and that a dominant clone, later diversifying into the classical members of the *M. tuberculosis* complex (*M. tuberculosis*, *M. africanum*, *M. microti*, *M. bovis*) then spread throughout the world, possibly carried by waves of human migration. Moreover, from inspection of the genome sequences of *M. canettii* strains it is apparent that a substantial part of the genetic diversity in *M. canettii* strains is due to interstrain recombination events and horizontal gene transfer (35, 55). For the moment, it remains unclear where, when, and by which mechanisms these apparent transfers might have been achieved. The presence of homoplastic regions in different *M. canettii* genomes, in some cases also involving sequences characteristic for *M. tuberculosis*, observed at multiple loci, suggests that smooth tubercle bacilli might have (had) frequent occasion to exchange or uptake DNA with or from other strains (35, 55). It is tempting to speculate that distributive conjugation, recently described for the
fast-growing *M. smegmatis* (98), could be involved. However, in *M. smegmatis*, conjugation is enabled by the ESX-1 secretion system of the bacterium, which is present and highly conserved in both *M. canettii* and *M. tuberculosis* and is responsible for secreting proteins involved in virulence, such as the early secreted antigenic target ESAT-6 (99). Alternatively, the specific presence of a large putative complete phage-encoding region in at least one *M. canettii* strain (35) suggests a possible involvement of phages in the horizontal gene exchanges. Thus, more experimental work is needed to get deeper insight into this matter for tubercle bacilli.

Apart from genetically diverse regions, the genome project of smooth tubercle bacilli also identified genomic regions that are largely conserved in *M. canettii* and *M. tuberculosis* strains. Among these, genes encoding proteins harboring recognized human T-cell epitopes have dN/dS ratios that are on average lower than those of genes classified as nonessential and similar or slightly lower than those of the essential genes (Table 1) (35).

Overall, these observations are in agreement with results from Comas and colleagues (60), who first reported that human T-cell antigens were more conserved in *M. tuberculosis* strains relative to the rest of the proteome, suggestive of purifying selection acting on these epitopes. While these results could be interpreted to suggest that *M. canettii* might benefit from recognition by human T cells, as was proposed for *M. tuberculosis* (60), an alternative interpretation might be more plausible. It could well be that the conservation of the concerned proteins is caused by purifying selection due to specific, important functions for the mycobacterial cell (e.g., cell wall functions, secretion systems involved in environmental competition) despite nonessentiality for in vitro growth. Indeed, many of the T-cell-epitope-containing proteins of *M. tuberculosis* are also conserved in mycobacteria that do not have a long-lasting coevolution/host-pathogen interaction with humans. As an example, the antigen 85 complex, well known for its strong recognition by human T cells, is also highly conserved in *M. marinum* and other mycobacterial species (100). The same is true for ESAT-6 and CFP-10, which are more than 90% conserved in *M. marinum* and some other mycobacteria not necessarily in contact with humans (15). Why the *M. tuberculosis* orthologs were targeted by the human immune system remains an open question; however, their importance for the mycobacterial cell and the implied conservation and reduced possibility of escape variants might have played a role.

Finally, *M. canettii* strains were found to be less virulent in two different mouse models and also showed reduced persistence compared to *M. tuberculosis* (35). It should be mentioned that *M. canettii* strains can cause very serious disease in humans (91–93). However, it is intriguing that the lower virulence in mice seems to correlate with the epidemiological situation in humans; i.e., the potentially less virulent *M. canettii* strains cause very few human infections compared to the millions of tuberculosis cases caused by the more virulent *M. tuberculosis* strains. If we consider *M. canettii* strains as the closest relatives of *M. prototuberculosis*, the proposed ancestor of the *M. tuberculosis* complex, it seems likely that the highly prevalent *M. tuberculosis* strains have emerged from low-virulence, potentially environmental, smooth strains by gaining additional virulence and persistence factors (35, 55, 91, 95).

This scenario also fits well with the above-described comparison of *M. tuberculosis* with *M. marinum* (Fig. 1), where the closest related, nontuberculous mycobacteria reside in the aquatic environment, possibly in contact with aquatic protozoa as potential hosts. The interaction with these organisms could have favored

### TABLE 1 Ratios of nonsynonymous versus synonymous SNPs in gene categories

<table>
<thead>
<tr>
<th>Strain</th>
<th>dN/dS in gene category&lt;sup&gt;a&lt;/sup&gt;</th>
<th>All</th>
<th>Essential</th>
<th>Nonessential</th>
<th>T-cell antigens</th>
<th>T-cell epitopes</th>
</tr>
</thead>
<tbody>
<tr>
<td>STR-A</td>
<td>0.19/0.15</td>
<td>0.14/0.11</td>
<td>0.21/0.17</td>
<td>0.14/0.12</td>
<td>0.18/0.14</td>
<td></td>
</tr>
<tr>
<td>STR-J</td>
<td>0.18/0.13</td>
<td>0.14/0.11</td>
<td>0.19/0.15</td>
<td>0.14/0.11</td>
<td>0.13/0.09</td>
<td></td>
</tr>
<tr>
<td>STR-D</td>
<td>0.20/0.16</td>
<td>0.16/0.12</td>
<td>0.22/0.17</td>
<td>0.15/0.12</td>
<td>0.10/0.08</td>
<td></td>
</tr>
<tr>
<td>STR-L</td>
<td>0.19/0.15</td>
<td>0.16/0.12</td>
<td>0.21/0.17</td>
<td>0.14/0.12</td>
<td>0.15/0.11</td>
<td></td>
</tr>
<tr>
<td>STR-K</td>
<td>0.17/0.13</td>
<td>0.14/0.10</td>
<td>0.19/0.15</td>
<td>0.15/0.12</td>
<td>0.13/0.09</td>
<td></td>
</tr>
<tr>
<td>MTBC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>ND</td>
<td>0.53</td>
<td>0.66</td>
<td>0.50</td>
<td>0.53–0.25&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>STB, smooth tubercle bacilli; MTBC, *M. tuberculosis* complex; ND, not done.

<sup>b</sup>dN/dS ratios were calculated on orthologs conserved in all smooth tubercle bacilli strains and *M. tuberculosis* H37Rv, based on pairwise, concatenated codon alignments and using SNAP (value on the left) and PAML maximum likelihood methods (value on the right) as reported in reference 35. *M. tuberculosis* H37Rv T-cell antigen, essential, and nonessential gene categories, as well as T-cell epitope codon concatenates, were constructed as in Comas et al. (60).

<sup>c</sup>dN/dS ratios calculated by Comas et al. (60) from SNPs identified across 21 MTBC strains.

<sup>d</sup>Lower value obtained after exclusion of epitopes of three antigens considered as outliers.

---

*ASMscience.org/MicrobiolSpectrum*
the development or adaptation of traits that have later also enabled coping with the intracellular environment in macrophages. These thoughts are, of course, still highly speculative, although the availability of the various genome and physiopathology data allows new insights into the genetic factors that might have contributed to the development of *M. tuberculosis* becoming an obligate, highly efficient pathogen. The availability of these data provide an excellent opportunity to design further experimental work in order to confirm or dismiss such hypotheses.

**MYCOBACTERIAL SYSTEMS DISCOVERED BY GENOMICS THAT ARE INVOLVED IN PATHOGENICITY**

Progress in mycobacterial genome research has enabled the development of genome-wide methods for identifying genes implicated in *in vitro* and *in vivo* growth. One of the first studies in this respect was carried out by Sassetti and colleagues, who adapted the mariner transposon for high-density mutagenesis in *M. tuberculosis* and developed a method named transposon site hybridization (TraSH). By this method these authors identified around 700 genes that were essential for optimal growth of *M. tuberculosis* H37Rv under *in vitro* conditions (101). In a similar approach, Lamichhane and coworkers identified numerous essential genes in the *M. tuberculosis* CDC1551 strain (102). In addition, almost 200 genes essential for survival under *in vivo* conditions in the mouse infection model were found by adaptation of the TraSH screen to select for *in vivo* growth mutants (103). These findings extended the previously established list of genes identified by gene knockout and signature tagged mutagenesis (104–107). The TraSH approach was recently refined by using next-generation sequencing (NGS): 453 genes were identified as essential for *in vitro* growth both by microarrays and NGS. However, the NGS-based study suggested around 700 genes that were essential for *in vitro* and *in vivo* growth, depending on the culture media used (44).

In parallel to gene knockout and transposon-insertion approaches, comparative genomics of virulent and attenuated strains from the *M. tuberculosis* complex also provided highly relevant information for the identification of virulence gene clusters. As mentioned above, initial studies identified genomic polymorphisms across different members of the complex. Certain RDs were present in the genome of *M. tuberculosis* H37Rv but missing from attenuated strains (39, 108, 109). As the most prominent example, the 9.5-kb ESAT-6-encoding RD1 region is deleted from the genome of *M. bovis* BCG (RD1*BCG*), the only currently used attenuated antituberculosis vaccine (108). However, an RD1*BCG* overlapping portion (RD1*mic*) is also deleted from *M. microti* strains (110), which were also used as live attenuated vaccines in the 1960s (111, 112). Complementation of BCG and *M. microti* with the extended RD1 locus from *M. tuberculosis* restored ESAT-6 secretion and partially increased the virulence of recombinant BCG and *M. microti* strains (8, 113, 114). Together with results from *M. tuberculosis* RD1 deletion/knockout mutants (115–117), it became clear that RD1 encoded proteins have an important impact on *M. tuberculosis* virulence. Their loss thus seems to have played a key role in the attenuation of BCG- and *M. microti*-based vaccines. However, it should be mentioned that some other rare members of the *M. tuberculosis* complex, similar to *M. microti*, seem to have naturally lost overlapping portions of the RD1 region, as reported for the dassic bacillus (69) and for *M. mungi* (67).

Soon after the genomic identification of the RD1 encoded gene cluster, experimental evidence for a novel specialized secretion system was obtained. The system was named the ESAT-6 secretion system (ESX-1) (113, 118, 119), the Smm system (117), and more recently, the type VII secretion system (120, 121). The four ESX-1 paralogs found in *M. tuberculosis* were named ESX-2 to ESX-5 (10, 37, 47). Because it was first discovered and shown to be a key virulence factor, ESX-1 can be considered the prototype of ESX systems (Fig. 3). Most of the current knowledge on ESX secretion comes from the characterization of the ESX-1 secretion machinery, although more mechanistic data became recently available for the ESX-5 system (30, 122). Each ESX secretion apparatus is a multiprotein complex, consisting of so-called ESX conserved components (Ecc) and ESX-secretion-associated proteins (Esp) as well as Esp and PE/PPE proteins (121). It is constituted by cytosolic and membrane-anchored ATP-binding proteins (EccA and EccC, respectively) and other proteins containing a number of transmembrane domains (EccB, EccD, EccE), which are thought to mediate the ATP-dependent export of ESX substrates across the cytoplasmic membrane (114, 121) (Fig. 3). To date, the protein pores responsible for translocation of ESX substrates across the mycobacterial outer membrane have not been experimentally identified. However, it has been recently proposed that one of the components of the membrane-anchored complex, namely EccE or EccC, might span both the inner and outer membrane, thus forming a
channel in the mycomembrane \( (122) \). All ESX systems also include membrane-bound mycobacteria-specific subtilisin-like serine proteases called mycosins \( \text{MycP}_1-\text{MycP}_5 \) \( (123, 124) \), which might be involved in proteolytic digestion of ESX substrates. The ESX-1-associated MycP1, the only mycosin characterized so far, plays a key role in modulation of ESX-1 secretion activity, via the proteolytic digestion of the ESX-1 substrate EspB \( (125) \).

A typical hallmark of Esx proteins and other ESX substrates is the absence of a classical N-terminal signal sequence for secretion \( (99, 126, 127) \). A conserved secretion signal has been identified in CFP-10 (EsxB), which is required for CFP-10 being recognized by the FtsK-SpoIIIE ATPase EccC and thus for targeting the protein together with its protein partner ESAT-6 (EsxA) to the corresponding ESX secretion machinery \( (128) \). This observation was recently extended by defining a
C-terminal domain (YxxxD/E) as a general sequence required for targeting proteins to the ESX/type VII secretion machineries (129). To date, the signal sequences that specifically target each ESX substrate to the corresponding ESX secretion apparatus are still unknown, but it has been hypothesized that selected EspG proteins (e.g., the M. marinum ESX-1- and ESX-5-encoded EspG) might act as chaperons in directing type VII substrates to the corresponding ESX machineries (130). In this respect it is intriguing that gene inactivation of EspG1 from the M. tuberculosis ESX-1 system results in attenuation of the mutant strain, although the secretion of two of the main substrates, ESAT-6 and CFP-10, was not impaired by the deletion process (131). The involvement of an ATPase in the contact between Exs substrates and the secretion apparatus shows similarity with type IV secretion systems in Gram-negative bacteria, where a membrane-bound SpoIIIE/FtsK-like ATPase recognizes an unstructured C-terminal sequence and directs the secreted substrate to the cytoplasmic membrane (132).

Phylogenetic analyses and comparative genomics suggest that the ESX-4 cluster is the most ancestral ESX locus in the genus Mycobacterium (47). The other ESX loci are supposed to have evolved from ESX-4-like systems by gene duplication events and, in some cases, insertion of additional genes, with the order ESX-1, ESX-3, ESX-2, and ESX-5 (47). One of the most important gene insertion events is related to pe and ppe genes, which are absent at the ESX-4 locus and have been inserted during and after the ESX-1 diversification. It is thought that the ESX system contained pe and ppe genes that are the ancestral representatives of these gene families, from which the other genes have evolved (32).

Although different segments of the ESX-1 locus are deleted in some strains and/or lineages of the M. tuberculosis complex (e.g., BCG, M. microti, the dassie bacillus, M. mungi) (67, 69, 108, 110), ESX-1 is highly conserved in M. tuberculosis as well as in the early branching, M. canettii strains (35), suggesting that the full ESX-1 encoding region is part of the ancestral genome organization of the tuberculosis-causing mycobacteria. In addition to the ESX-1 encoding genes at the RD1 region, the espACD locus, which is situated elsewhere in the genome, is also important for ESX-1 functions in M. tuberculosis (133, 134). This locus plays a fundamental role in modulation of ESX-1 (121, 133, 134), because the espACD gene products (ESX-1-associated EspA, EspC, and EspD proteins) are secreted in a codependent manner with ESAT-6 and CFP-10 (e.g., EspA) (134) or required for regulating the intracellular levels of other Esp proteins (e.g., EspD) (135). In contrast, the espACD operon is not necessary for the secretion of EspB, another ESX-1 substrate (136). The expression of genes of the espACD cluster is regulated by a mechanism that involves the activators PhoP, EspR, and/or MprAB (137–139). The association between the PhoP/PhoR two-component regulatory system and ESAT-6 secretion was demonstrated in the M. tuberculosis strain H37Ra (137), one of the most widely used attenuated variants of the H37 strain, which apparently lost its virulence during in vitro serial passages. Comparative sequence analysis of M. tuberculosis H37Rv and H37Ra genomes by microarray-based DNA sequencing techniques identified a point mutation in the predicted DNA binding region of the phoP gene in H37Ra (137), which interferes with the DNA binding capacities of the corresponding PhoP protein (140). The complementation of H37Ra with a wild-type copy of the phoP gene restored ESAT-6 secretion and partially increased the virulence of the recombinant H37Ra::phoP strain (137), thus demonstrating the role of PhoP as a two-component regulator acting indirectly on the modulation of ESX-1 secretion activity. More recently, the PhoP-mediated regulation cascade of ESAT-6 secretion was found to be linked with the nucleoid-associated regulator factor EspR (138), which seems to directly bind and activate the espACD operon (138, 141).

It is noteworthy that other ESX-1-associated Esp proteins have a species-specific impact on ESX secretion. This is the case for EspF1 or the above-mentioned EspG1: Unlike observations with the model organisms M. smegmatis or M. marinum, where orthologous genes are involved in EsxA/B secretion (142, 143), disruption of espF and espG1 in M. tuberculosis did not impact secretion, the post-translational modification, or T-cell recognition of ESAT-6/CFP-10, but still caused strong attenuation (131). These findings suggest that, despite the high homology of ESX systems in various mycobacterial species, substantial differences in the secretion machinery exist that might have evolved during the adaptation to the respective hosts (114, 131).

It has been recently reported that the presence of a functional ESX-1 system strongly influences the ability of M. tuberculosis to establish infection and its trafficking in host cells. Analysis by electron microscopy techniques of the subcellular localization of different mycobacterial species, carrying an intact ESX-1 system or not, demonstrated a link between ESX-1 and the ability of tubercle bacilli to escape from the phagovacuole of infected cells. While M. tuberculosis and a range of pathogenic mycobacterial species (M.
leprae, M. bovis, and M. marinum)—as well as recombinant BCG variants expressing an intact ESX-1 system—were reported to translocate from the phagosome into the cytosol of infected macrophages and dendritic cells, strains with an interrupted ESX-1 system remained enclosed in the phagovacuole of the cells (144–146).

An important factor for the ESX-1-mediated phagosomal rupture appears to be the ability of ESAT-6 to interact with biomembranes, causing their destabilization and lysis (115, 147). It was hypothesized that this process might be pH dependent (147), but it remains unclear if other factors are also involved, because the data obtained from spectroscopic analyses of recombinant ESAT-6 and CFP-10 preparations predicted stability of the complex even under low pH conditions (148). However, a recent study suggested that there were differences among ESAT-6 orthologs from pathogenic and nonpathogenic mycobacterial species in their ability to undergo a conformational change under acidic pH conditions (149). Furthermore, it is not known whether other proteins co-secreted by ESX-1 might also contribute to membrane perforation. For example, ESX-1 systems are widely distributed among pathogenic and nonpathogenic species, although the environmental, nonpathogenic mycobacterial species lack the EspACD region in contrast to pathogenic species (145). Thus, for the moment it remains unclear which effectors of the ESX-1 secretion systems contribute to the rupture of vacuoles containing pathogenic mycobacteria. It seems clear, though, that ESX-1-dependent pore formation and lysis of the vacuolar membrane allows pathogenic mycobacteria and/or bacterial components to gain access to the cytosolic compartment of the host cell at different stages of infection, which accounts for major differences observed between virulent M. tuberculosis and attenuated BCG, the latter lacking ESX-1 functions due to the RD1 deletion. Indeed, M. tuberculosis and BCG differ in a wide variety of features such as cell-to-cell spread (115, 150), apoptosis (151), autophagy induction, and/or impairment (152, 153), and access of mycobacterial proteins to the class I-processing machinery contained in the proteasome, with impact on NLRP3 inflammasome activation (154–156), type I interferon responses (157), and induction of CD8 T-cell responses (158).

Recent research demonstrated that in addition to ESX-1, ESX-5 is also a key virulence determinant of pathogenic mycobacteria. Insights into the functional role of the ESX-5 system have been obtained for M. tuberculosis and M. marinum (29–31, 159, 160), which harbors an orthologous ESX-5 system (47). Secretome analysis of a panel of M. tuberculosis and M. marinum strains, mutated in selected components of the ESX-5 system, demonstrated that ESX-5 mediates the secretion of EsxN/EsxM, the ESX-5-encoded EssA/EssB paralogs. ESX-5 also transports a number of PE and PE proteins, including the representative PE25-PPE41 proteins (29, 30, 122, 129, 160) that are co-secreted as a heterodimer, whose crystal structure is similar to that reported for the ESAT-6•CFP-10 complex. Further substrates of ESX-5 are proteins belonging to the PE-PGRS and PPE_MPTR subfamilies (3, 97), including the mycobacterial lipase LipY (51) that is involved in degradation of long chain triacylglycerols during late phases of infection (161). Species-specific or strain-specific differences among LipY orthologs of M. tuberculosis (LipYMt) and M. marinum (LipYMm) might exist due to the species-associated differences in glycolipid composition of the cell wall and differences in PE/PPE expression profiles (30, 122, 159). Interestingly, PE_PGRS and PPE_MPTR are the phylogenetically most recent subclasses of PE and PPE proteins, and their emergence and expansion are linked to the ancestral PE and PPE proteins encoded at the ESX-5 locus (32), the last ESX locus supposed to have appeared in the M. tuberculosis genome (see above). Consistent with its role as a secretion system specialized for the transport of PE and PPE proteins, ESX-5 is absent in fast-growing saprophytic mycobacterial species (47, 119), which possess only a very limited set of PE/PPE proteins. In contrast, it is present in various slow-growing human pathogenic species, such as M. leprae and M. ulcerans, which harbor a wide spectrum of PE/PPE proteins like M. tuberculosis.

An intact ESX-5 system is required for optimal in vitro growth of M. tuberculosis and is also essential for full virulence of tubercle bacilli. Initial attempts to delete a large portion of the ESX-5 locus that encodes key building blocks of the ESX-5 membrane-bound protein complex (e.g., the eccB5-eccC5 operon) were not successful. Deletion of the eccB5-eccC5 genomic segment was only obtained in a merodiploid strain, which carried an additional copy of eccB5-eccC5 genes inserted into the chromosome via an attB site (160). In contrast, disruption of another core component of the ESX-5 secretion apparatus, e.g., the predicted transmembrane channel EccD5, did not affect the growth properties of the mutant strain in liquid medium but caused strong attenuation of the mutant, which was not able to replicate in murine macrophages or in severely combined immunodeficiency (SCID) mice (30). The impact of ESX-5 on M. tuberculosis viability and virulence might be explained by the involvement of this system in maintaining cell
envelope stability and transport of PE and PPE proteins. ESX-5 inactivation causes extensive damage to the mycobacterial cell envelope, as revealed by the increased sensitivity of ESX-5 mutants to detergents and hydrophilic antibiotics to which mycobacteria normally are naturally resistant (30).

Interestingly, the putative nucleoid-associated regulator EspR binds to multiple sites in the ESX-1, ESX-2, and ESX-5 loci as well as upstream of a number of genes encoding enzymes involved in cell-wall biosynthesis (138). This finding provides evidence for a functional link between ESX-mediated secretion and cell-wall biogenesis. In addition, it is also possible that the perturbation of the transport of PE and PPE proteins caused by the inactivation of ESX-5 functions might contribute to the decreased viability of ESX-5 mutants: Accumulation of nonsecreted or falsely localized ESX-5 substrates might have a toxic effect on the mycobacterial cell and diminish fitness. To date, the role of many ESX-5 substrates on mycobacterial virulence has not been investigated. However, it seems clear that upon infection with M. tuberculosis the host mounts specific T-cell responses against ESX-5-associated PE/PPE proteins (PPE25, PE18, PPE26, PPE27, PE19) and a number of their non-ESX-5-encoded homologs (31), which are completely lost if the ESX-5 secretion system is non-functional, confirming a role for ESX-5 in the transport of these proteins during infection. ESX-5-encoded PE and PPE proteins have been demonstrated to be involved in virulence functions of M. tuberculosis. An M. tuberculosis Δppe25-pe19 mutant, deleted for the genomic segment encoding all ESX-5-associated PE/PPE proteins, is strongly attenuated in SCID mice and is unable to replicate and disseminate to the spleen in aerosol-infected immunocompetent C57BL mice (30, 31). The observed attenuated phenotype together with the retained ability of secreting important protective ESX-1 antigens makes this strain a potential candidate for further testing as a promising attenuated live vaccine (31). Although more and more aspects of ESX-related features have become known, further research is needed to better understand the many contributions of type VII systems to tuberculosis-related infection and immunity.

CONCLUSIONS AND PERSPECTIVES
The above-described proteins and systems that influence the outcome of M. tuberculosis infections represent only a small selection of the large spectrum of potential factors that intervene in host-pathogen interactions during the course of tuberculosis infection. There is a range of numerous lipids from M. tuberculosis that are important for the pathogenicity of M. tuberculosis (18, 162, 163), but this complex subject is beyond the scope of this chapter. It is also clear that many proteins other than those mentioned here affect the survival and in vivo growth of M. tuberculosis. Many of these additional virulence candidates were identified by the above-mentioned genetic knockout strategies such as gene deletion, transposon insertion, signature tagged mutagenesis, and TraSH screens (103, 105, 106, 164). It is likely that even more genes will be identified that might be specific for particular physiological states or growth phases, such as dormancy, resuscitation, etc., which are, however, more difficult to model or investigate (165–167). In addition, it is well known that besides the bacterial factors that contribute to virulence, the outcome of infection with M. tuberculosis also depends largely on the response of host immunity and underlying host genetics. These two substantial research domains are dealt with in more detail elsewhere and are not discussed further here.

We have described the evolution of M. tuberculosis and the genomic factors that seem to have played crucial roles as it has evolved to become such a devastating pathogen. The insights we are continuing to gain from genome data are substantial, particularly nowadays, because the cost to sequence a bacterial genome has become trivial. On this basis, we are living in a scientifically exciting era, progressively revealing long-hidden secrets and solving mysteries surrounding ancient mycobacterial diseases. However, the bottleneck now lies in the time and effort required for the analysis and experimental confirmation of hypotheses raised from genomic data. Our challenge is thus now to translate the new insights from genomics into practical approaches that have the power to prevent tuberculosis and other mycobacterial diseases. It remains to develop a novel vaccination concept that is more protective than the widely used BCG vaccine (14, 113, 168). While some promising candidates are entering clinical trials (169, 170), their protective efficacy in humans still needs to be confirmed. The example of the functional characterization of the ESX-5 locus shows the interest of rational pathogenomic approaches to developing potential additional vaccine candidates (see above).

Another real possibility by which progress in tuberculosis control can be made is the development of new antituberculosis drugs, which are urgently required for the treatment of the emerging drug-resistant variants of the tubercle bacillus. Several new molecules have been presented that are active against M. tuberculosis under in vitro and/or in vivo conditions (171–174).
Information on the genome content and “omic” approaches in general are helping to identify the molecular targets of these new molecules (172, 174–177). Preclinical and clinical investigations will then have to show whether some of the drug candidates can be confirmed and developed into new drugs (173, 178, 179). The road ahead appears long, but the darkness has begun to lift.

ACKNOWLEDGMENTS
Special thanks to our many colleagues who contributed the various original data summarized and discussed here.

This work was supported in part by the Fondation de Recherche Medicale FRM DEQ20130326471, the European Community’s Framework Programme 7 grants 241745 and 260872, the Institut Pasteur (PTR372 and 383), grant 09 BLAN 0400 01 from the Agence Nationale pour la Recherche (France), and the National Health and Medical Research Council of Australia.

REFERENCES
Mycobacterial Pathogenomics and Evolution


Mycobacterial Pathogenomics and Evolution


