# The nature and consequence of genetic variability within *Mycobacterium tuberculosis*

M. Kato-Maeda, … , B.N. Kreiswirth, P.M. Small

**Perspective**

Viewed in terms of its historical and contemporary disease burden, Mycobacterium tuberculosis (MTB) is undeniably the most successful human pathogen. Molecular paleoarchaeologic evidence clearly identified MTB as the cause of lesions in 5,000-year-old mummies. Over recorded history, the burden of MTB has been staggering. In Europe the mortality was between 200 and 300 per 100,000 at the beginning of the 19th century. Today, tuberculosis still remains an important cause of morbidity and has been displaced from the position of leading infectious cause of death only by the accounting policies of the World Health Organization, which attribute tuberculosis deaths to human immunodeficiency virus in the increasing number of coinfected persons. The future is not brighter. The most optimistic scenarios predict in excess of 80 million new cases and 20 million deaths in the coming decade, 95% of which will occur in resource-poor countries. After decades of neglect, a resurgence of tuberculosis in industrialized countries reinvigorated research efforts, and tuberculosis is now back on the cutting edge of science. Molecular epidemiologic approaches have provided novel insights into the transmission dynamics of tuberculosis and have helped to refocus and refine control practices. These studies highlighted old evidence that there is significant variability in the clinical and epidemiologic consequences of infection with MTB. In the near future, mycobacteria are likely to be one of […]

**Find the latest version:**

# The nature and consequence of genetic variability within *Mycobacterium tuberculosis*

M. Kato-Maeda,[1] P.J. Bifani,[2] B.N. Kreiswirth,[2] and P.M. Small[1]

[1]Divisions of Infectious Diseases and Geographic Medicine, Stanford University, Stanford, California, USA
[2]Public Health Research Institute (PHRI) Tuberculosis Center, PHRI, New York, New York, USA

Address correspondence to: Peter M. Small, Division of Infectious Diseases and Geographic Medicine,
Medical Center Room S-156, Stanford University School of Medicine, Stanford, California 94305, USA.
Phone: (650) 725-7908; Fax: (650) 498-7011; E-mail: peter@molepi.stanford.edu.

Viewed in terms of its historical and contemporary disease burden, *Mycobacterium tuberculosis* (MTB) is undeniably the most successful human pathogen. Molecular paleoarchaeologic evidence clearly identified MTB as the cause of lesions in 5,000-year-old mummies. Over recorded history, the burden of MTB has been staggering. In Europe the mortality was between 200 and 300 per 100,000 at the beginning of the 19th century. Today, tuberculosis still remains an important cause of morbidity and has been displaced from the position of leading infectious cause of death only by the accounting policies of the World Health Organization, which attribute tuberculosis deaths to human immunodeficiency virus in the increasing number of coinfected persons. The future is not brighter. The most optimistic scenarios predict in excess of 80 million new cases and 20 million deaths in the coming decade, 95% of which will occur in resource-poor countries.

After decades of neglect, a resurgence of tuberculosis in industrialized countries reinvigorated research efforts, and tuberculosis is now back on the cutting edge of science. Molecular epidemiologic approaches have provided novel insights into the transmission dynamics of tuberculosis and have helped to refocus and refine control practices. These studies highlighted old evidence that there is significant variability in the clinical and epidemiologic consequences of infection with MTB. In the near future, mycobacteria are likely to be one of the most deeply sequenced pathogenic groups. The complete genomes of two MTB isolates are now available, and the sequencing of isolates of *M. bovis*, bacille Calmette-Guérin (BCG), *M. smegmatis*, *M. paratuberculosis*, *M. leprae*, *M. marinum*, *M. ulcerans*, and *M. avium* are somewhere between contemplated and completed.

> **The most optimistic scenarios predict in excess of 80 million new cases and 20 million deaths in the coming decade.**

Although much is known about environmental and host factors that contribute to infection and disease variability, little is know about the bacteria's role. Given recent advances, we are now poised to explore the nature and consequences of genetic variability in MTB. Here, we review the data supporting phenotypic and genotypic diversity of laboratory and natural strains of MTB and our understanding of the relationship between them.

## Phenotypic diversity

A hallmark of the natural history of tuberculosis has long been variability in the outcome of exposure. Only 30% of exposed persons show evidence of infection, and, of those infected, only 10% become ill. Among those who become ill, there is variability in disease time course, severity, and anatomic distribution. The prevailing dogma is that this variability is due to host and environmental factors, but there is considerable evidence that bacterial determinants contribute as well.

*Comparison of laboratory strains in vivo and in vitro.* Early on, the availability of laboratory-attenuated MTB showed that there could be variability between strains. In 1905, MTB strain H37 was cultivated from the sputa of a 19-year-old male with pulmonary tuberculosis. Subsequently, it was grown in different media and used in guinea pig and rabbit experiments. Beginning in 1922, it was noted that some isolates, whose colony morphology had changed, were not able to produce disease. This attenuated progeny came to be known as H37Ra. Somewhat analogously, Calmette and Guérin passaged one *M. bovis* isolate (which causes tuberculosis in cattle) 230 times, to generate the attenuated bacterium known as BCG. This live attenuated organism is the world's most widely used vaccine.

Studies comparing H37Rv, H37Ra, and BCG have identified phenotypic differences in vivo and in vitro. For instance, the number of bacteria recovered from the lungs and spleens of infected guinea pigs after inoculation with H37Ra is 100-fold less than with H37Rv at 3 weeks and declines rapidly thereafter (1). In other studies, H37Ra fails to grow progressively in

severe combined immunodeficient mice unless the animals are treated with hydrocortisone (2). Virulence differences have been investigated in monocyte and macrophage models. A key step in the immune response is phagocytosis of MTB by alveolar macrophages, a process mediated in part by mannose receptor binding to a bacterial surface lipoglycan, lipoarabinomannan (LAM). Differences in mannose receptor–mediated uptake of LAM from H37Rv, H37Ra and other strains have been related to the tertiary structure of LAM, the length of capped arabinofurans, and the exact location of LAM in the cell wall of the bacteria (3). The efficiency of phagolysosome fusion and the rate of bacterial proliferation inside the macrophage also varies among these strains. McDonough et al. (4) reported that the number of intracellular H37Rv increases more than tenfold over 3 days, considerably more growth than is observed for BCG or H37Ra (4). In other studies, the intracellular doubling time in human-blood monocytes was shorter for H37Rv (36 hours) when compared with BCG (89 hours) and H37Ra (247 hours) (5). Production of TNF-α by infected monocytes is correspondingly high for H37Rv, intermediate for BCG, and low for H37Ra. TNF-α–dependent cytotoxicity is not observed in cells infected with the latter strains (6).

*Well characterized clinical isolates.* Efforts to compare virulence of clinical isolates of MTB date back to the 1950s. Dubos described strain differences in the ability to survive oxygen deprivation or excess, and concluded that a wide spectrum of virulence phenotypes exist (7). Certain isolates proved relatively poor at causing disseminated disease in guinea pigs, showed enhanced susceptibility to hydrogen peroxide in vitro (8) and a distinctive capsular lipid composition (9). Moreover, MTB isolates from patients with similar characteristics had remarkable variation in transmission rates when studied in animal models, again consistent with the idea of intrinsic biological differences between strains (10).

More recently, molecular strategies to distinguish naturally occurring clones of MTB have refocused attention on strains thought to be highly virulent, such as strain CDC1551 (11), or very prevalent, such as C strain (12), W or Beijing strain family (13, 14), subpopulations of the W family–like strain 210 (15), and New Jersey W4 strain (16).

Strain CDC1551 was isolated in an outbreak that occurred between 1994 and 1996 in a rural community with low risk of tuberculosis. On average, 80% of the contacts of patients with disease caused by this strain had tuberculin skin test reactivity, an observation that contributed to the mistaken notion that this strain is unusually virulent. Initial mouse infection studies demonstrated that CDC1551 replicates more rapidly in the lung than another reference strain Erdman (11). However, subsequently it was shown that CDC1551

replicates to a similar level as H37Rv, *M. bovis* (17), and other clinical isolates (18), and that it even grew more slowly than these other strains at later time points during infection. Patterns of host responses to these bacteria may be better markers of virulence than simple growth rate. In fresh human monocytes, CDC1551 induces higher TNF-α and IL-10 levels than does H37Rv (18), and Bishai et al. (19) found that the number of grossly visible primary tubercles was similar in rabbits infected with H37Rv or CDC1551, but that the latter strain produced smaller tubercles with fewer bacilli, perhaps a result of this more robust cytokine response (19). Indeed, mice survive CDC1551 infections longer than infections with other strains, again consistent with the conclusion that strain CDC1551 induces a stronger and earlier immune response in experimental animals than do other isolates. The high rate of tuberculin skin test reactivity among contacts of persons with disease due to CDC1551 suggests that the same is true in humans.

The C strain is a drug-susceptible organism prevalent in New York City that was responsible for a large outbreak in men's shelters in Manhattan. In one population-based study, this strain caused more than 20% of the drug-susceptible tuberculosis cases in New York and was very frequently isolated from users of injected drugs. In vitro studies demonstrated that this strain is particularly resistant to reactive nitrogen intermediates (RNI), which arise from nitric oxide (NO) following the induction of cytokines in macrophages. The relative resistance in vitro to RNI correlates well with the virulence of the C strain in guinea pigs. Friedman et al. (12) have postulated that injection drug users have high levels of NO and that the prevalence of this clone in such patients results from its capacity to resist NO-mediated clearance (12).

Genotyping of MTB isolates since the early 1990s have repeatedly identified a group of related organisms responsible for a high percentage of disease in certain populations and numerous outbreaks. These phylogenetically linked isolates belong to the so-called W-strain family, also known as the Beijing strain. The W strain refers to a group of clonally related multidrug resistant organisms that caused very large HIV-linked nosocomial outbreaks in New York City between 1991 and 1994 (16). These organisms carry identical mutations in *rpo*B, *rps*L, *kat*G, and *emb*B genes, causing resistance to rifampin, streptomycin, isoniazid, and ethambutol, respectively, and have evolved from a single MTB strain that has undergone sequential selection for resistance conferring mutations in several genes. The Beijing strain refers to an endemic cluster of related MTB genotypes causing disease in Beijing, China. This strain was first described in 1992, and subsequent studies found it to be genetically similar to the W strain (14). All members of the W-strain family and the Beijing

strain belong to principal genetic group 1 of MTB and share characteristic genetic fingerprints (20).

W-strain family isolates have caused outbreaks and established themselves in various localities, and from there they continue to disseminate and evolve multiple new progeny genotypes. The New Jersey W4 strain group (16) and the Los Angeles strain 210 (21) are well documented representative W variants, which have undergone recent clonal expansion. The molecular genetic mechanism underlying the successful dissemination of W-family organisms remains unknown, but the growth of these organisms in human macrophages is four- to eightfold higher than seen with MTB isolates that cause few or no secondary cases of tuberculosis. In vitro analysis found no differences in cytokine production or susceptibility to RNI when compared with other clinical isolates. These data suggested that the ability to replicate in human macrophages might contribute to the organisms' fitness in the population (15).

### Genetic variability

Advances in molecular biology and functional genomics have fostered major advances in our understanding of genetic variability among MTB. Automated DNA sequencing allows for the direct comparison of specific genes among large populations of isolates and the determination of the complete genomic sequence of two strains (http://www.tigr.org/tdb/CMR/gmt/htmls/Splash-Page.html) (22). DNA sequence information, together with new technologies like DNA microarrays and computational biology, have provided a whole-genome perspective on genomic content, gene regulation, and MTB metabolism.

*Genotypic diversity among laboratory strains*. The clear phenotypic differences between H37Ra and H37Rv and between BCG and *M. bovis* have inspired several groups to seek the genomic differences in hope of explaining the attenuation of H37Ra and BCG.

A 25 kb DNA fragment called *ivg* (in vivo growth-promoting locus) has been identified in H37Rv but not in the attenuated H37Ra (23–25). In addition, a 7.9 kb fragment containing genes coding for a putative sugar transferase, oxidoreductase, and a membrane protein (also present in *M. bovis*, therefore it likely plays no role in virulence) occur in H37Ra but not H37Rv (24, 25). Interestingly, the introduction into H37Ra of genes that are restricted to H37Rv does not correct the attenuated phenotype of H37Ra (23, 24).

There are several genomic differences between BCG and *M. bovis* that could explain the attenuation of BCG. One is RD1, a fragment of DNA containing regulatory genes that, when reintroduced into BCG, repress the expression of certain proteins. This result suggested that the loss of virulence was due to a regulatory mutation (26). More recently, Behr et al. (27) employed a microarray representing essentially all open reading frames (ORFs) from MTB strain H37Rv to compare the genomic content of MTB, *M. bovis*, and the family of BCG vaccine strains. This analysis showed that *M. bovis* lacks 11 regions, comprising 61 ORFs and 54,000 bp, that are present in H37Rv and that in addition to these deletions, BCG isolates uniformly lack one region and are polymorphic for four other deletions (27). The plot of these polymorphic deletions on a historical genealogy of BCG (reconstructed from the historical record) demonstrate the molecular events of a half century of in vitro evolution. Study of the biological and immunological impact of these deletions may well provide insights into the phenotypic consequences of mycobacterial mutations.

Similar analyses using bacterial artificial chromosome libraries identified two fragments of DNA missing in H37Rv but present in *M. bovis*. These regions include four ORFs encoding unknown proteins, as well as an oxidoreductase and a membrane protein (25). In summary, although some genetic differences have been described, the basis of the attenuation of H37Ra and BCG (or conversely, the virulence of H37Rv and *M. bovis*) remains unclear.

*Genomic diversity in clinical isolates*. In the past decade, polymorphisms in the number and genomic location of mycobacterial repetitive elements have been widely exploited in epidemiologic studies. Although the MTB genome contains several repetitive elements, only a few are polymorphic and widely studied (28). Four of the better characterized such polymorphisms are IS*6110* insertions, polymorphic guanine cytosine–rich (GC-rich) repetitive sequences (PGRSs), direct repeat (DR), and the recently identified mycobacterial interspersed repetitive units (MIRUs).

IS*6110* is an insertion element that appears to be confined to members of the MTB complex and is present in over 99% of clinical isolates. All copies of IS*6110* are nearly identical in sequence, but their copy number and location in the genome vary, leading to different restriction fragment-length polymorphism (RFLP) patterns (29). This method demonstrated genetic variability among MTB and, in conjunction with clinical epidemiologic data, has been used to track the transmission of MTB in communities (30).

The PGRS element appears to be the most abundant repetitive sequence in MTB. It occurs multiple times and has significant polymorphism. This sequence is characterized by a high GC content (80%) and has been successfully used as a secondary marker to discriminate among MTB isolates with a low number of IS*6110* copies (31).

The DR is composed of multiple numbers of a 36 bp sequence that are interspersed by unique spacers 35 to 41 bp in length. One DR and the adjacent spacer sequence are called direct variable repeat (DVR). Groenen et al. have developed a reverse blot technique,

termed spoligotyping, to determine the presence or absence of 43 known DVR polymorphisms, a technique that can be used to differentiate strains of MTB (32). DVR polymorphisms appear to evolve more slowly than changes in IS*6110* and PGRS patterns, and have been used to reconstruct a conserved sequence that is presumed to have existed in an ancestral strain that evolved by a series of deletions to generate the current set of polymorphic isolates (33).

MIRUs consist of repetitive elements found in tandem repeats scattered throughout the genome. Comparative sequence analysis of strain H37Rv, CDC1551, *M. bovis* strain AF212/97, and 31 MTB clinical isolates identified 41 MIRUs and showed that at least 12 are polymorphic. Variations in nucleotide sequence and copy number of the tandem repeats have been identified. The five exact tandem repeats (ETR-A to ETR-B) known as variable number of tandem repeats, previously used as a genotyping tool, are also part of the 41 MIRUs. It has been suggested that MIRUs are reminiscent of human minisatellites, which are often hypervariable sequences consisting of 10–100 bp repeats believed to have evolutionary function (34).

In contrast to the high level of diversity found in these repetitive elements, there is an extremely limited amount of unselected nucleotide sequence variation in structural genes. Sequence analysis of two megabases of 26 structural genes from a phylogenetically diverse group of MTB isolates revealed only 32 polymorphic nucleotide sites not directly selected by antibiotic resistance (35). Thirty of the thirty-two polymorphic sites were synonymous changes, and the most common polymorphisms occurred at *katG* codon 463 and *gyrA* codon 95. A trichotomous classification system based on the distribution of the polymorphisms at these two loci showed significant IS*6110* variation among the isolates in each group, indicating that IS*6110* variation arose after the divergence in the three principal genetic groups. Moreover, *M. africanum*, *M. bovis*, and *M. microtii* isolated from various hosts also lacked significant levels of unselected nucleotide variation. A more recent study of 26,700 bp representing 24 genes coding for proteins that are targets of the host immune system confirmed the absence of significant nucleotide polymorphism (36). The lack of silent nucleotide polymorphisms in MTB has been interpreted as indicating that MTB is a relatively young species, perhaps 15,000 to 20,000 years old. Alternative explanations, such as enhanced fidelity of DNA replication or clonal sweeps purging variability have been speculated but have no supporting evidence.

Recently, DNA microarrays have been used to analyze 16 MTB clones for which clinical and epidemiological information was available. Deletion patterns were found to be conserved among isolates from a given clone but differed between clones. Each clone had an average of 0.3% of the genome deleted when compared with H37Rv, comprising an average of 17 deleted ORFs per clone. This preliminary observation suggests that genomic deletions are another common source of genomic variability (37).

The presence of abundant polymorphisms associated with repetitive elements, coupled with the restricted number of single nucleotide polymorphisms, indicate that transposition and homologous recombination play an important role in the generation of genomic diversity in MTB. In addition, polymorphisms found with different markers tend to show strong mutual association, supporting the contention that MTB has a strongly clonal population structure (35).

## Better informed, but how much smarter?

The recent accumulation of copious data about the extent and nature of genetic and phenotypic variation in MTB provides an unprecedented opportunity to determine the phenotypic consequences of genetic polymorphism in this organism. However, currently, a coherent synthesis of this knowledge is lacking, reflecting the formidable challenges of this endeavor. Because environmental and host factors clearly contribute to the clinical and epidemiologic behavior of strains, these factors must be carefully integrated into the investigative process. Accurately measuring laboratory and clinically relevant phenotypes is a significant challenge. For example, early reports of the unusually high growth rate of CDC1551 in mice appear in retrospect to be largely due to the relative attenuation of the comparison strains. Current in vitro and in vivo systems will need to be improved and new ones will need to be developed. Additionally, the clinical manifestations of infection with a particular strain need to be observed in multiple patients and settings before they can be ascribed with certainty to the organism. Furthermore, while statistically significant increases in the prevalence of a clone in a community may reflect increased fitness, it may also simply be the consequence of a founder effect. Our limited understanding of the genetics of MTB makes it difficult to predict which genetic polymorphisms may be of consequence. For example, while it is now relatively easy to detect gross genetic variability, mutations as minor as a single bp mutation in a single σ factor have been demonstrated to alter virulence in laboratory models.

Even when specific mutations are described one must be cautious about inferring specific phenotypic changes to it because of the possibility of other unrecognized mutations. Fortunately, in contrast to other pathogens, there appears to be limited unselected structure gene nucleotide variation in MTB, suggesting that causality may be easier to deduce than in more genetically variable pathogens. Furthermore, the population structure of MTB is remarkably clonal. The organism can be con-

ceptualized as a group of genes moving together through time and space. Thus, multiple observations of a clone can be meaningfully analyzed together.

The significant recent advances in our understanding of genetic and phenotypic variability in MTB also pose a great challenge. In the absence of carefully conducted, thoughtful analysis, we run the risk of spending decades making irrelevant observations about poorly characterized strains. Alternatively, if we succeed in identifying the genetic events that confer increased virulence, this information could profoundly change our approach to the control and prevention of tuberculosis.

1. Alsaadi, A.I., and Smith, D.W. 1973. The fate of virulent and attenuated Mycobacteria in guinea pigs infected by the respiratory route. *Am. Rev. Respir. Dis.* **107**:1041–1046.
2. North, R.J., and Izzo, A.A. 1993. Mycobacterial virulence. Virulent strains of Mycobacteria tuberculosis have faster in vivo doubling times and are better equipped to resist growth-inhibiting functions of macrophages in the presence and absence of specific immunity. *J. Exp. Med.* **177**:1723–1733.
3. Schlesinger, L.S., Kaufman, T.M., Iyer, S., Hull, S.R., and Marchiando, L.K. 1996. Differences in mannose receptor-mediated uptake of lipoarabinomannan from virulent and attenuated strains of Mycobacterium tuberculosis by human macrophages. *J. Immunol.* **157**:4568–4575.
4. McDonough, K.A., Kress, Y., and Bloom, B.R. 1993. Pathogenesis of tuberculosis: interaction of Mycobacterium tuberculosis with macrophages. *Infect. Immun.* **61**:2763–2773.
5. Silver, R.F., Li, Q., and Ellner, J.J. 1998. Expression of virulence of Mycobacterium tuberculosis within human monocytes: virulence correlates with intracellular growth and induction of tumor necrosis factor alpha but not with evasion of lymphocyte-dependent monocyte effector functions. *Infect. Immun.* **66**:1190–1199.
6. Balcewicz-Sablinska, M.K., Keane, J., Kornfeld, H., and Remold, H.G. 1998. Pathogenic Mycobacterium tuberculosis evades apoptosis of host macrophages by release of TNF-R2, resulting in inactivation of TNF-alpha. *J. Immunol.* **161**:2636–2641.
7. Dubos, R.J. 1955. Properties and structures of tubercle bacilli concerned in their pathogenicity. *Proceedings of the Symposia of the Society for General Microbiology.* 103–125.
8. Mitchison, D.A., Selkon, J.B., and Lloyd, J. 1963. Virulence in the guinea-pig, susceptibility to hydrogen peroxide and catalase activity of isoniazid-sensitive tubercle bacilli from South Indian and British patients. *J. Pathol. Bacteriol.* **86**:377–386.
9. Goren, M.B., Brokl, O., and Schaefer, W.B. 1974. Lipids of putative relevance to virulence in Mycobacterium tuberculosis: correlation of virulence with elaboration of sulfatides and strongly acidic lipids. *Infect. Immun.* **9**:142–149.
10. Sultan, L., et al. 1960. Tuberculosis disseminators. A study of the variability of aerial infectivity of tuberculous patients. *Am. Rev. Respir. Dis.* **82**:358–369.
11. Valway, S.E., et al. 1998. An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. *N. Engl. J. Med.* **338**:633–639.
12. Friedman, C.R., et al. 1997. Widespread dissemination of a drug-susceptible strain of Mycobacterium tuberculosis. *J. Infect. Dis.* **176**:478–484.
13. Bifani, P.J., et al. 1996. Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. *JAMA.* **275**:452–457.
14. van Soolingen, D., et al. 1995. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. *J. Clin. Microbiol.* **33**:3234–3238.
15. Zhang, M., et al. 1999. Enhanced capacity of a widespread strain of Mycobacterium tuberculosis to grow in human macrophages. *J. Infect. Dis.* **179**:1213–1217.
16. Bifani, P.J., et al. 1999. Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. *JAMA.* **282**:2321–2327.
17. North, R.J., Ryan, L., LaCource, R., Mogues, T., and Goodrich, M.E. 1999. Growth rate of mycobacteria in mice as an unreliable indicator of mycobacterial virulence. *Infect. Immun.* **67**:5483–5485.
18. Manca, C., et al. 1999. Mycobacterium tuberculosis CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J. Immunol.* **162**:6740–6746.
19. Bishai, W.R., et al. 1999. Virulence of Mycobacterium tuberculosis CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. *Infect. Immun.* **67**:4931–4934.
20. Kurepina, N.E. et al. 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS6110 elements in Mycobacterium tuberculosis: non-random integration in the dnaA-dnaN region. *Tuber. Lung Dis.* **79**:31–42.
21. Barnes, P.F., et al. 1997. Patterns of tuberculosis transmission in Central Los Angeles. *JAMA.* **278**:1159–1163.
22. Cole, S.T., et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature.* **393**:537–544.
23. Pascopella, L., et al. 1994. Use of in vivo complementation in Mycobacterium tuberculosis to identify a genomic fragment associated with virulence. *Infect. Immun.* **62**:1313–1319.
24. Brosch, R., et al. 1999. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain. *Infect. Immun.* **67**:5768–5774.
25. 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.
26. Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., and Stover, C.K. 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *J. Bacteriol.* **178**:1274–1282.
27. Behr, M.A., et al. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science.* **284**:1520–1523.
28. Poulet, S., and Cole, S.T. 1995. Repeated DNA sequences in mycobacteria. *Arch. Microbiol.* **163**:79–86.
29. van Embden, J.D., et al. 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**:406–409.
30. Kato-Maeda, M., and Small, P.M. 2000. How molecular epidemiology has changed what we know about tuberculosis. *West. J. Med.* **172**:256–259.
31. Chaves, F., et al. 1996. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of Mycobacterium tuberculosis. *J. Clin. Microbiol.* **34**:1118–1123.
32. Groenen, P.M., Bunschoten, A.E., van Soolingen, D., and van Embden, J.D. 1993. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis: application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**:1057–1065.
33. van Embden, J.D., et al. 2000. Genetic variation and evolutionary origin of the direct repeat locus of Mycobacterium tuberculosis complex bacteria. *J. Bacteriol.* **182**:2393–2401.
34. Supply, P., et al. 2000. Variable human minisatellite-like regions in the mycobacterium tuberculosis genome. *Mol. Microbiol.* **36**:762–771.
35. Sreevatsan, S., et al. 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. USA.* **94**:9869–9874.
36. 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.
37. Kato, M.M., et al. 2001. Comparative genomics of M. tuberculosis: a snapshot of bacterial evolution. *Genome Res.* In press.