Published in final edited form as:

Tuberculosis (Edinb). 2009 March ; 89(2): 114–119. doi:10.1016/j.tube.2008.09.004.

Does the *lipR* gene of tubercle bacilli have a role in tuberculosis transmission and pathogenesis?

Katherine D. Sheline¹, Anne M. France¹, Sarah Talarico¹, Betsy Foxman¹, Lixin Zhang¹, Carl F. Marrs¹, Joseph H. Bates^{2,3}, M. Donald Cave^{4,5}, and Zhenhua Yang^{1,*}

¹Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan

²Arkansas Department of Health, College of Public Health, University of Arkansas for Medical Sciences

³Department of Epidemiology, College of Public Health, University of Arkansas for Medical Sciences

⁴Department of Neurobiology and Developmental Sciences, College of Medicine, University of Arkansas for Medical Sciences

⁵Central Arkansas Veterans Healthcare Center, Little Rock, Arkansas

Summary

Mycobacterium tuberculosis lipases, a diverse class of enzymes involved in lipid metabolism, may have an important role in tuberculosis (TB) pathogenesis. We explored the association of large sequence polymorphism (LSP) in one of the *M. tuberculosis* lipase-encoding genes, *lipR* (Rv3084), with patient characteristics using a population-based sample of clinical isolates to elucidate the potential role of *lipR* in TB pathogenesis. LSP in *lipR* were found in 104 (15.6%) of 665 isolates, of which 96% belonged to principal genetic group 3. When linkage by molecular type and epidemiologic evidence were compared, molecularly clustered cases infected with a *lipR* LSP isolate were more often epidemiologically linked than clustered cases infected with a *lipR* wild-type isolate. Further epidemiologic and functional studies are necessary to determine if the association between this *lipR* LSP and recent transmission we identified in this population reflects a functional role of *lipR* in TB transmission and pathogenesis, the *lipR* polymorphism acting as a marker for strains that have been recently introduced into this geographical region, or a yet unidentified mechanism.

Keywords

lipR; large sequence polymorphisms; tuberculosis transmission and pathogenesis; molecular epidemiology of tuberculosis

Introduction

Tuberculosis (TB) remains a major public health problem worldwide. Identifying and analyzing *Mycobacterium tuberculosis* virulence factors is an important first step toward

Corresponding author: Zhenhua Yang, Epidemiology Department, School of Public Health, University of Michigan, 109 S. Observatory Street, Ann Arbor, MI 48109-2029. Phone: 734-763-4296; Fax: 734-764-3192; Email: E-mail: zhenhua@umich.edu.

Conflict of Interest Statement: None of the authors of this manuscript has any financial and personal relationships with other people or organizations that could inappropriately influence (bias) his or her work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

developing new tools for TB control. Molecular epidemiologic investigations that can characterize strain-specific patterns of transmission and disease in human populations present a powerful approach to assess the importance of putative *M. tuberculosis* virulence factors, and to generate hypotheses that may later be tested in functional studies.

M. tuberculosis is characterized by a complex cell envelope that is exceptionally rich in unusual lipids [1] and the genes involved in lipid metabolism and cell-envelope modification are increasingly recognized to play an important role in virulence [2]. This complex, lipid-rich cell envelope is important in the formation of aerosolized droplet nuclei of *M. tuberculosis*, as well as its survival in the air [3]. Components of the cell envelope are also central to the organism's ability to evade the host immune system and survive in the harsh intracellular environment [4,5]. Lipases may be important in the formation and modification of the cell envelope and the metabolism of host cell membrane lipids, and thus the persistence of *M. tuberculosis* during latency [6].

lipR, a lipase-encoding gene of *M. tuberculosis*, [5] is a part of the *mymA* operon (*Rv3083-Rv3089*) that may alter the envelope of the organism in the intracellular environment [2]. The *mymA* operon is significantly up-regulated under acidic conditions in vitro [2,7,8]. However, *lipR* is not an essential gene; isolates with a large sequence polymorphism (LSP) within the *mymA* operon resulting in the complete deletion of *lipR*, and the partial deletion of adjacent genes *Rv3083* and *Rv3085*, have been recovered from patients with active TB disease in San Francisco [9,10].

In a recent whole-genome DNA microarray- based comparison of two clinical isolates that were judged to have a similar opportunity for transmission, but which differed substantially in the extent of observed transmission, with one strain being found in only one case and the other strain being isolated from 24 cases during a eight-year study time period, *lipR* was identified as present in the more widely transmitted strain and absent in the less successful strain (authors' unpublished data).

To explore the association of *lipR* LSP with patient demographic, clinical, and epidemiological characteristics, thereby providing new information about the potential role of *lipR* in the pathogenesis of *M. tuberculosis*, we investigated the *lipR* LSP in a well characterized population-based collection of clinical *M. tuberculosis* isolates.

Materials and Methods

M. tuberculosis isolates and patient information

This study includes a total of 665 clinical isolates, representing 92% of 721 incident cases of culture-confirmed TB diagnosed in the state of Arkansas between January 1, 1996 and December 31, 2000, of which sufficient DNA was available for this study. Epidemiologic, demographic and clinical information was collected from the CDC Report of Verified Case of Tuberculosis form for each case. Molecular data including the results of restriction fragment length polymorphism (RFLP) with the *IS6110* probe, pTBN12 DNA fingerprinting, and principal genetic group as defined by Sreevatsan et al.[11] were obtained from previous studies of the authors [12,13]. The distribution of demographic and clinical variables for the isolates included in the study did not differ significantly from isolates excluded from the study due to insufficient DNA quantity or quality (chi square test, evaluated at an $\alpha = 0.05$ level). The study protocols and procedures for the protection of human subjects were approved by the Health Sciences Institutional Review Board of the University of Michigan.

PCR

The *lipR* gene region was amplified by PCR to detect LSP, using a pair of primers described previously [10]. Each reaction mixture was composed of reagents from the BD Advantage-GC 2 PCR kit (BD Biosciences Clontech, Palo Alto, CA) in the following volumes: 10 µL 5× reaction buffer, 5µL GC-Melt, 1µL 50× deoxyribonucleoside triphosphate mix, 1µL 50× BD Advantage 2 polymerase mix, 2µL each of 2 primers at 10µM, 4µL DNA template at 10ng/µL, and 25µL PCR-grade water. The thermocycler program sequence was: 94°C for 1 minute, 29 cycles of 94°C for 30 seconds, 68°C for 30 seconds, 72°C for 4.5 minutes, then 72°C for 10 minutes.

DNA sequencing

PCR products with a size indicated by 0.8% agarose gel electrophoresis to be smaller than the 3.6 Kb expected for the wild-type PCR product were double strand DNA sequenced to determine the size and locations of LSPs. DNA sequencing was conducted at the University of Michigan Sequencing Core, with the same primers used for the PCR. The size and location of LSPs in the *lipR* gene region were determined by comparing the *lipR* gene region sequencing results of the study isolates to the *lipR* sequence of the laboratory reference strain H37Rv [5], using Edit Seq 5.02 and the clustal W method in the MegAlign 5.01 programs (DNASTar Inc., Madison, WI).

Cluster Definition

Molecular typing techniques are commonly used to differentiate TB cases resulting from recent infection from cases resulting from the reactivation of latent infections [14-16], with “clustered” cases (cases infected with *M. tuberculosis* isolates exhibiting the same pattern by molecular typing) considered to reflect disease due to recently acquired infection. To seek to understand the epidemiological and clinical relevance of the *lipR* LSP among clinical isolates of *M. tuberculosis*, study isolates were defined as clustered or unique based on available IS6110 and pTBN12 genotyping data, according to the methods described previously [17, 18]. For isolates with < 5 IS6110-hybridizing bands, having identical IS6110 and pTBN12 RFLP fingerprints was used as the cluster definition. For isolates having > 5 IS6110 copies, the cluster definition was having identical IS6110 RFLP fingerprint or having only a one band difference in IS6110 RFLP fingerprints and having identical pTBN12 RFLP fingerprints.

Case clustering by molecular typing alone, which we refer to as the “conventional clustering definition”, may not necessarily reflect recent transmission [18,19]. Because clustered cases that occur within a short time interval are more likely to be related by recent transmission than are cases separated by a long time interval [13,20], we also examined the association between the *lipR* LSP and clustering using a 1-year time-restricted definition of clustering. By our time-restricted definition, a case was considered to be clustered only if a case with the same pattern by molecular typing was diagnosed within 365 days before or after the diagnosis date of the given case. By using this time-restricted definition, we aimed at capturing clusters that were most likely to represent current transmission, allowing the transmissibility of the related *M. tuberculosis* strains to be better assessed.

Epidemiologic Linkage Definition

The identification of a potential transmission connection between two cases with clustered isolates provides compelling evidence that the given cases are truly related by recent transmission. Therefore, in order to provide a more specific measure of recent transmission, detailed patient interviews, medical records, and notes and documentation from contact tracing investigations were reviewed for each case that was part of a cluster in order to find evidence of epidemiologic linkages between clustered cases. Patient interviews were conducted using

an extensive and standardized questionnaire with the explicit purpose of identifying potential infection sources and connections among patients in a cluster. All contact tracing investigations and patient interviews were conducted by an experienced group of Arkansas Department of Health TB control staff following a standardized protocol, limiting concern that regional differences in cluster investigation could bias results.

A definite epidemiologic link was defined as one between patients within a cluster who lived in the same household or shared the same indoor airspace when at least one of the patients was judged to be infectious [21]. A probable epidemiologic link was defined as a link that did not fit the criteria for a definite epidemiologic link, but for which the interview data suggested either a direct exposure of one infectious patient to another in the cluster, or a circumstance whereby cluster patients were in the same location at the same time (drawing on questions regarding workplaces, places of education, places of worship, and locations where the patient spent a lot of time on a regular basis) [18]. If neither a definite nor probable link could be identified between a patient and any other patient in the same cluster, that patient was classified as having no known epidemiologic link. Epidemiologic linkage could not be classified for 3 cases infected with *lipR* LSP strains and 2 cases infected with *lipR* wild-type strains, as medical records and interview notes were not available.

Data analysis

We used contingency tables to assess whether or not infection with *lipR* LSP isolates versus wild type is associated with a given demographic or clinical/epidemiological category of the patients, determining statistical significance, using the χ^2 test or Fisher's exact test, as appropriate. In all analyses involving case clustering, we restricted our sample to those cases infected with isolates carrying >5 copies of *IS6110* for two reasons. First, due to the limited discriminatory power of the *IS6110* RFLP analysis for isolates with < 5 copies of *IS6110*, clustering between isolates with > 5 bands in the *IS6110* RFLP pattern is more likely to represent recent transmission than is clustering between isolates with 5 or fewer *IS6110* RFLP bands [17]. Second, all the isolates with *lipR* LSP had > 5 *IS6110*-hybridizing bands compared to 68.4% of wild-type isolates; even with pTBN12 as a secondary typing method, the level of discrimination between isolates with <5 bands and the others is likely to differ. We took this as a conservative approach to our analysis, as even a slightly lower resolution among isolates with <5 bands could, because of the association between *lipR* LSP and isolates with > 5 bands, create an artificial association.

Among these isolates, we generated crude odds ratios (ORs) and 95% confidence intervals (CIs) to test the association of infection by *lipR* LSP versus wild-type isolates with both molecular clustering and the identification of an epidemiologic link between clustered cases. In order to control for potential confounding by cluster size, we repeated this comparison on a sample restricted to one isolate per cluster (i.e. strain-based analysis). This was possible because, with one exception, *lipR* LSP were invariant within clusters. The variant cluster, consisting of 2 case isolates, was not included in this analysis.

We further analyzed the association between *lipR* LSP and demographic, clinical, and epidemiologic characteristics using a multivariate logistic regression model to adjust for potential confounders. Models were fitted to a dataset restricted to only clustered isolates with >5 *IS6110* RFLP bands, as epidemiologic linkage was not investigated for unique (non-clustered) isolates.

Clusters of TB cases which include homeless individuals present a potential source of bias in our analysis of the association between the *lipR* LSP and the identification of an epidemiologic link between clustered cases. Homeless shelters are well known as high-risk environments for tuberculosis transmission, and clusters associated with homeless shelters may therefore be

more rigorously investigated. Conversely, the investigation of clusters involving homeless individuals may sometimes be impeded by difficulties in locating and interviewing homeless case-patients. In order to address this potential source of bias, we conducted a stratified analysis in which clusters involving one or more cases reported as homeless in the year prior to diagnosis were analyzed separately from those clusters not involving any case identified as homeless. To address concern that our results were driven by any single *lipR* LSP cluster, we created a macro that re-ran our multivariate regression model, iteratively excluding each *lipR* LSP cluster from the sample.

Results

lipR LSP

PCR of the *lipR* gene region generated a product of ~0.7 kb, smaller than the expected wild-type product size (~3.6 kb), for 104 (15.6%) of 665 isolates, while the remaining 561 isolates had wild-type product size. In all 104 isolates with reduced PCR product size, DNA sequencing data analysis confirmed that the deletion extended from nucleotide 3448508 to nucleotide 3451398 of the H37Rv genome (~2.9 kb).

Association between *lipR* genotype and molecular typing

Of the 104 isolates with a LSP in the *lipR* region, 100 (96.2%) belonged to principal genetic group 3, one (1.0%) to genetic group 2, and three (2.9%) were of undetermined genetic group. This distribution differed significantly ($p < 0.0001$) from that among the 557 of 561 wild-type isolates that had a genetic group determined, of which 73 (13.1%) belonged to genetic group 3, 424 (76.1%) to genetic group 2, and 60 (10.8%) to genetic group 1. The ranges of the IS6110-hybridizing band numbers were 6-16 and 1-22 for isolates with and without *lipR* LSP, respectively, with 62 distinct IS6110 RFLP patterns among *lipR* LSP isolates and 326 among *lipR* wild-type isolates. The percentage of IS6110 high-copy isolates (>5 bands) was much higher among the isolates with the *lipR* LSP (100%) than among the *lipR* wild-type isolates (68.4%) (Figure 1).

Association between *lipR* genotype and patient demographic and clinical data

Patients infected with *lipR* LSP isolates were significantly more likely than patients infected with *lipR* wild-type isolates to report a history of homelessness and alcohol abuse in the year prior to diagnosis (Table 1). No statistically significant associations were found between infection with a *lipR* LSP isolate and clinical variables including disease site (thoracic/extrathoracic), lung cavitation, age, race/ethnicity and sputum smear positivity. In addition, we also compared the distribution of patient demographic and clinical data between the *lipR* LSP isolates and the *lipR* wild-type isolates within principal genetic group 3. This restricted analysis showed that the *lipR* LSP isolates were more likely to be from younger patients (e.g. < 65 year old) ($p < 0.0001$), and males ($p = 0.04$).

Association between *lipR* genotype and clustering

With one exception, all clusters (as defined by the conventional clustering definition) were exclusively composed of either wild-type isolates or isolates having the *lipR* LSP. The exception was a cluster of two isolates that shared an identical 11-band IS6110 RFLP pattern and included a *lipR* wild-type genetic group 2 isolate and a *lipR* LSP genetic group 3 isolate. No epidemiologic link could be identified between the two patients in this *lipR* discordant cluster.

Among the remaining 91 concordant clusters, there were 83 wild-type clusters and 8 clusters of isolates having the *lipR* LSP, with an average cluster size of 3.4 and 6.5 isolates, respectively.

The largest clusters, wild-type or having the *lipR* LSP respectively, were 17 and 16 isolates in size. The *lipR* LSP was not significantly associated with clustering of isolates using the conventional clustering definition, however, *lipR* LSP isolates were 1.8 times more likely than wild-type isolates to be involved in a cluster as defined by the 1-year time-restricted definition (Table 2).

Among non homeless-associated clusters, the mean cluster size among 75 wild-type clusters was 3.1, compared to a mean of 4.0 among 6 *lipR* LSP clusters. Among homeless-associated clusters, the mean cluster size among 8 homeless-associated clusters was 5.8, compared to 14 between 2 *lipR* LSP clusters.

Association between *lipR* genotype and epidemiologic data

A statistically significant association was observed between *lipR* genotype and type of epidemiologic linkage (Table 3). Epidemiologic linkages were identified 3.07 [95%CI: (1.6-5.9)] times more often for cases infected with a *lipR* LSP isolate. This association remained significant in a multivariate logistic model including all clustered, >5 IS6110 RFLP band isolates that adjusted for potential confounding by race/ethnicity, age, homelessness, and alcohol abuse. A model considering only clusters without an association with homelessness predicted the identification of an epidemiologic link 5 [95%CI: (1.7-14.5)] times more often among clustered cases infected with a *lipR* LSP (Table 4). However, in a model in which only homeless-associated clusters was considered, no variable was significantly associated with infection with a *lipR* LSP isolate (data not shown). The associations between homelessness, epidemiologic linkage, and *lipR* genotype were not driven by any single cluster of *lipR* LSP isolates: effect estimates and statistical significance of the associations between the *lipR* LSP and both epidemiologic linkage type and homelessness were consistently robust as each cluster of *lipR* LSP isolates was iteratively excluded from the model.

Discussion

We investigated the *lipR* LSP in a large, population-based sample of clinical isolates of *M. tuberculosis*, analyzing the association between this LSP and clinical, demographic, and epidemiological characteristics of TB patients. The major findings from the study include: 1) approximately 16% of the isolates for which a PCR product was generated had a *lipR* deletion; 2) the majority (100/104) of isolates with a LSP in the *lipR* region belonged to principal genetic group 3, and 3) epidemiologic linkages were identified more often for clustered cases infected with a *lipR* LSP isolate than for clustered cases infected with a *lipR* wild-type isolate, even after adjusting for potential biases of large clusters and confounding by host risk factors for TB transmission, including being homeless, excessive alcohol use, and non-Hispanic black race.

In a previous investigation of genetic variation among *M. tuberculosis* clinical isolates collected in San Francisco, deletions of the *lipR* region were observed in 8 of 100 (8%) isolates which had been selected from that population to represent 50 clustered and 50 unique strains, such that no transmission-linked isolates would be included [10]. The higher frequency of *lipR* deletions observed in the population-based sample described here (16%) may reflect a true difference in the frequency of *lipR* deletions between Arkansas and San Francisco, or may reflect differences in sampling approaches. Additionally, the near-perfect correlation we observed between the *lipR* deletion and principal genetic group 3 suggests that the frequency of *lipR* deletions may depend on the principal genetic group composition of a given study sample. Genetic group 3 made up 26.0% of the isolates in our total study sample; unfortunately, genetic group was not reported in the San Francisco study. The high correlation with genetic group 3, combined with the observed difference in the IS6110 band number distributions for wild-type isolates as compared to isolates with the *lipR* deletion, suggests that the LSP in

lipR may have resulted from a unique deletion event followed by proliferation and divergence. This is consistent with the phylogenetic analysis in the San Francisco study [10] that grouped all 8 of their study isolates with RD219 disrupting the *lipR* region, based on the distribution of 68 large genomic deletions [10,22].

The most interesting finding in this study is that clustered isolates with *lipR* deleted were associated with an increased likelihood of the patient having an identified epidemiologic link, as compared with clustered isolates with an intact *lipR*. Numerous reports have described clustered cases for which no epidemiologic link can be found [18,23,24]. These cases may reflect cases that are not related by recent transmission [18,24]. They may also reflect cases that are related by recent transmission, but for which the TB control staff was unable to identify a linkage. Because all contact-tracing investigations included in this study were conducted by a dedicated group of centrally based Arkansas Department of Health TB control staff following a standardized and detailed protocol, it is unlikely that the rigor of contact-tracing investigations (and therefore the probability of identifying an epidemiologic linkage) varied across local health jurisdictions. However, because the well established risk of TB transmission in homeless shelters may result in more rigorous contact investigations for clusters associated with these settings, concern arose that epidemiologic links might have been identified more often among cases involved in homeless-associated clusters due to heightened investigator awareness. The robustness of the association between the infection with a *lipR* LSP and the identification of an epidemiologic link observed in a multivariate model restricted to non-homeless associated cases indicates that such a bias does not explain this association. Furthermore, because the *lipR* LSP isolates belonged nearly exclusively to genetic group 3, one may ask is the observed association with the *lipR* LSP or with the principal genetic group 3. However, no associations between genetic group 3 and any of the clinical, demographic, or epidemiologic variables have been reported by others, nor were evident in our sample.

These results provide strong evidence of an association between infection with a *lipR* LSP isolate and recent transmission among incident cases of culture-positive TB reported in Arkansas between 1996 and 2000. A number of plausible hypotheses may explain this finding. First, it may reflect the dynamics of TB transmission in Arkansas, with the *lipR* LSP acting as a genetic marker for a lineage of strains that were recently introduced to the population, and are therefore most prevalent in the high-risk populations among which the bulk of recent TB transmission occurs, or this association could reflect a functional relationship between the *lipR* gene and the ability of *M. tuberculosis* to be transmitted by the airborne route and its ability to survive and multiply intracellularly in the human host [2,8,25]. The range of IS6110 band numbers (6 to 16) among the isolates with *lipR* LSP indicates the significant genotype variation among the isolates, suggesting that it is unlikely that the bulk of the *lipR* LSP isolates were part of a single outbreak. The impact of variation in *lipR* or *mymA* on the survival of the organism while suspended in air as a droplet nucleus has not been investigated. However, decreased survival of *lipR* deletion strains in the environment, and a consequent impairment of transmission, is plausible, given the importance of the lipid content of the *M. tuberculosis* cell envelope in the survival of the organism in the air [3]. Alternatively, enhanced intracellular survival may increase the probability of progression from infection to disease and result in disease with a heavier bacterial load, prolonging and increasing infectiousness and increasing the chance for transmission. The other possibility could be that strains lacking *lipR* are less able to establish latent infection and thus cases in an infection network are more closely linked in time and more easily identified by contact tracing. Alternatively, it is also possible that the *lipR* LSP is a marker for a strain lineage with other genetically linked mutations that contribute to the observed association.

Epidemiologic linkages are more likely to be identified between close friends or relatives than between more distant or casual acquaintances. If the *lipR* LSP impairs transmissibility of the

organism, requiring relatively more extensive contact between an infectious individual and a susceptible contact to achieve transmission, while wild-type strains may be transmitted after only limited or casual contact, in such a condition, epidemiologic linkages would be easier to identify for cases infected with *lipR* LSP isolates. This explanation is also supported by the association of *lipR* LSP isolates with homelessness and excess alcohol use, both factors that facilitate transmission. Given the limitations of the available data, we were not able to categorize patients into levels of contact.

While the *lipR* gene has been identified as a probable lipase-encoding gene, the functional effect of the LSP we identified is unknown. The effect of the deletion may not be limited to an interruption of the *lipR* gene because *lipR* is the second gene in the *mymA* operon; it is possible that this deletion also results in the interruption of downstream *mymA* genes. Further work to elucidate the effect of this deletion is needed. The inherent limitations of observational studies make it impossible to determine if the dynamics of strain transmission in our population, or a functional disruption of the *lipR* gene, are responsible for the associations we observed in this study. Both possibilities have implications for TB control, and warrant further exploration through studies of this LSP in other populations as well as in functional studies.

Acknowledgments

The genotyping data and contact information of the study patients used in this study were obtained, respectively, by the Mycobacteriology Laboratory at the Central Arkansas Veterans Healthcare Center and the Tuberculosis Epidemiological Investigation Team at the Arkansas Department of Health as a part of the National Tuberculosis Genotyping and Sentinel Surveillance Program sponsored by the interagency agreement 98FED10318 between the Veterans Administration and the Centers for Disease Control and Prevention. The investigation of the *lipR* LSP and all the epidemiological data analyses were supported by the National Institutes of Health grant NIH-R01-AI151975. The authors are indebted to Leonard N. Mukasa, Iram Bakhtawar, and the TB control staff at Arkansas Department of Health for their help in retrieving the contact investigation data from the Arkansas Department of Health's epidemiological information archive.

References

1. Kolattukudy PE, Fernandes ND, Azad AK, Fitzmaurice AM, Sirakova TD. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol* 1997;24:263–70. [PubMed: 9159514]
2. Singh A, Jain S, Gupta S, Das T, Tyagi AK. *mymA* operon of *Mycobacterium tuberculosis*: its regulation and importance in the cell envelope. *FEMS Microbiol Lett* 2003;227:53–63. [PubMed: 14568148]
3. Heinsohn, P. Respiratory tract infections -- bacteria. In: Flannigan, B.; Samson, RA.; Miller, JD., editors. *Microorganisms in home and indoor work environments : diversity, health impacts, investigation and control*. London ; New York, NY: Taylor & Francis; 2001. p. 155-194.
4. Kremer L, de Chastellier C, Dobson G, et al. Identification and structural characterization of an unusual mycobacterial monomeromycetyl-diacylglycerol. *Mol Microbiol* 2005;57:1113–26. [PubMed: 16091048]
5. Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;393:537–44. [PubMed: 9634230]
6. Cotes K, Dhoub R, Douchet I, et al. Characterization of an exported monoglyceride lipase from *Mycobacterium tuberculosis* possibly involved in the metabolism of host cell membrane lipids. *Biochem J* 2007;408:417–27. [PubMed: 17784850]
7. Singh A, Gupta R, Vishwakarma RA, et al. Requirement of the *mymA* operon for appropriate cell wall ultrastructure and persistence of *Mycobacterium tuberculosis* in the spleens of guinea pigs. *J Bacteriol* 2005;187:4173–86. [PubMed: 15937179]
8. Fisher MA, Plikaytis BB, Shinnick TM. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol* 2002;184:4025–32. [PubMed: 12081975]

9. Kato-Maeda M, Rhee JT, Gingeras TR, et al. Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* 2001;11:547–54. [PubMed: 11282970]
10. Tsolaki AG, Hirsh AE, DeRiemer K, et al. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci U S A* 2004;101:4865–70. [PubMed: 15024109]
11. Sreevatsan S, Pan X, Stockbauer KE, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 1997;94:9869–74. [PubMed: 9275218]
12. Kong Y, Cave MD, Zhang L, et al. Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of *M. tuberculosis* clinical isolates. *J Clin Microbiol* 2007;45:409–14. [PubMed: 17166963]
13. France AM, Cave MD, Bates JH, Foxman B, Chu T, Yang Z. What's Driving the Decline in Tuberculosis in Arkansas? A Molecular Epidemiologic Analysis of Tuberculosis Trends in a Rural, Low-Incidence Population, 1997–2003. *Am J Epidemiol* 2007;166:662–71. [PubMed: 17625223]
14. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis* 1999;180:726–36. [PubMed: 10438361]
15. Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* 1994;330:1703–9. [PubMed: 7910661]
16. Ellis BA, Crawford JT, Braden CR, McNabb SJ, Moore M, Kammerer S. Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis* 2002;8:1197–209. [PubMed: 12453343]
17. Yang Z, Chaves F, Barnes PF. Evaluation of method for secondary DNA typing of *Mycobacterium tuberculosis* with pTBN12 in epidemiologic study of tuberculosis. *J Clin Microbiol* 1996;34:3044–8. [PubMed: 8940446]
18. Braden CR, Templeton GL, Cave MD, et al. Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. *J Infect Dis* 1997;175:1446–52. [PubMed: 9180185]
19. Nguyen D, Brassard P, Menzies D, et al. Genomic characterization of an endemic *Mycobacterium tuberculosis* strain: evolutionary and epidemiologic implications. *J Clin Microbiol* 2004;42:2573–80. [PubMed: 15184436]
20. Jasmer RM, Hahn JA, Small PM, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. *Ann Intern Med* 1999;130:971–8. [PubMed: 10383367]
21. Cave MD, Yang ZH, Stefanova R, et al. Epidemiologic import of tuberculosis cases whose isolates have similar but not identical IS6110 restriction fragment length polymorphism patterns. *J Clin Microbiol* 2005;43:1228–33. [PubMed: 15750088]
22. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A* 2004;101:4871–6. [PubMed: 15041743]
23. Lambregts-van Weezenbeek CS, Sebek MM, van Gerven PJ, et al. Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung Dis* 2003;7:S463–70. [PubMed: 14677839]
24. Oelemann MC, Diel R, Vatin V, et al. Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J Clin Microbiol* 2007;45:691–7. [PubMed: 17192416]
25. Cheruvu M, Plikaytis BB, Shinnick TM. The acid-induced operon Rv3083–Rv3089 is required for growth of *Mycobacterium tuberculosis* in macrophages. *Tuberculosis (Edinb)* 2007;87:12–20. [PubMed: 16893682]

Figure 1.

Comparison of IS6110-hybridizing band number distribution between *lipR* wild-type isolates and *lipR* LSP isolates based on a population-based sample obtained from Arkansas during the time period from 1996 to 2000 (n=665). Shaded bars indicate wild-type; white bars represent LSP.

Table 1

Distribution of key demographic and behavioral characteristics among patients infected by a *lipR* LSP or wild-type *lipR* isolate, respectively, based on analyzing 665 incident cases of TB diagnosed in Arkansas from 1996 to 2000.

Variable ^a	<i>lipR</i> LSP n = 104 No. (%)	<i>lipR</i> wild-type n = 561 No. (%)	p-value
Age (years)			
0-19	0	10 (1.8)	0.156
20-44	29	121 (21.6)	
45-64	30	136 (24.2)	
65-84	34	198 (35.3)	
85+	11	96 (17.1)	
Sex			
Male	70	352 (62.8)	0.375
Female	34	209 (37.2)	
Race/ Ethnicity			
Non-Hispanic white	52	321 (57.2)	0.132
Non-Hispanic black	43	186 (33.2)	
Hispanic	7	21 (3.7)	
Asian / Pacific Islander	2	30 (5.4)	
American Indian / Alaska Native	0	3 (0.5)	
Excess alcohol use within past year			
No	71	461 (87.3)	0.013
Yes	22	67 (12.7)	
Homeless within past year			
No	93	531 (98.0)	0.001
Yes	8	11 (2.0)	
US Born			
No	7	41 (7.3)	0.834
Yes	97	520 (92.7)	
HIV status			
Known positive	5	20 (3.6)	0.097
Known negative	57	250 (44.6)	
Not tested/Unknown	42	291 (51.9)	

Variable ^a	<i>lipR</i> LSP n = 104 No. (%)	<i>lipR</i> wild-type n = 561 No. (%)	p-value
-----------------------	---------------------------------------	---	---------

^a Other demographic and clinical variables for which the association with *lipR* LSP was analyzed and found to be non-significant were residence at a correctional facility, residence in a Metropolitan Statistical Area, disease site (thoracic/extra-thoracic), lung cavitation, previous TB disease, and sputum smear positivity.

Table 2

Relative odds that infection with a *lipR* LSP isolate, as compared to a *lipR* wild-type isolate, was clustered by IS6110 RFLP and pTBN12 typing, determined by analyzing 489 isolates with >5 IS6110 copies, using the conventional cluster definition and a 1-year time restricted cluster definition, respectively.

	Clustered Nb. (%)	Unique Nb. (%)	OR (95% CI)
Conventional cluster definition			
Wild-type <i>lipR</i>	168	n=270	
<i>lipR</i> LSP	51	217	1.0
		(56.4)	
		53	1.24 (0.81, 1.92)
		(51.0)	
1-year time restricted cluster definition			
Wild-type <i>lipR</i>	120	n=321	
<i>lipR</i> LSP	48	265	1.0
		(68.8)	
		56	1.89 (1.22, 2.94)
		(46.2)	

Table 3
Relative odds that an epidemiologic linkage was identified between clustered cases infected with a *lipR* LSP isolate, as compared with those infected with a *lipR* wild-type isolate, determined by analyzing 214 clustered isolates with >5 IS6110 copies.

	Linkage identified No. (%)		No linkage identified No. (%)		OR (95% CI)
All clustered isolates					
Wild-type <i>lipR</i>	49	n=76 (12.7)	117	n=138 (30.6)	1.0
<i>lipR</i> LSP	27	(26.7)	21	(20.8)	3.07 (1.59, 5.94)
Non-homeless associated clustered isolates					
Wild-type <i>lipR</i>	33	n=46 (24.4)	102	n=113 (75.6)	1.0
<i>lipR</i> LSP	13	(54.2)	11	(45.8)	3.65 (1.49, 8.93)
Homeless- associated clustered isolates					
Wild-type <i>lipR</i>	16	n=30 (51.6)	15	n=25 (58.3)	1.0
<i>lipR</i> LSP	14	(58.3)	10	(41.7)	1.31 (0.45, 3.84)

Table 4

Relative odds that an epidemiologic linkage was identified between clustered cases infected with a *lipR* LSP isolate, as compared to a *lipR* wild-type isolate, adjusting for potential confounders in a multivariate logistic regression model restricted to isolates with > 5 IS6110 copies.

Variable	Adjusted OR (95% CI)	
	Using all clustered cases (n=214)	Using clustered cases infected with a non-homeless associated strain ^a (n=159).
No epidemiologic linkage identified	1.0 ^b	1.0 ^b
Epidemiologic linkage identified	2.74 (1.32, 5.68)	5.02 (1.73, 14.46)
Race/ Ethnicity		
Non-Hispanic white	1.0 ^b	1.0 ^b
Non-Hispanic black	1.47 (0.99, 2.18)	1.71 (1.01, 2.90)
Other	0.48 (0.05, 4.23)	0.67 (0.07, 6.77)
Excess alcohol use within past year		
No	1.0 ^b	1.0 ^b
Yes	1.30 (0.56, 2.99)	0.90 (0.25, 3.20)
Homeless within past year		
No	1.0 ^b	1.0 ^b
Yes	2.81 (0.77, 10.24)	-- ^d
Age (years)		
Age 0 to 19	-- ^c	-- ^c
Age 20 to 44	1.0 ^b	1.0 ^b
Age 45 to 64	1.39 (0.61, 3.14)	1.12 (0.35, 3.60)
Age 65 up	1.77 (0.60, 5.21)	2.69 (0.64, 11.29)

^aIn order to control for the possibility that clusters associated with homeless shelters may be more rigorously investigated than clusters not associated with homeless shelters, we fit our multivariate model to a sample set restricted to non-homeless associated strains. A non-homeless associated strain was defined as a characteristic, >5 band IS6110 RFLP pattern, which was not isolated from any case-patient reported to be homeless in the year prior to diagnosis.

^bReference group.

^cNo cases in this group were infected with a *lipR* LSP isolate, so an OR could not be calculated.