Clinical Relevance of *Mycobacterium tuberculosis plcD* Gene Mutations

Zhenhua Yang, Dong Yang, Ying Kong, Lixin Zhang, Carl F. Marrs, Betsy Foxman, Joseph H. Bates, Frank Wilson, and M. Donald Cave

Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan; Arkansas Department of Health; Department of Epidemiology, College of Public Health, and Department of Neurobiology and Developmental Sciences, College of Medicine, University of Arkansas for Medical Sciences; and Central Arkansas Veterans Healthcare Center, Little Rock, Arkansas

To identify *Mycobacterium tuberculosis* virulence factors, we integrated comparative genomics and epidemiologic data analysis to investigate the relationship between certain genomic insertions and deletions in the phospholipase-C gene D (*plcD*) with the clinical presentation of tuberculosis (TB). Four hundred ninety-six well-characterized *M. tuberculosis* clinical isolates were studied. Approximately 30% (147) of the isolates had an interruption of the *plcD* gene. Patients infected with the *plcD* mutant were twice as likely to have extrathoracic disease as those infected by a strain without an interruption (adjusted odds ratio, 2.19; 95% confidence interval, 1.27, 3.76). When we limited the analysis to the 275 isolates with distinct DNA fingerprint patterns, we observed the same association (adjusted odds ratio, 2.74; 95% confidence interval, 1.35, 5.56). Furthermore, the magnitude of the association appeared to differ with the type of extrathoracic TB. Our findings suggest that the *plcD* gene of *M. tuberculosis* is potentially involved in the pathogenesis of TB, and the clinical presentation of the disease may be influenced by the genetic variability of the *plcD* region.

Keywords: extrathoracic tuberculosis; pathogenesis; virulence factors

Tuberculosis (TB) is a major cause of morbidity and mortality worldwide (1). The lack of an effective vaccine to prevent those who are latently infected with *Mycobacterium tuberculosis* from developing active disease, the variable results obtained in bacillus Calmette-Guérin vaccine efficacy trials, the emergence of multidrug-resistant TB worldwide, and the increasing coinfection of HIV with *M. tuberculosis* in many regions of the world (2–4) highlight the need for a highly effective vaccine and new antimicrobial agents. Identification of the *M. tuberculosis* factors that contribute to the pathogenesis of TB is necessary if the goal to develop a more effective vaccine is to be reached.

To date, identification of *M. tuberculosis* virulence factors has been based primarily on *in vitro* and animal studies (5). Although these studies have increased our understanding of the function of a number of *M. tuberculosis* genes and the impact of gene products on the organisms' behavior in various *in vitro* and *in vivo* models, correlation of *in vitro* findings and animal studies with the pathogenesis of human disease remains challenging (5). An alternative strategy to identify *M. tuberculosis* virulence factors (6–11) combines comparative genomics to identify genome alteration with epidemiologic methods, to assess associations between genetic polymorphisms and clinical characteristics of the disease.

Previous comparative genomic studies have identified large sequence deletions in multiple regions of the *M. tuberculosis* genome (6, 11, 12). However, the epidemiologic and clinical phenotypes of these genomic alterations remain largely unknown. Of the previously reported genomic deletions, the region containing the phospholipase-C gene D (*plcD*) (6) is of particular interest because of the role of *plc* in the pathogenesis of disease caused by a number of intracellular bacteria (13–15). Furthermore, the expression of *M. tuberculosis plc* genes is strongly upregulated during the first 24 hours of macrophage infection, and *plc* gene mutants are attenuated in the late phase of the infection in a mouse model (16).

This article reports an observational study that integrates comparative genomics with epidemiologic methods to investigate the genetic diversity in the *plcD* gene region among a large set of *M. tuberculosis* isolates and to assess the association between the observed genetic polymorphisms and the clinical presentation of the patients. Some of the results have been previously reported in the form of a meeting abstract (17).

METHODS

Study Design and Sample

This study was performed in two steps. First, we performed a molecular characterization of the *plcD* gene region of a sample of *M. tuberculosis* clinical isolates. Then, we assessed the association of the genetic alterations of the *plcD* gene with certain patient characteristics, particularly the clinical presentation of the patients. The study sample included 496 isolates of *M. tuberculosis* obtained from 496 patients with culture-confirmed TB diagnosed in Arkansas between January 1, 1996, and December 31, 2000. During the study period, a total of 973 TB cases were diagnosed in Arkansas. Of these 973 cases, 847 (87.05%) were defined as thoracic TB and 126 (12.95%) as extrathoracic TB using the definitions described later. The proportions of thoracic and extrathoracic TB being culture confirmed were 74.62 (632/847) and 69.05% (87/126), respectively, resulting in a total of 719 culture-confirmed cases. Of these 719 cases, 705 (98%) had a viable isolate at the Mycobacteria Research Laboratory at the Central Arkansas Veterans Healthcare Center. At the laboratory of the University of Michigan, the genomic DNA of 496 retrievable isolates was successfully extracted. All of these isolates were included in this study. Isolates were obtained from extrathoracic sites in 63 of the 68 (92.65%) extrathoracic TB cases studied.

Genomic DNA from the isolates was extracted from Lowenstein-Jensen cultures using standard procedures (18). All the isolates were genotyped using a combination of IS6110 fingerprinting and pTBN12 secondary typing (19). Patient data were obtained from the Arkansas Department of Health surveillance records, as described previously (20). This study was approved by the Health Sciences Institutional Review Boards of the University of Michigan and the University of Arkansas for Medical Sciences.

Polymerase Chain Reaction Assays

All 496 isolates were screened for DNA polymorphisms of the *plcD* gene region using a polymerase chain reaction (PCR) assay designated *plcD*-PCR1. The *plcD*-PCR1 amplifies the region extending from 107 bp left-hand side to 240 bp right-hand side of the *plcD* gene (1977761bp–

(Received in original form August 31, 2004; accepted in final form March 23, 2005)

Supported by the National Institutes of Health (grant NIH-R01-AI151975).

Correspondence and requests for reprints should be addressed to Zhenhua Yang, M.D., Ph.D., Epidemiology Department, School of Public Health, University of Michigan, 109 S. Observatory Street, Ann Arbor, MI 48109-2029. E-mail: zhenhua@umich.edu

Am J Respir Crit Care Med Vol 171. pp 1436–1442, 2005

Originally Published in Press as DOI: 10.1164/rccm.200408-1147OC on April 1, 2005

Internet address: www.atsjournals.org

TABLE 1. ANATOMIC SITES OF 68 CULTURE-CONFIRMED EXTRATHORACIC CASES

Site of Extrapulmonary TB	Extrathoracic Involvement Only		Extrathoracic and Thoracic Involvement		Total
	No.	%	No.	%	
Cervical lymph node	7	10.29	2	2.94	9
Other lymph node	3	4.41	3	4.41	6
Bone and joint	15	22.06	6	8.82	21
Genitourinary	6	8.82	1	1.47	7
Miliary	3	4.41	5*	7.35	8
Meningeal	1	1.47	3	4.41	4
Peritoneal	1	1.47	3	4.41	4
Pancreas	1	1.47	2	2.94	3
Small intestine	1	1.47	1	1.47	2
Laryngeal		0	1†	1.47	1
Breast			1	1.47	1
Chest wall			1	1.47	1
Pericardial			1†	1.47	1
Total	38	55.88	30	44.12	68

Definition of abbreviation: TB = tuberculosis.

* Of the five miliary cases, two isolates were obtained from sputum and one from lung.

† The isolate was from sputum.

1979655bp). When an isolate failed to produce a detectable product, a control PCR targeting the single-copy 16S rRNA gene was performed to confirm the quantity and quality of the DNA templates (21). These isolates were tested further using a secondary PCR, designated *plcD*-PCR2, which amplifies an approximately 20-kb region extending from 116 bp left-hand side to 18.175 kb right-hand side of the *plcD* gene (1977753bp–1997587bp); this generated a PCR fragment for DNA sequencing to describe actual sequence alteration of the *plcD* gene (12). For isolates that remained negative in *plcD*-PCR2, a third PCR, designated *plcD*-PCR3, was performed to amplify the region extending from 642 bp left-hand side to 432 bp right-hand side of the *plcD* gene (1977226bp–1979844bp).

Genomic DNA of the *M. tuberculosis* laboratory strain H37Rv and the clinical strain CDC1551 served as a negative and a positive control, respectively, in PCR experiments as the *plcD* gene region is truncated and interrupted in strain H37Rv (6) and is intact in strain CDC1551 (<http://www.tigr.org>) (22).

Southern Blot Analysis

For isolates that generated a *plcD*-PCR1 product of altered size as compared with that of strain CDC1551, Southern hybridization of the PCR product using the IS6110-3' probe was performed to confirm the IS6110 insertion (23). For isolates that failed to produce positive results in any of the PCRs, Southern blotting of the *Pvu*II-restricted genomic DNA was done using the *plcD* gene as a probe to confirm the interruption of the *plcD* gene sequence. The *plcD* probe was the purified *plcD*-PCR1 product of CDC1551.

DNA Sequencing

The sites of the IS6110 insertion and the insertion-associated deletions were determined by automated DNA sequencing. Sequence comparisons were performed using the software Edit Seq 5.02 and MegAlign 5.01 (DNASTar, Inc., Madison, WI).

Definitions

Isolates with an interruption or deletion of the *plcD* gene region were designated as mutant-type, and those with an intact *plcD* gene as wild-type. To assess the association of the *M. tuberculosis plcD* gene mutation with the capacity for the infecting organism to move distinctly beyond the original anatomic site of disease to another anatomic site outside the lung, pleural surface, and lymph nodes within and immediately adjacent to the lung, we designated those patients whose disease sites were confined to the lung, pleura, and intrathoracic lymph nodes as thoracic TB, reasoning that migration of infection and disease to these intrathoracic sites is very common in cases of thoracic TB, although it is frequently not noted clinically. We defined patients who had extrathoracic

racic disease with or without concurrent disease within the thoracic cavity as being in the extrathoracic group.

Statistical Analysis

Using the χ^2 , or Fisher's exact test, as appropriate, we compared the characteristics of the study patients with all the culture-confirmed patients who had viable isolates ($n = 705$) to address potential selection bias. Then, we tested the differences in demographic and clinical characteristics between the wild and mutant types. The magnitude of the association between the *plcD* gene region genotype and the disease presentation was estimated using the odds ratio and 95% confidence intervals. To control for potential confounding of previously reported host-related risk factors for extrathoracic TB (20), multivariate logistic regression was performed. Because 44% of the study isolates were clustered by DNA fingerprinting, we also performed both bivariate and multivariate logistic regression analyses on 275 isolates with unique IS6110 fingerprint patterns to confirm that the association observed was independent of the clustering of the isolates. To gain a better understanding of the specificity of the potential effect of the *plcD* gene mutation on the occurrence of extrathoracic TB, we performed an association analysis for three different types of extrathoracic TB (i.e., extrathoracic

Figure 1. Agarose gel electrophoresis (A) and IS6110 Southern hybridization (B) of *plcD*-PCR1 products. Lane 1: 1-kb molecular size standards (Bio-Rad, Hercules, CA); Lanes 2 through 6: representatives of CDC1551, groups I, III, IV, and II isolates.

lymphadenitis, isolated extrathoracic organ involvement, and disseminated disease, including those with miliary disease as well as those having at least two distinct anatomic sites of extrathoracic disease). All the statistical analyses were done using SAS version 8.0 (SAS Institute, Cary, NC) (24).

RESULTS

Characteristics of Study Patients

Of the 496 patients studied, 428 were defined as having thoracic TB and 68 as having extrathoracic disease, representing, respectively, 78.16 (68/87) and 67.72% (428/632) of the patients with culture-

confirmed thoracic and extrathoracic disease in Arkansas during the study period. The anatomic sites of the 68 extrathoracic cases are set out in Table 1. Of the 443 patients who had a chest radiograph report available, 173 (39.05%) were found to have cavitory involvement. Of the 496 study isolates, 221 (44.56%) were contained within 58 fingerprint clusters based on a combination of IS6110 fingerprinting and pTBN12 secondary typing (19). The size of the clusters ranged from 2 to 16 patients. A comparison of the patients' sociodemographic information, including age, sex, race/ethnicity, geographic location and type of residency, homelessness, alcohol consumption, and drug use, showed

Figure 2. (A, B) Mapping of the IS6110 insertion and sequence deletion of the *plcD* gene and adjacent region. *Section A:* gene map of CDC1551 for the region studied (GenBank, NC-002755); *Section B:* locations of polymerase chain reaction (PCR) primers (sequences of the primers [5'–3']: *plcD*-PCR1F, TCG CCC GGA CAG GTC AAC AAG GTG; *plcD*-PCR1R, CGC GCG CCG CCG CGC CGA AATA; *plcD*-PCR2F, GTC CAC TGT CGC CCG GAC AGG TCA ACA AGG TGT; *plcD*-PCR2R, CGC TGG TAC ACC TGG GGA ATC TGG TGA CGT AGA; *plcD*-PCR3F, CAA TGC GGA TAT CAG TGG AC; and *plcD*-PCR3R, ACA CGA GAC ATT TGG GCT AG); *Section C:* results of DNA sequencing of PCR fragments. Triangles indicate the position of each IS6110 insertion (open triangles: forward insertion; closed triangles: reverse insertion). Horizontal open bracket represents the sequence deletion in each isolate. Numbers in parentheses are numbers of isolates having a given deletion and insertion pattern. (A, *Section C*) Insertion and deletion mapping of groups II, III, IV, and Va. (B, *Section C*) Insertion and deletion mapping of Group Vb isolates.

no significant difference between the study sample and all the 705 patients with culture-confirmed disease who had viable isolates ($p > 0.05$). The two groups also had comparable clinical characteristics, including HIV status, clinical forms of disease (extrathoracic vs. thoracic), sputum smear positivity, and chest radiography findings (cavitary vs. noncavitary; $p > 0.05$).

Diversity of *plcD* Gene

On the basis of the results of the *plcD*-PCR1, the 496 isolates fell into five groups, designated groups I through V. Of the five groups, group I (the wild-type) was the largest, comprising 349 (70.36%) of the 496 isolates. The *plcD*-PCR1 product size of group I isolates was 1.9 kb, identical to that of the CDC1551 product (Figure 1A). The remaining 147 isolates (29.64%) were designated as mutant-type and categorized into four groups (groups II–V). Isolates in groups II to IV all had a PCR product of different size, ranging from 2.4 to 3.3 kb (Figure 1A). Groups II, III, and IV contained 21 (4.23%), five (1.01%), and one (0.20%) isolate, respectively (Figure 2A). Group V included 120 (24.19%) isolates that, like strain H37Rv, failed to generate a product in *plcD*-PCR1, despite the fact that the DNA served as a target for the 16S rRNA gene control PCR. Both mutant and wild types were observed among isolates in 20 (34.48%) of the 58 clusters.

Characterization of Groups II, III, and IV

Isolates in groups II, III, and IV were examined for the insertion of IS6110 into the *plcD* gene sequence and compared with the wild-type group and CDC1551 by Southern hybridization using IS6110 as a probe. IS6110 hybridization was observed in all the isolates in these three groups (Figure 1B), whereas group I and CDC1551 showed no IS6110 hybridization. DNA sequence analysis of the *plcD*-PCR1 products of all the isolates in groups II, III, and IV confirmed that a complete copy of IS6110 was inserted in each isolate. The insertion of IS6110 resulted in a partial deletion of the *plcD* gene in the isolates in groups III and IV, but an IS6110 insertion without a deletion was found for the isolates sequenced for group II. The site and orientation of the IS6110 insertions varied from isolate to isolate (Figure 2A).

Characterization of Group V

The region adjacent to the *plcD* gene was studied by *plcD*-PCR2 in all 120 group V isolates. On the basis of these results, group V isolates were divided into three subgroups designated Va, b, and c. Subgroup Va included 46 (38.33%) isolates that had a positive *plcD*-PCR2 result. The size of the PCR fragments generated by these 46 isolates ranged between 3.0 and 5.4 kb, markedly smaller than the predicted size (19.8 kb) found in the wild type. DNA sequencing of these 46 PCR products revealed that all had an IS6110 insertion within the *plcD* gene, each at a slightly different site and in both orientations (Figure 2A). Each insertion resulted in a deletion of a portion of the *plcD* gene on the right-hand side of the insertion site followed by a deletion of several additional adjacent genes. The size of the deletions found in this approximately 20-kb region ranged from 14.4 to 16.8 kb.

Nineteen (15.83%) of the 120 group V isolates were in subgroup Vb. These isolates were not amplified by *plcD*-PCR2 because of the deletion of the sequence complementary to the forward primer (Figure 2B). However, they were successfully amplified by *plcD*-PCR3. DNA sequencing of the *plcD*-PCR3 product of these 19 isolates identified an IS6110 insertion of dual orientations at the right-hand-side end of gene MT1797 followed by the deletion of MT1798 and a partial *plcD* deletion of 171 to 689 bp long, thereby resulting in the loss of the *plcD*-PCR1F and *plcD*-PCR2F priming sites (Figure 2B).

The remaining 55 (45.83%) of the group V isolates were placed in subgroup Vc. These isolates failed to be amplified by any of the primers used in the study. However, Southern blotting of the *Pvu*II-restricted genomic DNA using the *plcD* gene as a probe confirmed that all 55 isolates had an interruption of the *plcD* gene. Of the 55 isolates, 35 (63.64%) appeared to have a complete deletion of the *plcD* gene, whereas the remaining 20

TABLE 2. DISTRIBUTION OF SELECTED DEMOGRAPHIC, BEHAVIORAL, AND CLINICAL CHARACTERISTICS AMONG 496 PATIENTS WITH TUBERCULOSIS DIAGNOSED IN ARKANSAS BETWEEN JANUARY 1, 1996, AND DECEMBER 31, 2000, BY THE *plcD* GENOTYPE OF THE ISOLATES

Characteristics	<i>plcD</i> Mutant-Type (n = 147)		<i>plcD</i> Wild-Type (n = 349)		p Value
	Cases	%	Cases	%	
Sex					0.53
Male	93	63.27	231	66.19	
Female	54	36.73	118	33.81	
Age, yr					0.59
≤ 60	73	49.66	164	46.99	
> 60	74	50.34	185	53.01	
Race/ethnicity					0.80
Non-Hispanic white	81	55.10	195	55.87	
Non-Hispanic black	54	36.73	120	34.38	
Other	12	8.16	34	9.74	
Homelessness					1.00†
No	140	95.24	330	94.56	
Yes	6	4.08	16	4.58	
Unknown	1	0.68	3	0.86	
Geographic location					0.01
Within city limit	120	81.63	241	69.05	
Outside city limit	25	17.01	91	26.07	
Unknown	2	1.36	17	4.87	
Residency in long-term care facility					0.10†
No	139	94.56	310	88.83	
Yes	8	5.44	38	10.89	
Unknown	0	0.00	1	0.29	
Injection drug use					0.72†
No	143	97.28	333	95.42	
Yes	0	0.00	2	0.57	
Unknown	4	2.72	14	4.01	
Non-injection drug use					0.79
No	139	94.56	326	93.41	
Yes	2	1.36	8	2.29	
Unknown	6	4.08	15	4.30	
Excessive alcohol use					0.35
No	122	82.99	287	82.23	
Yes	23	15.65	49	14.04	
Unknown	2	1.36	13	3.72	
HIV serologic test					0.86
Negative	68	46.26	165	47.28	
Positive	5	3.40	15	4.30	
Unknown	74	50.34	169	48.42	
Cavitary TB*					0.10
No	83	64.34	187	56.84	
Yes	45	34.88	128	38.91	
Unknown	1	0.78	14	4.26	
Positive sputum smear*					0.74
No	70	54.26	170	51.67	
Yes	53	41.09	138	41.95	
Unknown	6	4.65	21	6.38	
Extrathoracic involvement					0.005
No	117	79.59	311	89.11	
Yes	30	20.41	38	10.89	

For definition of abbreviation, see Table 1.

* Only 458 of the 496 cases who had pulmonary involvement were included in the analysis of this variable analysis.

† The p value was determined by Fisher's exact test.

were found to have a partial deletion of the *plcD* gene of approximately 830 to 1,250 bp long.

Clinical Relevance of *plcD* Gene Region

A χ^2 analysis of the selected demographic, behavioral, and clinical characteristics of the 496 study patients showed that characteristics of the patients harboring the *plcD* mutant were comparable to those of patients infected with wild-type isolates with some exceptions (Table 2). Patients infected with *plcD* mutants had a higher proportion of extrathoracic involvement (20.41% for *plcD* mutant group vs. 10.89% for the *plcD* wild-type group), and were more likely to live in cities (81.63% for the *plcD* mutant group vs. 69.34% for the *plcD* wild-type group). After stratification to control for potential confounding by geographic location, the association between the infection with a *plcD* mutant isolate and extrathoracic involvement remained statistically significant (adjusted odds ratio, 2.14; 95% confidence interval, 1.26–3.65). After adjusting for previously identified host-related risk factors for extrathoracic TB (20), infection with a mutant-type *M. tuberculosis* remained significantly associated with extrathoracic TB; this observation held true when we included only the 275 isolates with unique DNA fingerprint patterns in the analysis (Table 3). Patients with extrathoracic involvement, with and without concurrent thoracic TB, had a higher proportion of *plcD* mutant isolates (47.37% for extrathoracic only, 40.00% for concurrent extrathoracic and thoracic involvement, and 27.34% for only thoracic involvement; $p = 0.02$). Further analysis by dividing the extrathoracic TB into three types showed a strong association of the *plcD* mutation with isolated organ involvement, a marginal association with disseminated disease, and no association with extrathoracic lymphadenitis (Table 4). No association was found between infection with *plcD* mutant isolates and the occurrence of cavitory TB or positive sputum smear.

DISCUSSION

This study found that approximately 30% of the study isolates had an interruption of the *plcD* gene by either an insertion of IS6110 inside the *plcD* gene or an insertion of IS6110 followed by a partial deletion of the *plcD* gene that most frequently involved a deletion of adjacent genes. The IS6110 insertion and genomic deletion in the *plcD* gene region of *M. tuberculosis* was

observed in two earlier studies using small selected samples involving 24 and 25 isolates, respectively (12, 25), and in a recent study using a hospital-based sample that included 106 isolates (26). However, the relationship between the genotype of the *plcD* gene of *M. tuberculosis* and the clinical phenotype of the infection has not been described previously. Because extensive data regarding the isolates were available for the present study, we were able to assess the association between the *plcD* gene mutation and the clinical presentations of TB, while confirming the previously reported genetic diversity in this region using a much larger set of isolates. After adjustment for host risk factors (e.g., being female, non-Hispanic black, and HIV positive) for extrathoracic TB identified in a previous study using the same patient dataset (20), we found that infection by a strain having an interrupted *plcD* gene is associated with the occurrence of extrathoracic TB. One limitation of our study is a lack of information on HIV infection status for about half of the study subjects; thus, the adjustment for the confounding of HIV status in our analysis might be imperfect. However, having unknown HIV status was not found to be associated with extrathoracic TB by multivariable analyses (Table 3).

The different degrees of association found between the *plcD* mutation and the different types of extrathoracic TB may suggest that the pathogenesis of TB in different anatomic sites could be affected differently by genetic changes in *M. tuberculosis*. Future studies with larger samples of clinical isolates collected from different anatomic sites would generate useful information to enhance our understanding of the mechanisms of host and *M. tuberculosis* interaction in the pathogenesis of different forms of TB.

In the present study, the mutant *plcD* genotypes were associated with the insertion of IS6110. IS6110 can upregulate downstream genes through an outward-directed promoter in its 3' end (27). Sequence analysis of the mutant isolates showed that all the IS6110 insertions, either within the coding sequence of the *plcD* gene, or in the flanking region of *plcD*, resulted in a partial deletion of the coding sequence of the *plcD* gene. Thus, it is unlikely that the orientation of IS6110 made any difference in *plcD* gene function of the mutants. However, it is possible that the orientation of the insertion did make a difference in the expression of the adjacent genes, which might contribute to defining the clinical phenotype of the study patients. In addition, the mutant types in our study were attributable to only insertion

TABLE 3. MULTIVARIATE LOGISTIC REGRESSION ANALYSIS SHOWING THAT THE EFFECT OF THE *plcD* GENE MUTATIONS OF *M. TUBERCULOSIS* ON EXTRATHORACIC TUBERCULOSIS ARE INDEPENDENT OF KNOWN HOST RISK FACTORS

Variables in the Model	Entire Sample (n = 496)			Unique Isolates Alone (n = 275)		
	Adjusted OR	95% CI of OR	p Value	Adjusted OR	95% CI of OR	p Value
<i>plcD</i> genotype			0.005			0.005
Wild-type	1			1		
Mutant-type	2.19	1.27, 3.76		2.74	1.35, 5.56	
Sex			0.002			0.11
Male	1			1		
Female	2.39	1.39, 4.11		1.80	0.88, 3.66	
Race/ethnicity			0.10			0.16
Non-Hispanic white	1			1		
Non-Hispanic black	1.93	1.07, 3.51		2.12	0.98, 4.59	
Other	1.46	0.54, 3.92		1.29	0.38, 4.39	
HIV serologic test			0.003			0.13
Negative	1			1		
Positive	5.80	2.09, 16.08		4.73	1.03, 21.85	
Unknown	1.18	0.66, 2.14		1.16	0.53, 2.55	

Definition of abbreviations: CI = confidence interval; OR = odds ratio.

TABLE 4. ASSOCIATIONS BETWEEN *plcD* GENOTYPE OF *M. TUBERCULOSIS* CLINICAL ISOLATES AND THE OCCURRENCE OF DIFFERENT TYPES OF EXTRATHORACIC TUBERCULOSIS (n = 496)

Disease Type	<i>plcD</i> Mutant-Type		<i>plcD</i> Wild-Type		p Value	Crude OR (95% CI)	Adjusted OR (95% CI)*
	No.	%	No.	%			
Thoracic TB alone	117	79.59	311	89.11	0.02	1	1
Extrathoracic lymphatic TB	3	2.04	7	2.01		1.14 (0.29, 4.48)	1.46 (0.35, 6.13)
Disseminated TB	14	9.52	20	5.73		1.86 (0.91, 3.81)	1.88 (0.91, 3.90)
Isolated extrathoracic TB	13	8.84	11	3.15		3.14 (1.37, 7.21)	3.17 (1.36, 7.38)

Definition of abbreviations: CI = confidence interval; OR = odds ratio; TB = tuberculosis.

* Adjusted for potential confounding of age, sex, and HIV status using multivariate logistic regression model.

and deletion events. Thus, it is possible that some wild-type isolates had small genetic alterations, such as point mutations, or small deletions that could reduce gene function. If this did occur, it could cause misclassification of the *plcD* genotype, which, in turn, might have weakened the strength of the association. Future investigations of single nucleotide changes and small deletions in this region and further functional analysis of different mutants in comparison with the wild-type isolates will add to our understanding of the role of the deleted genes in the virulence of *M. tuberculosis*.

An interesting observation in this study is that a significantly higher proportion of patients having isolates with *plcD* mutations lived within a city limit. This raises the question of whether or not phospholipase C is involved in the airborne transmissibility of *M. tuberculosis*. The *plcD* mutations were not found significantly more frequently in any zip code, city, or county in this study (data not shown). The possibility of other confounders that might differ between city and non-city dwellers was explored by analyzing the geographic distribution of thoracic and extrathoracic TB and their respective proportion included in this study; no significant differences were found in the comparison ($p = 0.84$). A future study using epidemiologic linkage between/among clustered patients would allow us to study this question more thoroughly.

Several host-related risk factors for extrathoracic TB have been reported previously (20, 28–30). However, this report is the first to observe a microbial change acting as an independent risk factor for extrathoracic TB. Like all epidemiologic studies, this study does not prove causal relationship. However, it does provide a rationale for the selection of gene targets for future functional studies aimed at identifying *M. tuberculosis* virulence factors. The usefulness of combining comparative genomics with epidemiologic data to study the pathogenesis of tuberculosis is exemplified by this report.

Conflict of Interest Statement: Z.Y. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; D.Y. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; Y.K. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; L.Z. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; C.F.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; B.F. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; J.H.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; F.W. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; M.D.C. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank Annadell H. Fowler, Leonard N Mukasa, Bill Starrett, Deborah Witoski, Peter J. Boldenow, and Patricia C. Juliao for their valuable efforts during the study. They acknowledge Dr. Kashef Ijaz's contribution to the establishment of the Arkansas Department of Health's surveillance database that was used for the study. They also thank Drs. Jack T. Crawford and Laura S. Cowan at the Centers for Disease Control and Prevention, Atlanta, GA, for providing the DNA preparation of CDC1551.

References

1. World Health Organization. Global tuberculosis control surveillance, planning, financing: a report of WHO. Geneva, Switzerland: World Health Organization; 2004. WHO Publication No. WHO/HTM/TB/2004.331.
2. Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. *Lancet* 1995;346:1339–1345.
3. Espinal MA. The global situation of MDR-TB. *Tuberculosis (Edinb)* 2003;83:44–51.
4. Godfrey-Faussett P, Ayles H. Can we control tuberculosis in high HIV prevalence settings? *Tuberculosis (Edinb)* 2003;83:68–76.
5. Smith I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* 2003;16:463–496.
6. Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol* 1999;32:643–655.
7. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 1999;284:1520–1523.
8. Kato-Maeda M, Bifani PJ, Kreiswirth BN, Small PM. The nature and consequence of genetic variability within *Mycobacterium tuberculosis*. *J Clin Invest* 2001;107:533–537.
9. Brosch R, Pym AS, Gordon SV, Cole ST. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol* 2001;9:452–458.
10. Mostowy S, Behr MA. Comparative genomics in the fight against tuberculosis: diagnostics, epidemiology, and BCG vaccination. *Am J Pharmacogenomics* 2002;2:189–196.
11. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere YO, Aman K, Kato-Maeda M, Small PM. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci USA* 2004;101:4865–4870.
12. Ho TB, Robertson BD, Taylor GM, Shaw RJ, Young DB. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast* 2000;17:272–282.
13. Segal W. Growth dynamics of *in vivo* and *in vitro* grown mycobacterial pathogens. In: Kubica GP, Wayne LG., editors. *The mycobacteria: a sourcebook*. New York: Dekker; 1984. pp. 547–573.
14. Wheeler PR, Bulmer K, Ratledge C. Enzymes for biosynthesis de novo and elongation of fatty acids in mycobacteria grown in host cells: is *Mycobacterium leprae* competent in fatty acid biosynthesis? *J Gen Microbiol* 1990;136:211–217.
15. Songer JG. Bacterial phospholipases and their role in virulence. *Trends Microbiol* 1997;5:156–161.
16. Raynaud C, Guilhot C, Rauzier J, Bordat Y, Pelicic V, Manganelli R, Smith I, Gicquel B, Jackson M. Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* 2002;45:203–217.
17. Yang ZH, Yang D, Kong Y, Zhang L, Marrs CF, Foxman B, Bates JH, Wilson F, Cave MD. Clinical implication of *Mycobacterium tuberculosis* phospholipase C gene (*plcD*) polymorphisms [abstract]. Presented at the ASM General Meeting, New Orleans, LA, 2004.
18. Murray MG, Thompson WF. Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Res* 1980;8:4321–4325.
19. Yang Z. Molecular epidemiology of tuberculosis. *Front Biosci* 2003;8:d440–d450.
20. Yang Z, Kong Y, Wilson F, Foxman B, Fowler AH, Marrs CF, Cave

- MD, Bates JH. Identification of risk factors for extrapulmonary tuberculosis. *Clin Infect Dis* 2004;38:199–205.
21. Boddingtonhaus B, Rogall T, Flohr T, Blocker H, Bottger EC. Detection and identification of mycobacteria by amplification of rRNA. *J Clin Microbiol* 1990;28:1751–1759.
22. Valway SE, Sanchez MPC, Shinnick TF, Orme I, Agerton T, Hoy D, Jones JS, Westmoreland H, Onorato IM. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* 1998;338:633–639.
23. Thierrv D, Brisson-Noel A, Vincent-Levv-Frebault V, Nguven S, Guesdon J, Gicquel B. Characterization of a *Mycobacterium tuberculosis* insertion sequence, *IS6110*, and its application in diagnosis. *J Clin Microbiol* 1990;28:2668–2673.
24. Cody RPSJ. Applied statistics and the SAS programming language. Upper Saddle River, NJ: Prentice-Hall; 1997.
25. Viana-Niero C, De Haas PE, Van Soolingen D, Leao SC. Analysis of genetic polymorphisms affecting the four phospholipase C (plc) genes in *Mycobacterium tuberculosis* complex clinical isolates. *Microbiology* 2004;150:967–978.
26. Talarico S, Durmaz R, Yang Z. Insertion and deletion-associated genetic diversity of *Mycobacterium tuberculosis* phospholipase C-encoding genes among 106 clinical isolates from Turkey. *J Clin Microbiol* 2005; 43(2):533–538.
27. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, Howard ST. IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol Microbiol* 2004;52:999–1012.
28. Rieder HL, Snider DE Jr, Cauthen GM. Extrapulmonary tuberculosis in the United States. *Am Rev Respir Dis* 1990;141:347–351.
29. Gonzalez OY, Adams G, Teeter LD, Bui TT, Musser JM, Graviss EA. Extra-pulmonary manifestations in a large metropolitan area with a low incidence of tuberculosis. *Int J Tuberc Lung Dis* 2003;7:1178–1185.
30. Ong A, Creasman J, Hopewell PC, Gonzalez LC, Wong M, Jasmer RM, Daley CL. A molecular epidemiological assessment of extrapulmonary tuberculosis in San Francisco. *Clin Infect Dis* 2004;38:25–31.