EMERGING INFECTIOUS DISEASES

EID Online
www.cdc.gov/eid

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends Vol.8, No.11, November 2002

Reemerging Tuberculosis

In Index Medicus/Medline, Current Contents, Excerpta Medica, and other databases

Editorial Board

Dennis Alexander, Addlestone Surrey, United Kingdom
Ban Allos, Nashville, Tennesee, USA
Michael Apicella, Iowa City, Iowa, USA
Abdu F. Azad, Baltimore, Maryland, USA
Ben Beard, Atlanta, Georgia, USA
Barry J. Beaty, Ft. Collins, Colorado, USA
Martin J. Blaser, New York, New York, USA
S.P. Borriello, London, United Kingdom
Donald S. Burke, Baltimore, Maryland, USA
Charles Calisher, Ft. Collins, Colorado, USA
Arturo Casadevall, Bronx, New York, USA
Thomas Cleary, Houston, Texas, USA
Anne DeGroot, Providence, Rhode Island, USA
Vincent Deubel, Lyon, France
J. Stephen Dumler, Baltimore, Maryland, USA
Durland Fish, New Haven, Connecticut, USA
Richard L. Guerrant, Charlottesville, Virginia, USA
Scott Halstead, Arlington, Virginia, USA
Seyed Hasnain, Hyderabad, India
David L. Heymann, Geneva, Switzerland
Sakae Inouye, Tokyo, Japan
Peter B. Jahrling, Frederick, Maryland, USA
Mohamed A. Karmali, Guelph, Ontario, Canada
Charles King, Cleveland, Ohio, USA
Keith Klugman, Atlanta, Georgia, USA
Bruce R. Levin, Atlanta, Georgia, USA
Myron Levine, Baltimore, Maryland, USA
Stuart Levy, Boston, Massachusetts, USA
Thomas J. Marrie, Edmonton, Alberta, Canada
John E. McGowan, Jr., Atlanta, Georgia, USA
Patrick S. Moore, New York, New York, USA
Philip P. Mortimer, London, United Kingdom
Fred A. Murphy, Davis, California, USA
Barbara E. Murray, Houston, Texas, USA
P. Keith Murray, Ames, Iowa, USA
James M. Musser, Hamilton, Missouri, USA
Rosanna W. Peeling, Geneva, Switzerland
David H. Persing, Seattle, Washington, USA
Gianfranco Pezzino, Topeka, Kansas, USA
Richard Platt, Boston, Massachusetts, USA
Didier Raoult, Marseille, France
Leslie Real, Atlanta, Georgia, USA
David Relman, Palo Alto, California, USA
Pierre Rollin, Atlanta, Georgia, USA
Nancy Rosenstein, Atlanta, Georgia, USA
Connie Schmaljohn, Frederick, Maryland, USA
Ira Schwartz, Valhalla, New York, USA
Robert Shope, Galveston, Texas, USA
Bonnie Smoak, Bethesda, Maryland, USA
Rosemary Soave, New York, New York, USA
P. Frederick Sparling, Chapel Hill, North Carolina, USA
Jan Svoboda, Prague, Czech Republic
Robert Swanepoel, Sandringham, South Africa
Phillip Tarr, Seattle, Washington, USA
Timothy Tucker, Cape Town, South Africa
Elaine Tuomanen, Memphis, Tennessee, USA
David Walker, Galveston, Texas, USA
Mary E. Wilson, Cambridge, Massachusetts, USA

Editors

D. Peter Drotman, Interim Editor-in-Chief
Atlanta, Georgia, USA

Stephen S. Morse, Perspectives Editor
New York, New York, USA

Brian W.J. Mahy, Perspectives Editor
Atlanta, Georgia, USA

Phillip J. Baker, Synopses Editor
Bethesda, Maryland, USA

Stephen Ostroff, Dispatches Editor
Atlanta, Georgia, USA

Patricia M. Quinlisk, Letters Editor
Des Moines, Iowa, USA

Polyxeni Potter, Managing Editor
Atlanta, Georgia, USA

Joseph E. McDade, Founding Editor
Atlanta, Georgia, USA

International Editors

Patrice Courvalin
Paris, France

Takeshi Kurata
Tokyo, Japan

S.K. Lam
Kuala Lumpur, Malaysia

John S. Mackenzie
Brisbane, Australia

Hooman Momen
Rio de Janeiro, Brazil

Sergey V. Netesov
Novosibirsk Region, Russian Federation

Diana Walford
London, United Kingdom

Production Editors

Kevin Burlison
Mary Anne Castranio
Teresa M. Hood
Anne D. Mather
Carol D. Snarey
Reginald S. Tucker
Cathy E. Young

Electronic Access

Retrieve the journal electronically on the World Wide Web (WWW) at http://www.cdc.gov/eid or from the CDC home page (http://www.cdc.gov).

Announcements of new table of contents can be automatically e-mailed to you. To subscribe, send an e-mail to listserv@cdc.gov with the following in the body of your message: subscribe EID-TOC.

Emerging Infectious Diseases

Emerging Infectious Diseases is published monthly by the National Center for Infectious Diseases, Centers for Disease Control and Prevention (CDC), 1600 Clifton Road, Mailstop D61, Atlanta, GA 30333, USA. Telephone 404-371-5329, fax 404-371-5449, e-mail eideditor@cdc.gov.

All material published in Emerging Infectious Diseases is in the public domain and may be used and reprinted without special permission; proper citation, however, is appreciated.

Use of trade names is for identification only and does not imply endorsement by the Public Health Service or by the U.S. Department of Health and Human Services.

∞ Emerging Infectious Diseases is printed on acid-free paper that meets the requirements of ANSI/NISO 239.48-1992 (Permanence of Paper)

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated.

EMERGING INFECTIOUS DISEASES

Full text free online at www.cdc.gov/eid

The print journal is available at **no charge** to public health professionals

☐ YES, I would like to receive Emerging Infectious Diseases.

Please print your name and business address in the box and return by fax to 404-371-5449 or mail to
EID Editor
CDC/NCID/MS D61
1600 Clifton Road, NE
Atlanta, GA 30333

Moving? Please give us your new address (in the box) and print the number of your old mailing label here_______________

EID Online
www.cdc.gov/eid

EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends Vol.8, No.11, November 2002

On the Cover: Amedeo Modigliani (1884-1920). Self-Portrait, 1919. Oil on canvas 100 cm x 65 cm.
Museu de Arte Contemporanea da Universidade de Sao Paulo, Brazil

About the Cover
see pg 1364

Tuberculosis Genotyping Network

The Continued Threat of Tuberculosis 1187
T. Navin et al.

Rationale and Methods for the National Tuberculosis Genotyping and Surveillance Network 1188
K.G. Castro and H.W. Jaffe

National Tuberculosis Genotyping and Surveillance Network: Design and Methods 1192
J.T. Crawford et al.

Molecular Epidemiology of Tuberculosis in a Sentinel Surveillance Population 1197
B.A. Ellis et al.

Quality Assessment of *Mycobacterium tuberculosis* Genotyping in a Large Laboratory Network 1210
C.R. Braden et al.

Identifying the Sources of Tuberculosis in Young Children: A Multistate Investigation 1216
S.J. Sun et al.

DNA Fingerprinting of *Mycobacterium tuberculosis* Isolates from Epidemiologically Linked Case Pairs 1224
D.E. Bennett et al.

Molecular Epidemiology of Multidrug-Resistant Tuberculosis, New York City, 1995–1997 1230
S.S. Munsiff et al.

Genotyping Analyses of Tuberculosis Cases in U.S.- and Foreign-Born Massachusetts Residents 1239
S. Sharnprapai et al.

Transmission of *Mycobacterium tuberculosis* in a Rural Community, Arkansas, 1945–2000 1246
J.A. Dillaha et al.

Cross-Jurisdictional Transmission of *Mycobacterium tuberculosis* in Maryland and Washington, D.C., 1996–2000, Linked to the Homeless 1249
M. Lathan et al.

Use of DNA Fingerprinting To Investigate a Multiyear, Multistate Tuberculosis Outbreak 1252
P.D. McElroy et al.

Mycobacterium tuberculosis Transmission between Cluster Members with Similar Fingerprint Patterns 1257
K. Ijaz et al.

A Prospective, Multicenter Study of Laboratory Cross-Contamination of *Mycobacterium tuberculosis* Cultures 1260
R.M. Jasmer et al.

Estimated Costs of False Laboratory Diagnoses of Tuberculosis in Three Patients 1264
J.M. Northrup et al.

Molecular Epidemiology of Tuberculosis in a Low- to Moderate-Incidence State: Are Contact Investigations Enough? 1271
W.A. Cronin et al.

Skin-Test Screening and Tuberculosis Transmission among the Homeless 1280
P.-M. Kong et al.

This issue of Emerging Infectious Diseases was made possible through a partnership with the CDC Foundation with financial support provided by **The Ellison Medical Foundation**.

EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends Vol.8, No.11, November 2002

Impact of *Mycobacterium tuberculosis* Genotyping on Public Health Practice in Massachusetts 1285
A.C. Miller et al.

Human Exposure Following *Mycobacterium tuberculosis* Infection of Multiple Animal Species in a Metropolitan Zoo 1290
P. Oh et al.

Genotype Analysis of *Mycobacterium tuberculosis* Isolates from a Sentinel Surveillance Population 1294
L.S. Cowan and J.T. Crawford

Molecular Typing of *Mycobacterium tuberculosis* Strains with a Common Two-Band IS*6110* Pattern 1303
K.H. Lok et al.

Spoligologos: A Bioinformatic Approach to Displaying and Analyzing *Mycobacterium tuberculosis* Data 1306
J.R. Driscoll et al.

Molecular Differentiation of *Mycobacterium tuberculosis* Strains without IS*6110* Insertions 1310
K.H. Lok et al.

DNA Fingerprinting of *Mycobacterium tuberculosis*: Lessons Learned and Implications for the Future 1314
S.J.N. McNabb et al.

Research

Rifampin- and Multidrug-Resistant Tuberculosis in Russian Civilians and Prison Inmates: Dominance of the Beijing Strain Family 1320
Francis Drobniewski et al.

Tuberculosis-Related Deaths within a Well-Functioning DOTS Control Program 1327
M. de Lourdes García-García et al.

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated.

Molecular Analysis of Sarcoidosis Tissues for *Mycobacterium* Species DNA 1334
W.P. Drake et al.

Dispatches

Genomewide Pattern of Synonymous Nucleotide Substitution in Two Complete Genomes of *Mycobacterium tuberculosis* 1342
A. Hughes et al.

Global Distribution of *Mycobacterium tuberculosis* Spoligotypes 1347
I. Filliol et al.

Two Cases of Pulmonary Tuberculosis Caused by *Mycobacterium tuberculosis* subsp. *canetti* 1350
J. Miltgen et al.

Another Dimension

At the Deathbed of Consumptive Art 1353
D.M. Morens

Letters

Mycobacterium haemophilum: Emerging or Underdiagnosed in Brazil? 1359
J.L.M. Sampaio et al.

Children and Multidrug-Resistant Tuberculosis in Mumbai (Bombay), India 1360
S. Karande and S.B. Bavdekar

News and Notes

Book Review: Cestode Zoonoses: Echinococcosis and Cysticercosis: An Emergent and Global Problem (P. Craig and Z. Pawlowski, editors) 1362
F.O. Richards, Jr.

Book Review: Immunology of Infectious Diseases (S.H.E. Kaufmann et al., authors) 1362
T. Lazar

About the Cover 1364
P. Potter

The Continued Threat of Tuberculosis

Thomas R. Navin,* Scott J.N. McNabb,* and Jack T. Crawford*

Why would a journal that tracks and analyzes emerging infectious disease trends devote an entire issue to tuberculosis, a disease that emerged some 15,000 to 35,000 years ago (1,2)? The disturbing answer is that tuberculosis is reappearing in many countries as a public health crisis. Thus, if not an emerging disease, it is an important reemerging disease, and though ancient, it is not a disease of the past. A staggering 1.9 million around the globe die of tuberculosis each year—another 1.9 billion are infected with *Mycobacterium tuberculosis* and are at risk for active disease (3).

In the 20th century, the United States made impressive strides in tuberculosis control. From the early 1900s, when some areas began systematic reporting of death rates, tuberculosis rates steadily declined from approximately 200 deaths per 100,000 per year to less than 1 death per 100,000 in 1985. In 1953, a national surveillance system was established for reporting new cases of tuberculosis disease; that year, reported annual incidence was 53 cases per 100,000 population (4). From 1953 to 1984, tuberculosis disease incidence dropped steadily at an average rate of 5.8% per year to 9.4 cases per 100,000.

In 1985, however, the United States saw a reversal in this long-standing downward trend, and tuberculosis reemerged as a public health threat. From 1985 to 1992, not only did the number of cases increase from 22,201 to 26,673, but also large outbreaks were reported. Many of these, especially in hospitals and other health-care settings in large cities (5), were caused by multidrug-resistant *M. tuberculosis*. Several factors contributed to this increase, including the emergence of the HIV epidemic and large influxes of immigrants from countries in which tuberculosis was common. Perhaps the major reason for the reemergence, however, was the end in 1972 of categorical federal funding for control activities and the subsequent deterioration of public health infrastructure for tuberculosis control.

In response to the crisis of reemerging tuberculosis, categorical grants were restored and federal funding was increased. The funding, modest at first, rose sharply in 1992 and again in 1993 and 1994. The Centers for Disease Control and Prevention (CDC) transfers most of its appropriated funds to tuberculosis control programs in states and large cities. These funds support clinics and laboratories, administer directly observed therapy, intensify investigation of latent infection in persons at high risk for active disease, sponsor clinical and epidemiologic research, and expand surveillance to monitor the impact of these efforts. Renewed investments paid off, and after a peak in 1992, tuberculosis incidence in the United States has declined each year. From 1992 to 2001, the annual decline averaged 7.3%, even greater than before 1985. But future success is not guaranteed. The National Academy of Sciences Institute of Medicine, in its 2000 report on tuberculosis control efforts in the United States, warned against the "complacency and neglect" that come with declining numbers of cases and reaffirmed the goal of TB elimination (annual incidence of <1 case per 1,000,000 population) in the United States (6).

In 2001, the 15,989 tuberculosis cases reported to CDC represented only a 2% decline from 2000, the smallest decline in 9 years. Although data from a single year do not constitute a trend, these numbers may be the first sign of stagnation in our control efforts. The proportion of cases in persons born outside the United States is growing; in 2001, that figure reached 50%. Efforts to reduce tuberculosis transmission in the United States have little effect on reducing risk for those infected elsewhere. The proportion of cases in persons born in other countries will probably continue to rise, unless domestic programs providing tuberculosis services for immigrants are strengthened and international programs are expanded. Another risk, in the current climate of bioterrorism, is the possible intentional spread of multidrug-resistant *M. tuberculosis.* This risk requires new tools for detection and rapid and effective response. Currently strengthened surveillance systems closely monitor changes in disease epidemiology. If tuberculosis elimination progress in the United States slows, we are prepared to respond quickly.

References

1. Kapur V, Whittam TS, Musser JM. Is *Mycobacterium tuberculosis* 15,000 years old? J Infect Dis 1994;170:1348–9.
2. Hughes AL, Friedman R, Murray M. Genomewide pattern of synonymous nucleotide substitution in two complete genomes of *Mycobacterium tuberculosis*. Emerg Infect Dis 2002;8;1342–46.
3. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC, for the WHO Global Surveillance and Monitoring Project. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. JAMA 1999;282:677–86.
4. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2001. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, September 2002.
5. Cantwell MF, Snider DE, Cauthen GM, Onorato IM. Epidemiology of tuberculosis in the United States, 1985 through 1992. JAMA 1994;272:535–9.
6. Institute of Medicine. Ending neglect: the elimination of tuberculosis in the United States. Washington: National Academy Press, 2000.

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Rationale and Methods for the National Tuberculosis Genotyping and Surveillance Network

Kenneth G. Castro* and Harold W. Jaffe*

Our understanding of tuberculosis (TB) transmission dynamics has been refined by genotyping of *Mycobacterium tuberculosis* strains. The National Tuberculosis Genotyping and Surveillance Network was designed and implemented to systematically evaluate the role of genotyping technology in improving TB prevention and control activities. Genotyping proved a useful adjunct to investigations of outbreaks, unusual clusters, and laboratory cross-contamination.

"In the future, the battle against this plague of mankind will not just be concerned with an uncertain something but with a tangible parasite, about whose characteristics a great deal is known and can be explored."

—Robert Koch, 1882

Molecular genetic typing (or genotyping) of *Mycobacterium tuberculosis* strains has revolutionized the field of tuberculosis (TB) research, prevention, and control (1–3). The subtypes characterized by molecular genetic typing methods provide a greater power and ability to differentiate strains than previous methods, such as comparisons of patterns of drug resistance or phage-typing (4,5). When molecular genotyping technology is applied to outbreaks or unusual clusters of disease, persons sharing *M. tuberculosis* strains can be identified, which can lead to important clues about the pattern and dynamics of transmission.

Methods of molecular genotyping have been increasingly applied to the epidemiology of TB. Because of its natural history, the transmission of *M. tuberculosis* is difficult to study; *M. tuberculosis* is spread by airborne droplets of respiratory secretions expelled by an infectious person to a susceptible host, who may or may not be known to the source (6). The bacterium can remain latent as an asymptomatic infection for years, and the source of such infections can be difficult to ascertain. Thus, the places and persons involved in a chain of transmission may be puzzling to identify or exclude. Molecular typing of *M. tuberculosis* adds important pieces to the construction of such a transmission puzzle; persons who harbor the same strain of *M. tuberculosis* are likely to have shared that strain in a common chain of transmission; by contrast, persons who are infected by unique and distinct strains were probably infected by means of a different exposure. Among its applications, genotyping has served to elucidate the poorly understood role of relapses and exogenous reinfection of persons with recurrent TB after cure. Several reports have relied on DNA genotyping to describe and document the occurrence of exogenous reinfection with distinct strains of *M. tuberculosis* as the cause of TB following successful treatment (7–9).

The usefulness of molecular typing was also confirmed in several epidemiologic investigations of HIV-associated multidrug-resistant TB outbreaks in hospices, hospitals, and prisons during the late 1980s and early 1990s and provided compelling evidence of institutional transmission (10–16). Unique features of these outbreaks included prolonged infectiousness of patients who were not recognized to harbor multidrug-resistant TB until months after their TB was diagnosed and the relatively rapid progression from latent infection to active TB disease in persons with HIV-associated immunosuppression. These data were also used to state the need to implement effective interventions to halt such outbreaks (17,18).

In addition to the use of DNA genotyping during outbreak investigations, the technology has been applied as a complementary tool to conventional methods in TB control (19–22). In two of the earliest studies conducted in the United States, one in San Francisco and one in New York City, the authors assumed that *M. tuberculosis* isolates with matching DNA fingerprints were epidemiologically related and represented recent transmission of *M. tuberculosis* among the patients involved (i.e., within 2 years before diagnosis) (19–20). In these studies, 30% to 40% of the patients had *M. tuberculosis* isolates with DNA fingerprint patterns that matched at least one other isolate. This finding led the authors to conclude that as many as 40% of TB cases in these two cities were the result of recent transmission and that TB control practices in San Francisco and New York were not effectively decreasing *M. tuberculosis* transmission. The observations in these two reports were useful and innovative. However, the findings could not be generalized to other geographic areas in the United States because the study populations were exclusively urban residents, a large proportion were HIV infected, and detailed epidemiologic information

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

was incomplete for either study, which limited the ability to corroborate actual contacts or exposures among the TB patients studied.

A separate study evaluated the use of DNA genotyping in TB patients from a large, rural population in the state of Arkansas (21). Analysis of *M. tuberculosis* isolates from TB patients for a 2-year period (1992–1993) found that more than half the isolates matched at least one other patient isolate. Epidemiologic investigation of the patients with matching *M. tuberculosis* isolates (i.e., clustered cases) revealed 24 persons who had documented latent TB infection or active TB many years in the past but which produced disease with matching isolates at the time of the study. In five of these patients, a remote epidemiologic connection (i.e., common exposures to a person with TB) was identified that occurred 20–25 years earlier. The authors also reported that some TB patients had isolates showing identical specific patterns, yet their geographic, social, and medical histories were so disparate that transmission among them was highly unlikely. Further investigation by genotyping with an independent method revealed that the capacity of the IS*6110* method to differentiate strains was roughly proportional to the number of bands present in the original fingerprint pattern. Thus, isolates with a low number of bands (e.g., fewer than five) required a second method for appropriate differentiation. These findings suggested the need for additional assessments and evaluations of emerging assumptions in the interpretation of DNA genotyping of *M. tuberculosis*.

The fruitful use of DNA genotyping to confirm and refine our understanding of *M. tuberculosis* transmission in outbreaks provided the major impetus to evaluate the use of this technology in other settings and to determine its broader application as a tool in TB prevention and control. Specifically, we sought to assess the usefulness of this technology in searching for unidentified outbreaks, identifying risk factors for TB at the population level, and identifying and monitoring laboratory cross-contamination (23). Consequently, the Centers for Disease Control and Prevention (CDC) funded six laboratories in 1993 to develop regional databases of DNA genotype fingerprint patterns and undertake regional molecular epidemiologic studies. This genotyping network was expanded in 1996 to include sentinel surveillance sampling. Ultimately, the National Tuberculosis Genotyping and Surveillance Network comprised CDC, seven laboratories, and seven sentinel surveillance sites in the United States.

Objectives and Composition of the National Tuberculosis Genotyping and Surveillance Network

Successful applicants for a cooperative agreement with CDC formed part of the genotyping network. The following six potential objectives were to be the focus of activities by the network:

1. Determine the Relative Frequency of *M. tuberculosis* Strains on the Basis of DNA Fingerprint Patterns by Using the IS*6110* Method in Specific Geographic Areas. This determination was meant to characterize the diversity of strains in any one area or region and allow for more accurate interpretation of results of DNA fingerprint analysis.

2. Determine the Extent of Spread of Related *M. tuberculosis* Strains in Communities. As a secondary objective, the identification of common and potentially more transmissible strains could aid in the study of *M. tuberculosis* pathogenesis and host immunity.

3. Describe the Geographic Mobility of Related *M. tuberculosis* Strains and the Mode in Which Strains Spread. The characterization of places and activities involved in potential transmission could enable TB control programs to design interventions accordingly.

4. Determine the Relatedness of *M. tuberculosis* Isolates in Patients Who Are Identified as Being a High Risk for TB through Conventional Epidemiologic Studies. DNA fingerprint clustering of isolates among groups at high risk could represent a relatively more specific marker of recent transmission when compared to clustering identified in the general population.

5. Develop the Capacity of Local TB Controllers To Identify Patients with Related *M. tuberculosis* Organisms Who Deserve Careful Consideration and Investigation To Identify Ongoing Transmission. A secondary objective was to assess the role of fingerprinting in helping to prioritize and focus contact investigations.

6. Assess Use of DNA Fingerprinting Analyses in Guiding TB Control Activities, such as Targeted Testing and Treatment of Latent TB Infection and Monitoring Possibilities for Transmission in Congregate Settings such as Hospitals and Prisons.

Federal funds were provided for the TB genotyping network to establish a core set of databases at each of the laboratories and sentinel surveillance sites and a national database at CDC. The laboratory databases included computerized images of DNA fingerprint patterns from all *M. tuberculosis* isolates analyzed for their region and for all isolates analyzed for their sentinel surveillance site. The databases at the sentinel sites included a record for each sentinel area resident who was diagnosed with culture-positive TB. The record contained information collected as part of routine TB national surveillance activities; the identification of source or secondary cases, if known; and the DNA fingerprint pattern designation. Routine surveillance for TB included information for each patient concerning demographics, social and occupational risk factors for TB; clinical and radiologic details of disease; culture, strain, and histology results; susceptibility testing of isolates; and antibiotic treatment regimens and clinical outcome. At CDC, DNA fingerprint images, surveillance, and epidemiologic information were combined from all laboratories and sites to create national databases of sentinel site patients and a library of all unique DNA fingerprint patterns among isolates from sentinel surveillance site patients.

Sentinel Surveillance Sites and Regional Laboratories

Sentinel surveillance sites in the network included the states of Arkansas, Maryland, Massachusetts, Michigan, and

New Jersey; six counties in California (Alameda, Contra Costa, Marin, San Mateo, Santa Clara, and Solano); and four counties in Texas (Dallas, Tarrant, Cameron, and Hidalgo). All patients within those areas were included on a prospective basis. The sentinel surveillance sites were selected on the basis of applications by state and large-city departments of health and by characteristics of the proposed sentinel populations.

The following criteria were used to evaluate sites applying through competitive proposals: 1) understand the use of *M. tuberculosis* DNA fingerprinting in the epidemiology of TB; 2) report at least 250 TB cases per year, and submit one *M. tuberculosis* isolate for 75% of culture-positive patients in their areas for DNA fingerprinting; 3) conduct active surveillance of TB cases; 4) review information from the national TB surveillance database and make every effort to ensure that data are complete; 5) establish and maintain a surveillance site database; and 6) maintain records of activities performed as part of the genotyping network.

The network sentinel surveillance site relied on local mycobacteriology and hospital infection control records for all facilities in the surveillance site areas for case finding. Other sources of information included hospital IDC-9 discharge codes for TB, pharmacy records for prescriptions of a combination of two or more anti-TB drugs, coroners' records that showed TB as a diagnosis, and AIDS surveillance reports that indicated a diagnosis of TB. Sentinel surveillance site personnel reviewed the information in their national surveillance database to ensure that all information was complete to the extent possible.

The regional laboratories were also selected on a competitive basis. They were responsible for providing DNA fingerprint analysis of *M. tuberculosis* isolates from health departments in their region and for their assigned sentinel surveillance site. Regions were assigned to laboratories on the basis of history of their work and approximate numbers of TB patients in the regions. Each culture-positive TB case normally reported for national TB surveillance (on the form, Report of a Verified Case of Tuberculosis) within the sentinel site area was included as a sentinel surveillance case. An isolate from each culture-positive TB patient was sent for DNA fingerprinting at the designated regional DNA fingerprinting laboratory. *M. tuberculosis* isolates were shipped from sentinel sites to laboratories in approved shipping containers that were appropriately labeled and handled in accordance with bio-safety level 3 conditions. Subcultures of isolates were to be stored by the regional laboratories at –70°C in duplicate indefinitely.

At the regional laboratories, images of individual DNA fingerprint patterns of *M. tuberculosis* isolates were generated by using standardized procedures for DNA extraction, purification, digestion, electrophoresis, hybridization, and computerization. Individual images were electronically transmitted to CDC to be assigned a national DNA fingerprint designation. Each submitted fingerprint pattern was compared to previously submitted patterns. Unique patterns were added to the database and assigned consecutive five-digit numbers as their national designations. Results of analysis of isolates and DNA fingerprint designations were reported back to the regional DNA fingerprinting laboratories on a regular basis. The results of DNA fingerprint analysis of sentinel surveillance TB case isolates, including the national DNA fingerprint database pattern designations, were transmitted from the regional laboratories to the sentinel surveillance sites.

This special issue of Emerging Infectious Diseases contains the reports of the various analyses resulting from this highly productive collaboration of the National Tuberculosis Genotyping and Surveillance Network. The results represent a remarkable accomplishment and provide the scientific basis for future potential applications of DNA genotyping as part of population-based TB prevention and control activities in the United States. The results also highlight both the strengths and limitations of DNA genotyping as an adjunct to TB outbreak and contact investigations and assessments of laboratory cross-contamination. Although the network has been very productive, additional technologic advances are necessary as well as improvements in the understanding and use of principles and practices from other disciplines, such as social network analyses, before we can reliably obtain real-time laboratory results and improve our understanding of events facilitating the transmission of *M. tuberculosis* in modern societies. Such a comprehensive approach that combines both basic and operational research must be supported so that our efforts will ultimately result in the elimination of TB (24).

Acknowledgments

We are indebted to Christopher Braden, Jack Crawford, Ida Onorato, and Barbara Schable for their roles in the early design and implementation of the National Tuberculosis Genotyping and Surveillance Network.

Dr. Castro is director, Division of Tuberculosis Elimination, in the National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention. His research interests include the epidemiology, transmission dynamics, and clinical and diagnostic trials of TB, multidrug-resistant TB, and HIV-related TB.

Dr. Jaffe has served as director of the Division of AIDS, STD, and TB Laboratory Research in the National Center for Infectious Diseases. He is currently the director of the National Center for HIV, STD, and TB Prevention. His research interests include the epidemiology and molecular epidemiology of HIV, sexually transmitted diseases, and TB.

References

1. Cave MD, Eisenach KD, McDermott PF, Bates JH, Crawford JT. IS*6110*: Conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. Mol Cell Probes 1991;5:73–80.
2. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Giquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.

3. Mazurek GH, Cave MD, Eisenach KD, Wallace RJ, Bates JH, Crawford JT. Chromosomal DNA fingerprint patterns produced with IS*6110* as strain-specific markers for epidemiologic study of tuberculosis. J Clin Microbiol 1991;29:2030–1.
4. Bates JH, Mitchinson DA. Geographic distribution of bacteriophage types of *Mycobacterium tuberculosis*. Am Rev Respir Dis 1969;100:189–93.
5. Rado TA, Bates JH, Engel HW, Mankiewicz E, Murohashi T, Mizuguchi Y, et al. World Health Organization studies on bacteriophage typing of mycobacteria: subdivision of the species *Mycobacterium tuberculosis*. Am Rev Respir Dis 1975;111:459–68.
6. Riley R, Mills C, Nyka W. Aerial dissemination of pulmonary tuberculosis—a two-year study of contagion in a tuberculosis ward. Am J Hyg 1959;70:185–96.
7. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, et al. Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. N Engl J Med 1993;328:1137–44.
8. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, et al. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med 1999;341:1174–9.
9. Sonnenberg P, Murray J, Glynn JR, Shearer S, Kambashi B, Godfrey-Faussett P. HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. Lancet 2001;358:1687–93.
10. Centers for Disease Control and Prevention. Transmission of multidrug-resistant tuberculosis among HIV-infected persons—Florida and New York, 1988-1991. MMWR Morb Mortal Wkly Rep 1991;40:585–91.
11. Daley CL, Small PM, Schecter GF, Schoolnick GK, McAdam RA, Jacobs WR Jr, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. N Engl J Med 1992;326:231–5.
12. Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams J, Sordillo EM, et al. An outbreak of multidrug resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
13. Beck-Sague C, Dooley SW, Hutton MD, Otten J, Breeden A, Crawford JT, et al. Hospital outbreak of multidrug-resistant *Mycobacterium tuberculosis* infections. JAMA 1992;268:1280–6.
14. Coronado VG, Beck-Sague CM, Hutton MD, Davis BJ, Nicholas P, Villareal C, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* among persons with human immunodeficiency virus infection in an urban hospital. J Infect Dis 1993;168:1052–5.
15. Centers for Disease Control and Prevention. Transmission of multidrug-resistant tuberculosis among immunocompromised persons in a correctional system—New York, 1991. MMWR Morb Mortal Wkly Rep 1992;41:507–9.
16. Valway SE, Richards SB, Kovakovich J, Greifinger RB, Crawford JT, Dooley SW. Outbreak of multidrug-resistant tuberculosis in a New York State Prison, 1991. Am J Epidemiol 1994;140:113–22.
17. Centers for Disease Control and Prevention. Guidelines for preventing the transmission of tuberculosis in health-care settings with special focus on HIV-related issues. MMWR Morb Mortal Wkly Rep 1990;39(No. RR-1):1–29.
18. Centers for Disease Control and Prevention. Guidelines for preventing the transmission of *Mycobacterium tuberculosis* in health-care facilities, 1994. MMWR Morb Mortal Wkly Rep 1994;43(No. RR-13):1–132.
19. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
20. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City—an analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–16.
21. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, 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.
22. van Deutekom H, Gerritsen JJ, van Soolingen D, van Ameijden EJ, van Embden JD, Coutinho RA. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. Clin Infect Dis 1997;25:1071–7.
23. Braden CR, Templeton GL, Stead WW, Bates JH, Cave MD, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
24. Institute of Medicine. Ending neglect: the elimination of tuberculosis in the United States. Washington: National Academy Press; 2000. p. 1–269.

Address for correspondence: Kenneth G. Castro, Division of Tuberculosis Elimination, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E10, Atlanta, GA 30333, USA; fax: 404-639-8604; e-mail: kcastro@cdc.gov

Antique poster from the collection of Joel and Jahan Montague. Joel Montague is Board Chairman of Partners for Development, a nonprofit organization. Jahan Montague is associate professor of nephrology, University of Massachusetts Medical School in Worcester.

National Tuberculosis Genotyping and Surveillance Network: Design and Methods

Jack T. Crawford,* Christopher R. Braden,* Barbara A. Schable,* and Ida M. Onorato*

The National Tuberculosis Genotyping and Surveillance Network was established in 1996 to perform a 5-year, prospective study of the usefulness of genotyping *Mycobacterium tuberculosis* isolates to tuberculosis control programs. Seven sentinel sites identified all new cases of tuberculosis, collected information on patients and contacts, and obtained patient isolates. Seven genotyping laboratories performed DNA fingerprinting analysis by the international standard IS*6110* method. BioImage Whole Band Analyzer software was used to analyze patterns, and distinct patterns were assigned unique designations. Isolates with six or fewer bands on IS*6110* patterns were also spoligotyped. Patient data and genotyping designations were entered in a relational database and merged with selected variables from the national surveillance database. In two related databases, we compiled the results of routine contact investigations and the results of investigations of the relationships of patients who had isolates with matching genotypes. We describe the methods used in the study.

Jack T. Crawford, Guest Editor

Dr. Crawford is chief of the Immunology and Molecular Pathogenesis Section, Tuberculosis/Mycobacteriology Branch, Division of AIDS, STD, and TB Laboratory Research, Centers for Disease Control and Prevention. His research interests include application of molecular methods to epidemiology and diagnostics of mycobacterial diseases.

Molecular characterization of strains of *Mycobacterium tuberculosis* has been used for more than a decade to study the epidemiology of tuberculosis (TB). DNA fingerprinting, with IS*6110* as a probe, has been used successfully to trace transmission of *M. tuberculosis* in outbreaks, confirm laboratory cross-contamination, and identify risk factors for disease among populations of patients with TB (1). Before 1993, epidemiologic studies that used molecular characterization of *M. tuberculosis* were focused on populations, such as persons who were HIV positive or were in nosocomial settings (2–4). The standardization of methods provided an opportunity to examine *M. tuberculosis* strains from TB patients throughout the world and encompassed a variety of settings and populations (5).

In response to the upsurge of TB cases in the United States, the Centers for Disease Control and Prevention (CDC) funded regional laboratories to provide fingerprinting services to support TB control programs in outbreak investigations and to conduct studies on using DNA fingerprinting in TB epidemiology and control. This network was expanded to include sentinel surveillance sampling in 1996, when CDC established the National Tuberculosis Genotyping and Surveillance Network as a 5-year project. The genotyping network involved seven sentinel surveillance sites[1] that were paired with seven genotyping laboratories.[2] The historical background and objectives of the genotyping network have been discussed elsewhere (6). In this paper, we summarize the specific project methods.

Network Participants

The sentinel surveillance sites included the states of Arkansas, Maryland, Massachusetts, Michigan, and New Jersey; six counties in California surrounding San Francisco (Alameda, Contra Costa, Marin, San Mateo, Santa Clara, and Solano); and four counties in Texas representing two regions of the state with distinct demographics (Dallas, Tarrant, Cam-

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[1]Sentinel surveillance sites and personnel: 1) Arkansas Department of Health: Joseph Bates, Donald Cunningham, Kashef Ijaz, Bill Starrett, William Stead, and Frank Wilson; 2) California Department of Health Services: Rocio Agraz, Dan Chin, Jennifer Flood, Peter Oh, Sarah Royce, and Sumi Sun; 3) Maryland Department of Health and Mental Hygiene: Wendy Cronin, Jonathan Golub, David Hooper, Nancy Hooper, Monica Lathan, Leonard Mukasa, and Jafar Razeq; 4) Massachusetts Department of Public Health: Edward Corkren, Paul Elvin, Sue Etkind, Ann Miller, Edward Nardell, Jill Northrup, Alissa Scharf, Sharon Sharnprapai, Alexander Sloutsky, Robert Suruki, and Debra Thimas; 5) Michigan Department of Community Health: Bernie Benecke, Matthew Boulton, Norm Keon, Michael Kucab, Dennis Minnice, Jolynn Montgomery, and Charles Williams II; 6) New Jersey Department of Health and Senior Services: Nancy Buono, Mary Dillon, Zary Liu, Marcia Localio, Christian Nwigwe, Felicidad Santos, Mary Ellen Schulman, and Kenneth Shilkret; 7) Texas Department of Health: Dennis Ashworth, Dale Dingley Jr., Miguel Garza, Jeff Taylor, and Patricia Thickstun.

[2]Genotyping laboratories and personnel: 1) Alabama Department of Public Health and University of Alabama, Birmingham: William Benjamin Jr., Nancy Dunlap, Donna Hafner, Nancy Keenan, Kerry Lok, Donna Mulcahy, Virginia Pruitt, and Nancy Robinson; 2) Central Arkansas Veterans Healthcare Services: M. Donald Cave, Kathleen Eisenach, Ndingsa Fomukong, and ZhenhuaYang; 3) California Department of Health Services: Jason Coloma, Jonna De Leon, Edward Desmond, Athena Fye, Marguerite Griffith, Melvin Javonillo, Travis Jobe, Rose Longoria, Cynthia Sanders, and Lourdes Seli; 4) Michigan Department of Community Health: Dale Berry, Steve Church, Stephen Dietrich, Frances Pouch Downes, Jeffrey Massey, Teresa Miller, Laura Mosher, and Patricia Somsel; 5) New York State Health Department,Wadsworth Center: Jeff Driscoll, Robert Jovall, Mike McGarry, Max Salfinger, and Harry Taber; 6) Public Health Research Institute, New York: Pablo Bifani, Barry Kreiswirth, Natalia Kurepina, Barun Mathema, Soraya Moghazeh, Alex Ravikovitch, and Elena Shashkina; and 7) Texas Center for Infectious Diseases: Yvonne Camarce, Rebecca Cox, J. Seb Gillette, Oscar Gonzalez, Kenneth Jost, D. Mitch Magee, Teresa Quitugua, Ishmael I. Rosas, and Jeffery Taylor.

eron, and Hidalgo) (Figure). State departments of health TB-control programs conducted the project activities for each site. The seven laboratories provided DNA fingerprint analyses of *M. tuberculosis* isolates for their sentinel surveillance sites. In addition, genotyping network laboratories were assigned regions for which they provided genotype analysis of *M. tuberculosis* isolates in support of investigations by state departments of health. These isolates and associated patients were not included in the network databases. Regions were delineated on the basis of the approximate number of patients with TB in the regions; regions did not follow the boundaries of the U.S. Public Health Service Regions.

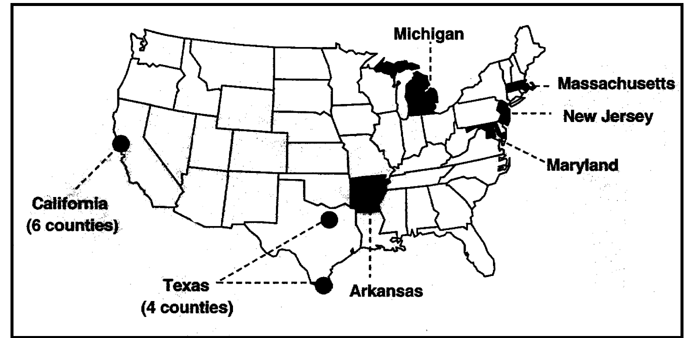

Figure. Map showing locations of the sentinel surveillance sites in the National Tuberculosis Genotyping and Surveillance Network, United States.

Case Finding and Isolate Collection

Sources of information to identify TB cases within surveillance sites included hospital and clinic records from all facilities serving patients in the area, records of laboratories performing mycobacteriology services, hospital ICD-9CM discharge codes for TB, pharmacy records for prescriptions of anti-TB drugs, medical examiners' or coroners' records, death certificates, and AIDS surveillance reports that listed a diagnosis of TB. Culture-positive, verified TB cases reported from January 1996 through December 2000 for national TB surveillance to CDC (reported on the Report of a Verified Case of Tuberculosis) were included as sentinel surveillance cases. Patients with recurrent TB occurring >1 year after completion of therapy are considered new case-patients; a small number of such cases were included. Patients later identified as residing outside of the surveillance area and those with positive cultures as the result of laboratory cross-contamination were excluded.

Every effort was made to acquire an isolate from each study patient. This task was difficult because many clinical samples were sent for culturing to laboratories other than the state public health laboratory, including hospital and commercial laboratories, some out of state. Isolates identified as *M. tuberculosis* complex, i.e., *M. tuberculosis*, *M. bovis,* and *M. africanum,* were included in the study. Isolates were sent to the assigned genotyping laboratory for typing. Stocks of all isolates were maintained at –70°C, and all isolates were transferred to CDC for long-term storage and future studies.

Laboratory Protocols

All National Tuberculosis Genotyping and Surveillance Network laboratories used standardized protocols. Isolates were typed by the international standard IS*6110* method, i.e., digestion of DNA with endonuclease *Pvu*II and hybridization with a probe containing the right end of IS*6110* (5). Growth was harvested directly from Lowenstein-Jensen slants, or isolates were subcultured in Middlebrook 7H9 broth. Cell suspensions were heat-killed before DNA was extracted. DNA was digested with *Pvu*II and run on a 20-cm, 1% agarose gel without ethidium bromide. Electrophoresis was conducted at 100 V for a short time to allow the DNA to enter the gel and then overnight at lower voltage until the dye front had run approximately 17 cm. An external size standard was run in wells 1, 10, and 20 on a 20-well gel. The gel was stained with ethidium bromide and photographed on a UV transilluminator to check for quality, including complete digestion of the DNA samples, degradation of DNA, and uneven running or smiling of the gel. Gels were blotted and hybridized in the standard fashion with chemiluminescent probes (ECL kit, Amersham Biosciences, Piscataway, NJ) and x-ray film, with the exception of one laboratory that used ^{32}P-labeled probes and a phosphoimager to generate direct digital images.

Two different size standards were used. Initially, total genomic DNA from *M. tuberculosis* strain 2650, provided by the genotyping laboratory at the Public Health Research Institute, New York, was used. This standard was replaced by a recombinant size standard prepared at CDC; this standard contains 25 fragments of known sequences, ranging in size from 700 to 15,000 bp. All fragments contain at least one copy of a 500-bp segment of the right side of IS*6110* and are detected by the IS*6110* probe. When the recombinant standard was available, the sizes of the fragments in the original 2650 standard were determined relative to the new standard, and these sizes were used to reanalyze all of the prior images containing the 2650 standard.

Computer-Assisted Pattern Analysis

IS*6110* patterns were scanned and then analyzed by regional laboratory personnel by using BioImage Whole Band Analyzer software, version 3.4 (BioImage, Ann Arbor, MI), a UNIX-based program run on Sun workstations (Sun Microsystems, Inc., Santa Clara, CA). This version of the software included enhancements developed specifically for the TB genotyping network. The software performed automatic band identification, but each image was visually evaluated and edited, as needed. The greatest difficulty was deciding if bands of greater than average width or intensity represented single or multiple fragments. The protocol called for assigning multiple bands only when separate bands were clearly evident, which often required observation of multiple exposures of the original films. Bands with lower than average intensity were counted only if the peak height in the scanned image was at least two thirds that of adjacent bands. The logarithmic scale method was used to size bands.

For pattern matching, BioImage software used the Jaccard coefficient of similarity between two patterns, A and B (100 x number of matched bands [number of bands in A + number of bands in B – number of matched bands]). Pairs of patterns were compared for matching bands by a deviation of ±2.5% for molecular weight (not distance migrated). All matches were verified by visual comparisons. Only patterns that were identical, i.e., had the same number of bands of the same sizes (±2.5%), were given the same designation.

Each regional laboratory maintained a database of fingerprint patterns for all isolates from the laboratory's sentinel site. New patterns were submitted to CDC for assignment of a genotyping network database designation. Any subsequent isolates with the identical fingerprint pattern were assigned the same designation and were not submitted to CDC. Files of single gel lanes were transmitted to a secure site at CDC. These files contained the original scanned image and the data file of calculated band sizes. In these images, bands included in the analysis were highlighted, and the locations of the bands in the standard were indicated. The database manager examined the images to determine if they were of acceptable quality but did not make changes in the submitted files. Apparent discrepancies were referred to the submitting laboratory for resolution, and the final decision about band identification rested with the laboratory that performed the analysis and had the original films. Submitted patterns were compared to the patterns in the existing TB genotyping network database by using the parameters described above. New patterns were added to the database and assigned consecutive 5-digit designations. When a submitted pattern matched a pattern already in the database, the submitted pattern was reported back with the existing designation. Thus, the TB genotyping network fingerprint database contained single examples of all distinct patterns from all seven regional laboratories during the 5-year study period. The designation 00000 was assigned to isolates having zero copies of IS*6110*, i.e., gave no bands. Some patterns were dropped from the study, e.g., if the isolate was the result of cross-contamination or the patient was from outside the study area. These designations were not reused; therefore, the total number of patterns (6,128) in the final database is lower than the highest designation (pattern 07193).

Secondary Typing

During the early part of the study, polymorphic GC-rich sequence (PGRS) typing, also referred to as pTBN12 (7), was used for secondary typing. However, this method was difficult to standardize, and spoligotyping (spacer oligotyping) was used for secondary typing (8). Isolates with six or fewer bands on IS*6110* typing, including most isolates that had been typed previously by using PGRS, were spoligotyped by using a standard method (9). The spoligotypes were recorded by using a 15-digit octal code to represent the binary result for the 43 spacers (10). Briefly, the 43-digit binary result, representing the 43 spacers (where 1=positive, 0=negative), was divided into 14 sets of 3 digits (spacers 1–42) plus 1 additional digit (spacer 43). Each 3-digit set was converted to octal code (000=0, 001=1, 010=2, 011=3, 100=4, 101=5, 110=6, and 111=7) with the final digit remaining either 1 or 0. This scheme yielded a 15-digit octal designation. To simplify database entries, the CDC database manager assigned consecutive arbitrary 4-digit designations to the octal designations.

Quality Assurance

Several sets of isolates were prepared at CDC and distributed to the typing laboratories for fingerprinting and analysis. Their results and results of typing at CDC were analyzed to determine reproducibility. Because such challenge sets may receive special handling, a second approach was used in the final 3 years of the project. For each laboratory, the database manager selected IS*6110* patterns at random from those previously submitted by that laboratory to the CDC database. The corresponding isolates, 10 per year per laboratory, were sent to CDC for typing, and results were compared to prior patterns.

Epidemiologic Databases

Sentinel surveillance sites each maintained a database of their patients and patient isolate genotype designations in Epi-Info 6.04d software (case file) (11). Patients were identified by their state surveillance case numbers. Complete files from each sentinel site were transmitted to CDC on a bimonthly basis and concatenated. These data were merged with select variables from the national TB surveillance database (SURVS-TB and TIMS) in an SAS dataset (SAS Institute, Inc., Cary, NC).

For each sentinel surveillance patient, data were collected concerning any TB case-patients identified as epidemiologically related through routine contact investigations conducted as part of TB-control programs. Epidemiologic connections were defined as direct exposure to an infectious TB patient (e.g., named contacts) or a circumstance in which patients were in the same location at the same time (e.g., the same jail). Related cases identified though routine contact investigations were included in the study if they had been diagnosed on or after January 1, 1990. Data collected for each related case-patient included the following: direction of transmission (source, secondary); type of relationship (household, non-household relatives or friends, co-worker, common source); type of setting of exposure (e.g., correctional facility, school or day-care center, co-worker, emergency shelter, group quarters, hospital, nursing home, other long-term care facility); and date that the epidemiologic relationship was discovered. Information for each related case identified through routine contact investigation was entered into a supplemental Epi-Info file, which was also routinely submitted to CDC. This database was merged with the case file data in an SAS dataset.

Genotype Cluster Investigation

Case-patients with isolates demonstrating indistinguishable genotypes, i.e., the same IS*6110*-pattern designation for isolates with more than six IS*6110* bands or the same IS*6110*

pattern and spoligotype for isolates with six or fewer IS*6110* bands, were defined as genotype clusters. All cases in clusters required review of epidemiologic information obtained from contact investigations before isolate genotyping to determine known epidemiologic connections. If one or more patients in a cluster had no known epidemiologic connections with others in the cluster, a cluster investigation was initiated. Patients participated with informed consent. Cluster investigations were conducted prospectively from January 1, 1998, to December 31, 2000. Data from medical record reviews and interviews with the patients were collected by using standardized forms. Information collected from interviews included demographics, TB medical history, any TB exposures, and at least a 2-year history for residence and social, work, and recreational activities. Information collected from medical records and interviews was compared for patients in each cluster to identify epidemiologic connections. The results of cluster investigations were entered into a third Epi-Info file. This file also contained related positive skin-test results, date of initial skin tests, direction of transmission, types of relationships, exposure settings, start dates of likely exposure, and information used to identify relationships (record review, routine interview, contact investigation, and isolate genotyping information). Data were submitted to CDC and converted into an SAS dataset for future analyses.

Weaknesses in Study Design

From the beginning of the project, we recognized that compiling results from multiple laboratories would be challenging because of the variability in the IS*6110* fingerprinting method. However, assigning a designation to the IS*6110* pattern for each isolate was necessary to allow entry of the result into the epidemiologic database. Two flaws in the design of the national fingerprint database were identified during the project; both primarily affected the analysis of isolates with low-copy numbers for IS*6110*. The first flaw arose from the computer algorithm for pattern matching. A newly submitted pattern (B) was considered distinct from an existing pattern (A) if any band in the pattern differed by more than the allowed 2.5% variation in size. However, a third pattern with a band of intermediate size can match both A and B. Thus, the software could identify a set of patterns that matched at 100%, even though some individual pairs within the set did not match. The second flaw was procedural. When a laboratory submitted a pattern that matched one already in the database, the existing designation was reported back to that laboratory. In such situations, the submitted image became the reference in that laboratory's database for that pattern designation. In retrospect, we realize that more consistent results would have been obtained if we had transmitted the prototype image from the national database to the regional laboratories so that everyone would have been using the same patterns for matching, i.e., the same set of band sizes.

Quality assurance results (12) highlighted the need for experienced judgment in evaluating and editing patterns and matching results. In addition, isolates with few copies of IS*6110* and common spoligotypes may not be well discriminated with these genotyping methods. The subjective nature of genotype interpretation and lack of specificity for some isolates may have resulted in some patients being misclassified as clustered or not.

Determination of epidemiologic connections based on routine contact investigations was problematic. Methods used by different jurisdictions varied considerably, as did the completeness of contact investigations. Therefore, interpretation and comparison of the proportion of cases with epidemiologic connections established by contact investigations among the sentinel sites should be done with caution. Genotype cluster investigations proved particularly difficult in several situations. Though cluster investigations were conducted prospectively, some genotype results were delayed for prolonged periods. Patients who had completed therapy were often difficult to locate for interviews. Delay in identifying clusters also occurred when a prolonged period had elapsed between diagnoses and cultures of patient specimens; delays also varied among the sites. The success of identifying epidemiologic links among clustered patients in these circumstances was limited.

Acknowledgments

We thank Charles Woodley and Lauren Cowan for their invaluable laboratory support and Ken Castro, Harold Jaffe, and Thomas Shinnick for their scientific guidance.

References

1. Small PM, van Embden JDA. Molecular epidemiology of tuberculosis. In: Bloom B, editor. Tuberculosis: pathogenesis, protection, and control. Washington: American Society for Microbiology; 1994. p. 569–82.
2. Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams EM, Sordillo KR, et al. An outbreak of multi-drug resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
3. Pearson ML, Jereb JA, Freiden TR, Crawford JT, Davis BJ, Dooley SW, et al. Nosocomial transmission of multi-drug resistant *Mycobacterium tuberculosis*: a risk to patients and healthcare workers. Ann Intern Med 1992;117:191–6.
4. Beck-Sague C, Dooley SW, Hutton MD, Otten J, Breeden A, Crawford JT, et al. Hospital outbreak of multi-drug resistant *Mycobacterium tuberculosis* infections: factors in transmission to staff and HIV-infected patients. JAMA 1992;268:1280–6.
5. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
6. Castro KG, Jaffe HW. Rationale and methods for the National Tuberculosis Genotyping and Surveillance network. Emerg Infect Dis 2002; 8;1188–91.
7. Ross BC, Raios K, Jackson K, Dwyer B. Molecular cloning of a highly repeated DNA element from *M. tuberculosis* and its use as an epidemiological tool. J Clin Microbiol 1992;30:942–6.
8. Kamerbeek J, Schouls L, Kolk A, van Agterveld D, van Soolingen D, Kuijpers S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.

9. Molhuizen HO, Bunschoten AE, Schouls LM, van Embden JDA. Rapid detection and simultaneous strain differentiation of *Mycobacterium tuberculosis* complex bacteria by spoligotyping. In: Parish T, Stoker NG, editors. Mycobacterial protocols. Totowa (NJ): Humana Press; 1998. p. 381–94.
10. Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, et al. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standard nomenclature. Int J Tuberc Lung Dis 2001;5:216–9.
11. Dean AG, Dean JA, Coulombier D, Burton AH, Brendel KA, Smith DC, et al. Epi-Info, Version 6.04a, a word processing, database, and statistics program for public health on IBM-compatible microcomputers. Atlanta: Centers for Disease Control and Prevention; 1996.
12. Braden CR, Crawford JT, Schable BA. Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network. Emerg Infect Dis 2002;8;1210–5.

All material published in Emerging Infectious Diseases is in the public domain and may be used and reprinted without special permission; proper citation, however, is appreciated.

Address for correspondence: Jack T. Crawford, Centers for Disease Control and Prevention, Mailstop F08, 1600 Clifton Rd., N.E., Atlanta, GA 30333, USA; fax: 404-639-1287; e-mail: jcrawford@cdc.gov

EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends

Detailed Instructions for Emerging Infectious Diseases Authors

Types of Articles

Perspectives. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author. Articles in this section should provide insightful analysis and commentary about new and reemerging infectious diseases and related issues. Perspectives may also address factors known to influence the emergence of diseases, including microbial adaptation and change, human demographics and behavior, technology and industry, economic development and land use, international travel and commerce, and the breakdown of public health measures. If detailed methods are included, a separate section on experimental procedures should immediately follow the body of the text.

Synopses. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author—both authors if only two. This section comprises concise reviews of infectious diseases or closely related topics. Preference is given to reviews of new and emerging diseases; however, timely updates of other diseases or topics are also welcome. If detailed methods are included, a separate section on experimental procedures should immediately follow the body of the text.

Research Studies. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author—both authors if only two. Report laboratory and epidemiologic results within a public health perspective. Although these reports may be written in the style of traditional research articles, they should explain the value of the research in public health terms and place the findings in a larger perspective (i.e., "Here is what we found, and here is what the findings mean").

Policy and Historical Reviews. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch. Articles in this section include public health policy or historical reports that are based on research and analysis of emerging disease issues.

Dispatches. Articles should be 1,000–1,500 words and need not be divided into sections. If subheadings are used, they should be general, e.g., "The Study" and "Conclusions." Provide a brief abstract (50 words); references (not to exceed 15); figures or illustrations (not to exceed two); and a brief biographical sketch of first author—both authors if only two. Dispatches are updates on infectious disease trends and research. The articles include descriptions of new methods for detecting, characterizing, or subtyping new or reemerging pathogens. Developments in antimicrobial drugs, vaccines, or infectious disease prevention or elimination programs are appropriate. Case reports are also welcome.

Commentary. Thoughtful discussions (500–1,000 words) of current topics. Commentaries may contain references but should not include figures or tables.

Another Dimension. Thoughtful essays, short stories, or poems on philosophical issues related to science, medical practice, and human health. Topics may include science and the human condition, the unanticipated side of epidemic investigations, or how people perceive and cope with infection and illness. This section is intended to evoke compassion for human suffering and to expand the science reader's literary scope. Manuscripts are selected for publication as much for their content (the experiences they describe) as for their literary merit.

Letters. This section includes letters that present preliminary data or comment on published articles. Letters (500–1,000 words) should not be divided into sections, nor should they contain figures or tables. References (not more than 10) may be included.

Book Reviews. Short reviews (250–500 words) of recently published books on emerging disease issues are welcome. The name of the book, publisher, and number of pages should be included.

News and Notes. We welcome brief announcements (50–150 words) of timely events of interest to our readers. (Announcements may be posted on the journal Web page only, depending on the event date.) In this sec-tion, we also include summaries (500–1,000 words) of emerging infectious disease conferences. Summaries may provide references to a full report of conference activities and should focus on the meeting's content.

Molecular Epidemiology of Tuberculosis in a Sentinel Surveillance Population

Barbara A. Ellis,* Jack T. Crawford,* Christopher R. Braden,* Scott J. N. McNabb,* Marisa Moore,* Steve Kammerer,* and the National Tuberculosis Genotyping and Surveillance Network Work Group[1]

We conducted a population-based study to assess demographic and risk-factor correlates for the most frequently occurring *Mycobacterium tuberculosis* genotypes from tuberculosis (TB) patients. The study included all incident, culture-positive TB patients from seven sentinel surveillance sites in the United States from 1996 to 2000. *M. tuberculosis* isolates were genotyped by IS*6110*-based restriction fragment length polymorphism and spoligotyping. Genotyping was available for 90% of 11,923 TB patients. Overall, 48% of cases had isolates that matched those from another patient, including 64% of U.S.-born and 35% of foreign-born patients. By logistic regression analysis, risk factors for clustering of genotypes were being male, U.S.-born, black, homeless, and infected with HIV; having pulmonary disease with cavitations on chest radiograph and a sputum smear with acid-fast bacilli; and excessive drug or alcohol use. Molecular characterization of TB isolates permitted risk correlates for clusters and specific genotypes to be described and provided information regarding cluster dynamics over time.

Barbara A. Ellis, Guest Editor

Dr. Ellis is a senior microbiologist with the National Center for Infectious Diseases, Centers for Disease Control and Prevention. Her research interests focus on the molecular epidemiology of infectious diseases, rodent-borne zoonotic diseases, and bioterrorism preparedness. Her work has included disease ecology studies of rodent-borne hemorrhagic fever viruses, molecular characterization of novel bartonellae, and molecular epidemiologic studies of *Mycobacterium tuberculosis*.

Since 1990, characterization of *Mycobacterium tuberculosis* isolates by molecular methods has been useful in confirming suspected laboratory contamination and as an adjunct to epidemiology-based contact investigation (1–3). Most studies used the restriction fragment length polymorphism (RFLP) technique, based on IS*6110* and specific to the *M. tuberculosis* complex. This genetic element may be present in different positions on the chromosome, resulting in a unique genotype useful for characterizing the strain of *M. tuberculosis* infecting a patient. Although RFLP has disadvantages (e.g., cost, time required to culture the organism, and specialized training and laboratory equipment), IS*6110*-based RFLP is the established method considered most discriminatory for genetic characterization of *M. tuberculosis* strains worldwide (4).

In 1996, the Centers for Disease Control and Prevention (CDC) established seven sentinel surveillance sites in the United States (National Tuberculosis Genotyping and Surveillance Network) to assess the utility of molecular genotyping for improving tuberculosis (TB) prevention and control. The TB genotyping network used standardized protocols for molecular characterization of *M. tuberculosis* isolates from patients in all sentinel sites. The network was designed to address specific epidemiologic questions regarding the natural history, transmission, and potential applicability of molecular genotyping of *M. tuberculosis* strains to augment TB control activities (5). Two objectives were to identify and determine the prevalence of specific *M. tuberculosis* genotype clustering in populations of sentinel surveillance TB patients and to describe the demographic characteristics of these populations and the genotypic characteristics of *M. tuberculosis* strains in clustered and nonclustered TB cases. We describe demographic and risk factor correlates for the most frequently occurring *M. tuberculosis* genotypes in isolates collected from sentinel TB patients.

Methods

This population-based sentinel study included all incident culture-positive TB patients from sentinel sites from January 1996 to December 2000. In brief, the seven sentinel surveillance sites included the states of Arkansas, Maryland, Massachusetts, Michigan, and New Jersey; Dallas, Tarrant, Cameron, and Hidalgo Counties in Texas; and Alameda, Contra Costa, Marin, San Mateo, Santa Clara, and Solano Counties in California. A detailed description of the study's design, participants, population, and laboratory and epidemiologic methods is provided elsewhere (6).

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[1]Members of the National Tuberculosis Genotyping and surveillance Network Work Group, in addition to the listed authors, included Joseph Bates, William Benjamin, Pablo Bifani, M. Donald Cave, Rebecca Cox, Wendy Cronin, Ed Desmond, Jeffrey Driscoll, Nancy Dunlap, Jennifer Flood, Kashef Ijaz,, Michael Kucab, Barry Kreiswirth, Zary Liu, D. Mitchell Magee, Jeffrey Massey, Ann Miller, Donna Mulcahy, Robert Pratt, Teresa Quitugua, Barbara Schable, Kenneth Shilkret, Harry Taber, Jeffrey Taylor, Sharon Sharnprapai, Sumi Sun, Zhenhua Yang

All patients included in the study were reported to the CDC national TB case registry on the form Report of a Verified Case of Tuberculosis, a standardized electronic form submitted for TB surveillance to CDC by all state public health reporting areas. Data reported include patient demographics, laboratory test results, drug susceptibilities, information on chest radiographs, and treatment outcomes (7).

Investigators from the sentinel surveillance sites submitted patient isolates to the corresponding regional laboratory for genotyping and conducted routine contact investigations. In addition, participants from the surveillance sites performed detailed epidemiologic investigations on groups of persons with *M. tuberculosis* isolates that had matching genetic patterns or clusters (see below). The regional genotyping laboratories conducted IS*6110* RFLP on isolates from sentinel patients. Since low-copy numbers of IS*6110* (i.e., six or fewer copies) reduce test specificity, spacer oligonucleotide typing (spoligotyping) was conducted on such isolates. A cluster, which was identified by analysis of the entire TB genotyping network database, was defined as two or more isolates with either identical RFLP patterns (at least seven copies of IS*6110*) or identical RFLP and spoligotype patterns for isolates with RFLP patterns that had six or fewer copies of IS*6110*.

Differences in the proportion of TB patients from the TB genotyping network population living in cities with populations of <100,000, 100,001 to 250,000, 250,001 to 500,000, and >500,000 were compared with those of the national TB patients for the year 2000 only. Statistics were obtained from the U.S. Census Bureau (available at: URL: http://www.census.gov/population/cen2000/phc-t6/tab04.pdf).

Correlation of average TB incidence among cases at the seven sentinel sites and percentage of cases with isolates that clustered genetically were examined by year by using the Spearman rank correlation statistic. Clustering was determined by examining each year's cases independently. A Mantel-Haenszel chi-square or Fisher exact test was used, as appropriate, to ascertain whether the sentinel population was representative of TB patients in the United States in terms of demographic, clinical, behavioral, or outcome characteristics.

We used multiple logistic regression to assess the importance of demographic, clinical, behavioral, or outcome variables in predicting the occurrence of a given genotype for those genetic clusters that occurred most frequently (≥20 isolates). The dependent variable was the presence or absence of a given genotype. The best-fit logistic regression model was determined by the strategy of Hosmer and Lemeshow (8). In brief, a univariate analysis of the categorical independent variables was done by using the Mantel-Haenszel chi-square or Fisher exact test, as appropriate; any variable with a significance value of ≥0.20 was included in a best subset, multivariate logistic regression model. Collinearity of independent variables was assessed by using the variance/covariance matrix from PROC LOGISTIC (SAS Institute, Inc., Cary, NC) to generate condition indices and a matrix of variance decomposition proportions to detect dependencies among the variables (9). Backward elimination of independent variables was performed if the probability of the independent variable was ≥0.20. Both the Wald statistic and 95% confidence interval were used on each coefficient to assess the significance of variables in each model; the log-likelihood ratio was used to assess the overall significance of the final models, and the Hosmer-Lemeshow statistic was used to evaluate the fit of each of the final models. Data were analyzed by SAS version 8.0 software (SAS Institute, Inc.) (10).

Results

Sentinel Population Characteristics

The incidence of TB cases in the sentinel surveillance sites varied within and among sites over time (Table 1). From 1996 to 2000, the overall incidence of TB in the United States declined from 8.0 to 5.8 per 100,000 inhabitants, and similar downward trends were observed in each of the TB genotyping network sites. The California, New Jersey, Arkansas, and Texas sites had a higher incidence of TB than the overall national rates. The incidence rates in California and Texas (sites that included only six and four counties from each state) were similar to the overall incidence rates for each state (data not shown).

In the surveillance area, 15,035 patients with verified TB represented 16% of the TB patients in the United States during the 5-year study period (Table 2). Overall, 11,923 TB patients were culture-positive (721 from Arkansas, 2,842 from California, 1,192 from Maryland, 1,022 from Massachusetts, 1,481 from Michigan, 2,599 from New Jersey, and 2,066 from Texas). Of TB patients in the surveillance areas, 79.3% (11,923) were culture positive, and RFLP results were available for 91.2% (10,883). However, spoligotyping results were not available for 131 of the isolates that had six or fewer copies of IS6110 (5%; n=2,638); thus, these patients were excluded from our analysis. Of 1,171 isolates not genotyped by RFLP or spoligotyping, 12 (1%) were from Michigan, 35 (3%) from Maryland, 40 (3%) from Massachusetts, 110 (9%)

Table 1. Incidence of tuberculosis cases in the United States and in the sentinel surveillance areas of the National Tuberculosis Genotyping Surveillance Network, 1996–2000[a]

Sentinel surveillance site	1996	1997	1998	1999	2000	Mean
Arkansas	9.0	7.9	6.7	7.1	7.4	7.6
California[b]	16.3	13.9	13.9	12.9	11.6	13.7
Maryland	6.3	6.7	6.3	5.7	5.3	6.1
Massachusetts	4.3	4.4	4.6	4.4	4.5	4.4
Michigan	4.6	3.8	3.9	3.6	2.9	3.8
New Jersey	10.3	8.9	7.9	7.0	6.7	8.2
Texas[b]	12.7	12.8	12.5	10.9	9.6	11.7
United States	8.0	7.4	6.8	6.4	5.8	6.9

[a]Number per 100,000 inhabitants.
[b]Sentinel surveillance areas for California and Texas did not include the entire states.

from Arkansas, 156 (13%) from Texas, 327 (28%) from California, and 491 (42%) from New Jersey. Primary reasons for lack of genotyping results included inability to obtain cultures from private health-care providers, contamination of cultures, or poorly growing or nonviable cultures.

Characteristics of the TB patient population from the genotyping network sentinel sites were comparable with those from the entire United States, with some exceptions (Table 2). Sentinel surveillance populations had higher proportions of women (42% for the genotyping network vs. 37% for the United States overall) and patients in the 15- to 44-year age category, and were more often homeless or lived in correctional or long-term care facilities. Higher proportions of genotyping network patients used intravenous drugs, but fewer patients used noninjecting drugs or alcohol excessively.

Of the study population, about 4% reported previous episodes of TB (652 of 15,035; Table 2). Of persons with a previous recent history of TB, 28 had TB after completing >1 year of therapy within the study period; genotyping data on isolates from both episodes were available for 22 of these persons. A higher number of persons from the TB genotyping network study population lived within city limits (97% vs. 87%). However, when compared with national averages, genotyping network populations were generally from smaller towns and cities: 1,446 (69%) of 2,099 genotyping network patients were from cities and towns with <250,000 inhabitants, compared with 10,093 (62%) of 16,377 TB patients nationwide (Mantel-Haenszel chi square= 41.8; p<0.0001).

The proportion of foreign-born patients was higher in genotyping network populations compared with the overall national average (50% for genotyping network vs. 41% for the United States). Numbers of foreign-born TB patients increased over time at about the same rate for both genotyping network populations and national TB patients. From 1996 to 2000, national proportions of foreign-born TB patients increased from 37% (7,725/21,045) to 47% (7,593/16,281); in the genotyping network populations, the proportions of foreign-born TB patients increased from 44% (1,153/2,642) to 58% (1,222/2,092). Characteristics of the genotyping network population between sites were similar, as were culture-positive genotyping network populations compared with the overall genotyping network case population.

Analysis of Genotyping Data

The distribution and diversity of RFLP and spoligotyping pattern results from the genotyping network have been discussed in detail (11). In contrast to that analysis, we used both RFLP and spoligotyping results to define genetic clusters. Overall, 6,609 distinct patterns were identified, including 1,029 that contained ≥2 isolates per cluster. When analyzed by site, 1,018 clusters were identified: 71 clusters were from Arkansas (611 cases genotyped, 2–16 cases per cluster), 233 from California (2,511 cases, 2–128 cases per cluster), 104 from Maryland (1,157 cases, 2–36 cases per cluster), 85 from Massachusetts (982 cases, 2–16 cases per cluster), 125 from Michigan (1,469 cases, 2–102 cases per cluster), 196 from New Jersey (2,112 cases, 2–40 cases per cluster), and 204 from Texas (1,910 cases, 2–96 cases per cluster). Overall, 970 distinct genotypes, including 235 representing clusters, had ≤6 copies (2,507 cases, 24% clustered, 2–93 cases per cluster). In contrast, 794 clusters from 5,639 distinct genotypes had ≥ 7 IS6110 copies (8,245 cases, 14% clustered, 2–105 cases per cluster). Most clusters included seven or fewer persons (85%; 900/1,029).

Longitudinal Analysis

Most clusters occurred in only a single site (66%; 680/1,029). However, 260 (25%) were found in two sites, 55 (5%) in three sites, 19 (2%) in four, 8 (1%) in five, and 7 (1%) in six sites. As expected, clusters that spanned multiple sites were larger. Clusters found at a single site averaged four persons per cluster (mean=3.65; standard error [SE] ± 0.22; n=680), in contrast to 61 persons per cluster for the genotypes found at six sites (mean=61.14; SE ± 23.6; n=7; Kruskal-Wallis test, p<0.0001). Most (62%) of the 34 clusters that occurred in at least four sites occurred in all 5 years of the study; 26% in 4 years; and 6% each in 3 and 2 years of the study.

Changes in proportions of patients with isolates that clustered were observed over time. In the first 2 years of the study, the percentage of the cumulative total number of cases that clustered increased from 28% to 45%; smaller increases occurred thereafter (Figure 1). Overall, the proportion of clustered cases was 48% (5,171/10,752). The percentages of clustered cases by sites were 28% (276/982) for Massachusetts; 34% (393/1,157) for Maryland; 41% (873/2,112) for New Jersey; 42% (1,046/2,511) for California; 44% (266/611) for Arkansas; 49% (720/1,469) for Michigan; and 57% (1,093/1,910) for Texas. Maximum cluster size and absolute numbers of cases with isolates that clustered continued to increase through the end of the study.

Overall, cases with isolates that clustered showed a concomitant decline with average incidence of TB over the 5-year period (Figure 2). A significant positive association was observed between the percentage of cases with clustered genotypes and TB incidence over time (Spearman rho=0.90; p=0.037).

Risk Factor Analyses of Genetic Clusters

Compared with persons whose isolates had unique genotypes, persons with isolates that clustered were more likely to be non-Hispanic, black men born in the United States. They were more likely to have pulmonary disease and abnormal chest radiographs with cavities; in addition, they more often had positive sputum smears; were HIV-positive, homeless, or residents of a correctional facility; and used drugs or alcohol excessively (Table 3). Patients with unclustered isolates were 5 years older on the average than those with isolates that clustered (44.8 years vs. 49.4 years, respectively; Table 3). Multiple logistic regression efforts resulted in models that were not robust (data not shown).

Table 2. Demographic and risk behavior factors and clinical, laboratory, and treatment outcomes for the sentinel surveillance patients (National Tuberculosis Genotyping and Surveillance Network) compared with factors and outcomes of all tuberculosis patients, United States, 1996–2000[a,b]

Variable	Category	All U.S. TB cases (n=93,097) (%)	All NTGSN cases (n=15,035) (%)	Probability[c]
Gender	Male	58,356 (62.7)	8,767 (58.3)	<0.001
	Female	34,734 (37.3)	6,266 (41.7)	
	Unknown	7 (0.0)	2 (0.0)	
Age (yrs)	≤4	3,289 (3.5)	518 (3.4)	NS
	5–14	2,397 (2.6)	393 (2.6)	NS
	15–24	7,988 (8.6)	1,462 (9.7)	<0.001
	25–44	32,433 (34.8)	5,413 (36.0)	0.005
	45–64	25,319 (27.2)	3,850 (25.6)	<0.001
	>64	21,662 (23.3)	3,397 (22.6)	NS
	Unknown	9 (0.0)	2 (0.0)	
Race/ethnicity	White, non-Hispanic	22,655 (24.3)	3,087 (20.5)	<0.001
	Black, non-Hispanic	30,201 (32.4)	4,775 (31.8)	NS
	Hispanic	20,475 (22.0)	2,923 (19.4)	<0.001
	American Indian/Native	1,280 (1.4)	38 (0.3)	<0.001
	Asian/Pacific Islander	18,346 (19.7)	4,195 (27.9)	<0.001
	Unknown	140 (0.2)	17 (0.1)	
Place of birth	U.S.-born	54,341 (58.4)	7,530 (50.1)	<0.001
	Foreign-born	38,252 (41.1)	7,468 (49.7)	
	Unknown	504 (0.5)	37 (0.2)	
Years in United States (foreign-born only)	<1	7,425 (19.4)	1,494 (20.0)	NS
	1	2,612 (6.8)	567 (7.6)	NS
	2	2,073 (5.4)	477 (6.4)	<0.005
	3	1,827 (4.8)	406 (5.4)	<0.05
	4	1,676 (4.4)	361 (4.8)	NS
	≥5	19,396 (50.7)	3,688 (49.4)	<0.001
	Unknown	3,243 (8.5)	475 (6.4)	
Country of origin[d]	Philippines	4,862 (12.7)	1,113 (14.9)	<0.0001
	Mexico	8,795 (23.0)	1,100 (14.7)	<0.0001
	Vietnam	3,824 (10.0)	968 (13.0)	<0.0001
	India	2,527 (6.6)	883 (11.8)	<0.0001
	China	1,930 (5.0)	370 (5.0)	NS
	Haiti	1,470 (3.8)	225 (3.0)	<0.0005
	Peru	636 (1.7)	207 (2.8)	<0.0001
	Republic of Korea	1,176 (3.1)	202 (2.7)	NS
	Ethiopia	578 (1.5)	153 (2.0)	<0.001
	Ecuador	627 (1.6)	115 (1.5)	NS
	Other	11,827 (30.9)	2,132 (28.5)	<0.0001
Status at diagnosis	Alive	90,141 (96.8)	14,611 (97.2)	0.02
	Dead	2,925 (3.1)	422 (2.8)	
	Unknown	31 (0.0)	2 (0.0)	
Site of disease	Pulmonary	68,611 (73.7)	10,576 (70.3)	<0.001

Table 2. (continued) Demographic and risk behavior factors and clinical, laboratory, and treatment outcomes for the sentinel surveillance patients (National Tuberculosis Genotyping and Surveillance Network) compared with factors and outcomes of all tuberculosis patients, United States, 1996–2000[a,b]

Variable	Category	All U.S. TB cases (n=93,097) (%)	All NTGSN cases (n=15,035) (%)	Probability[c]
Site of disease	Extrapulmonary	17,406 (18.7)	3,210 (21.4)	<0.001
	Pulmonary and Extrapulmonary	7,046 (7.6)	1,241 (8.3)	0.003
	Unknown	34 (0.0)	8 (0.1)	
Primary disease site	Pulmonary	73,157 (78.6)	11,365 (75.6)	<0.0001
	Lymph: cervical	4,312 (4.6)	1,020 (6.8)	<0.0001
	Pleural	3,842 (4.1)	674 (4.5)	<0.05
Primary disease site	Miliary	1,407 (1.5)	241 (1.6)	NS
	All other	10,345 (11.1)	1,727 (11.5)	NS
	Unknown	34 (0.0)	8 (0.0)	
Sputum smear for acid-fast organisms	Negative	36,912 (39.6)	5,995 (39.9)	<0.0001
	Positive	33,235 (35.7)	4,735 (31.5)	
	Not done/unknown	22,950 (24.6)	4,305 (28.7)	
TST at diagnosis	Negative	13,215 (14.2)	1,947 (12.9)	<0.001
	Positive	54,113 (58.1)	8,799 (58.5)	
	Not done/unknown	25,769 (27.6)	4,289 (28.6)	
Case verification criteria	Positive culture	74,940 (80.5)	11,967 (79.6)	<0.01
	Positive smear	765 (0.8)	136 (0.9)	NS
	Clinical case	11,286 (12.1)	1,858 (12.4)	NS
	Provider diagnosis	6,106 (6.6)	1,074 (7.1)	<0.01
Chest radiograph[e]	Cavitary	18,742 (24.8)	2,990 (25.3)	NS
	Noncavitary	50,652 (66.9)	7,897 (66.8)	NS
	Normal	2,495 (3.3)	360 (3.0)	NS
	Not done/unknown	3,802 (5.0)	578 (4.9)	
	Total	75,691	11,825	
HIV status[f]	Positive	6,062 (18.8)	884 (16.7)	NS
	Negative	16,525 (51.2)	2,406 (45.5)	
	Indeterminate	47 (0.1)	6 (0.1)	
	Refused	1,959 (6.1)	325 (6.1)	
	Not offered	4,130 (12.8)	899 (17.0)	
	Test done, unknown	714 (2.2)	115 (2.2)	
	Unknown	2,812 (8.7)	658 (12.4)	
	Total	32,249	5,293	
Homeless within past year	Yes	5,789 (6.2)	646 (4.3)	<0.001
	No	84,873 (91.2)	14,185 (94.3)	
	Unknown	2,435 (2.6)	204 (1.4)	
Resident of correctional facility at diagnosis	Yes	3,352 (3.6)	377 (2.5)	<0.001
	No	89,479 (96.1)	14,617 (97.2)	
	Unknown	266 (0.3)	41 (0.3)	
Correctional facility type	Federal prison	164 (4.9)	6 (1.6)	<0.005
	State prison	1,036 (30.9)	97 (25.7)	<0.05

Table 2. (continued) Demographic and risk behavior factors and clinical, laboratory, and treatment outcomes for the sentinel surveillance patients (National Tuberculosis Genotyping and Surveillance Network) compared with factors and outcomes of all tuberculosis patients, United States, 1996–2000[a,b]

Variable	Category	All U.S. TB cases (n=93,097) (%)	All NTGSN cases (n=15,035) (%)	Probability[c]
Correctional facility type	Total	3,352	377	
	Local jail	1,905 (56.8)	231 (61.3)	NS
	Juvenile facility	33 (1.0)	8 (2.1)	NS
	Other	161 (4.8)	34 (9.0)	<0.001
	Unknown	53 (1.6)	1 (0.3)	
	Yes	3,157 (3.4)	441 (2.9)	0.004
Resident, long-term care facility at diagnosis	Unknown	284 (0.3)	42 (0.3)	
Long-term care facility type	Nursing home	1,794 (56.8)	279 (63.3)	<0.01
	Hospital-based	441 (14.0)	66 (15.0)	NS
	Residential	356 (11.3)	34 (7.7)	<0.05
	All other	504 (16.0)	55 (12.5)	NS
	Unknown	62 (2.0)	7 (1.6)	
	Total	3,157	441	
Injecting drug use[g]	Yes	2,569 (2.8)	515 (3.4)	<0.001
	No	83,141 (89.3)	13,771 (91.6)	
	Unknown	7,387 (7.9)	749 (5.0)	
Noninjecting drug use[g]	Yes	6,557 (7.0)	811 (5.4)	<0.001
	No	78,622 (84.5)	13,367 (88.9)	
	Unknown	7,918 (8.5)	857 (5.7)	
Excessive alcohol use[h]	Yes	13,646 (14.7)	1,661 (11.0)	<0.001
	No	71,924 (77.3)	12,552 (83.5)	
	Unknown	7,527 (8.1)	822 (5.5)	
Drug resistance[i]				
First-line drugs	Yes	8,456 (11.7)	1,482 (12.6)	<0.001
	No	57,029 (79.0)	8,886 (75.5)	
	Not tested/unknown	6,703 (9.3)	1,399 (11.9)	
	Total	72,188	11,767	
Second-line drugs	Yes	1,341 (1.9)	208 (1.8)	<0.001
	No	175 (0.2)	78 (0.7)	
	Not tested/unknown	70,672 (97.9)	11,481 (97.6)	
	Total	72,188	11,767	
DOT	Yes—total DOT	40,511 (43.5)	4,936 (32.8)	<0.001
	Yes—both DOT and self-administered	20,555 (22.1)	3,648 (24.3)	<0.001
	No	23,337 (25.1)	5,326 (35.4)	<0.001
	Unknown	8,694 (9.3)	1,125 (7.5)	
Within city limits	Yes	80,775 (86.8)	14,603 (97.1)	<0.001
	No	10,916 (11.7)	374 (2.5)	
	Unknown	1,406 (1.5)	58 (0.4)	
Previous diagnosis of TB	Yes	4,794 (5.1)	652 (4.3)	<0.001
	No	87,567 (94.1)	14,336 (95.4)	

Table 2. (continued) Demographic and risk behavior factors and clinical, laboratory, and treatment outcomes for the sentinel surveillance patients (National Tuberculosis Genotyping and Surveillance Network) compared with factors and outcomes of all tuberculosis patients, United States, 1996–2000[a,b]

Variable	Category	All U.S. TB cases (n=93,097) (%)	All NTGSN cases (n=15,035) (%)	Probability[c]
Previous diagnosis of TB	Unknown	736 (0.8)	47 (0.3	
Duration of therapy (days)	Mean	246	245	NS
	Median	217	214	
	Std. dev.	135	130	
	No.	65,344	10,822	

[a]NTGSN, National Tuberculosis Genotyping Surveillance Network;TB, tuberculosis; DOT, directly observed therapy; TST, tuberculin skin test; Std. dev., standard deviation; NS, not significant (p>0.05).
[b]Subtotals for each category are listed if different from the total case numbers.
[c]Probability of significant differences between U.S. TB patients and all NTGSN surveillance patients (chi-square test; t-test for duration of therapy); referent group is all other groups combined, excluding not done or unknown categories, unless otherwise noted.
[d]Top 10 countries for foreign-born patients only.
[e]Excludes cases with extrapulmonary TB only.
[f]HIV cases from California are excluded because this site does not report HIV results on Report of a Verified Case of Tuberculosis forms; ages 15–44 years only.
[g]Injecting or noninjecting drug use within last year; includes use of licensed, prescription, or illegal drugs (not prescribed by a physician).
[h]Excessive use of alcohol within the past year as indicated by participation in alcohol treatment programs, diagnosis of alcoholism, or observation of intoxication during visits to health-care facilities.
[i]Drug resistance on initial testing of isolate. First-line drug resistance is resistance to at least one of the following: isoniazid, rifampin, ethambutol, or streptomycin. Second-line drug resistance is resistance to one or more of the following: ethionamide, kanamycin, cycloserine, capreomycin, para-amino salicylic acid, amikacin, rifabutin, ciprofloxacin, ofloxacin, or other drugs. Testing results for one or more of the drugs could have been missing.

Except for 4 genotypes, all 34 clusters with ≥ 20 isolates per cluster had significant demographic, clinical, and behavioral risk factors (Table 4). Race, ethnicity, and place of birth were frequently significant predictors for a given genotype. Other predictors included gender, age, site of disease, resistance to first-line drugs, and alcohol or drug abuse (Table 4). Twelve (40%) of 30 of these larger clusters were observed in four or more sites over a 5-year period. Lower percentages of foreign-born patients than U.S.-born patients clustered, regardless of the number of IS*6110* copies (Figure 3). More than 50% (1,025/1,825) of the foreign-born patients whose isolates clustered had been in the United States for ≥5 years. Clustering of isolates from foreign-born patients ranged from 15% (49/316) in Michigan to 38% (309/816) in Texas.

Discussion

This population-based study is the largest that has been conducted in the United States to assess risk factors related to specific *M. tuberculosis* genotypes. Generally, clustered isolates have been considered recently acquired infections (12). However, this assumption may not always be correct. Clustering does not prove that transmission occurred, and its demonstration depends on adequate sampling of the population, incidence of TB, and characteristics of the study population (e.g., age structure, population mobility, duration of residence, and immune status) (1,13). Only 25%–42% of patients in genetic clusters were shown to have epidemiologic connections with another member of the cluster (14–16). Conventional epidemiologic investigation of these TB patients (including interviews) was conducted, but inclusion in this analysis was outside the scope of this article. Thus, results that indicate clustered genotypes are representative of recent transmission should be interpreted with caution.

Given this caveat, our results nevertheless demonstrate several consistent patterns. Differences in demographic and other risk factors for persons with isolates that clustered corroborated those from smaller studies conducted in the United States and larger surveys in Europe. Extensive surveys from

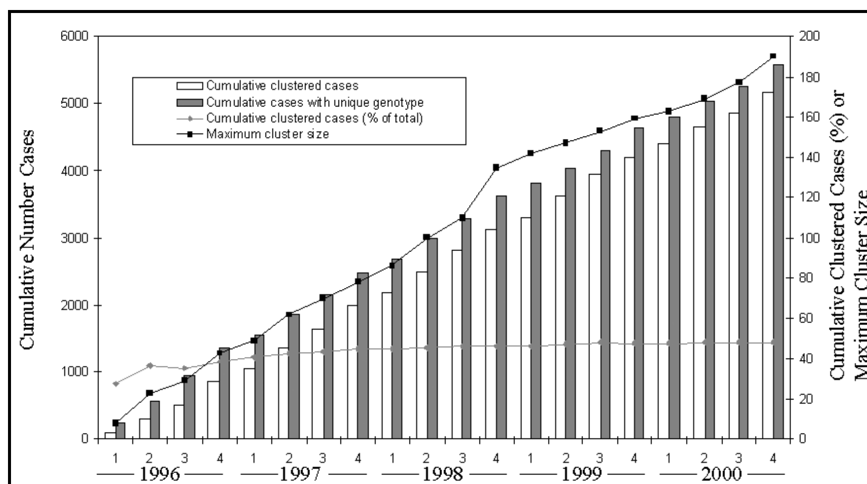

Figure 1. Numbers of tuberculosis cases, cumulative proportion of cases with isolates in genetic clusters, and maximum genetic cluster size from seven sentinel surveillance sites by quarter that verified case was counted, 1996–2000. Numbers of cases with isolates that had unique genotypes and those with isolates that were in genetic clusters are shown separately.

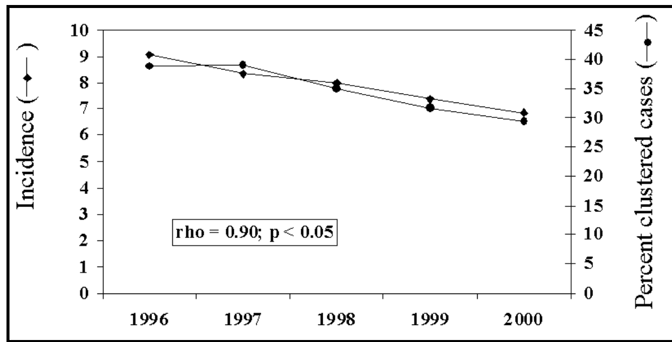

Figure 2. Average annual incidence of tuberculosis for seven sentinel surveillance sites and percentage of cases with isolates in genetic clusters, 1996 to 2000. Spearman correlation coefficient and probability of correlation between incidence and percentage of cases clustered are given.

the Netherlands (17) also demonstrated that persons with isolates that clustered genetically were younger than those with unique genotypes. Other risk factors for clustering included being male, born in the United States, non-Hispanic black, or homeless; using drugs and alcohol excessively; and having pulmonary disease and cavitations on chest radiograph, a sputum smear with acid-fast bacilli, and HIV infection. These risk factors have been observed for TB patients in different communities (12,18,19). The heterogeneity and diversity of the study population may account for our failure to produce a multivariate logistic model to predict clustering.

A third of the foreign-born cases were recent immigrants to the United States, and overall, the percentage of clustered isolates from foreign-born persons was lower than the percentage from nonimmigrants (Figure 3), indicating that at least a portion of these cases resulted from reactivation of latent disease or recent infection in the country of origin. In addition, for foreign-born persons, clustering of *M. tuberculosis* increased with the duration of residence in the United States. These results suggest that recently imported strains of *M. tuberculosis* from foreign-born persons may not commonly spread to U.S. residents or that transmission may be occurring after a lag time before the imported strains manifest as disease in contacts. Similar observations have been published in studies from San Francisco, New York, Switzerland, and Norway (20–24). These data may also reflect gaps in our knowledge of *M. tuberculosis* genotypes in circulation; a comparison of the U.S. TB genotyping network results with other databases worldwide may be warranted.

Logistic regression analysis of the most commonly occurring strains demonstrated that different risk factors were associated with specific genotypes. Several genotypes were associated with ethnic origin (e.g., Asian or Pacific Islander and Hispanic patients with six and three genotypes, respectively; Table 4). A recent study in Norway showed that several clusters consisted of patients of the same ethnic origin (23). An association has also been observed between the patient's ethnic origin and IS6110 copy number (25). These results, in conjunction with additional epidemiologic data, may be useful in tracking the geographic origin and spread of *M. tuberculosis* strains of public health importance (26).

A small proportion of clustered isolates were from persons from more than four sites spanning 5 years of study (Table 4).

Table 3. Comparison of demographic and behavioral risk factors and clinical and treatment outcomes of tuberculosis (TB) case-patients who have genetically clustered genotypes with factors and outcomes of patients who had unique genotype patterns[a]

Variable[b]		Clustered (%)	Unclustered (%)	Relative risk (95% CI)	Probability[c]
Total cases (n=10,752)		5.171(48.1)	5,581(51.9)		
Gender	Male	3,289 (63.6)	3,107 (55.7)	1.19 (1.14% to 1.24%)	<0.001
	Female	1,881 (36.4)	2,473 (44.3)		
	Unknown	1 (0.0)	1 (0.0)		
Mean age (yrs; ±S.E.)		44.8 (±0.26)	49.4 (±0.28)		<0.0001
Race/ethnicity	White, non-Hispanic	1,018 (19.7)	1,201 (21.5)	0.94 (0.90% to 0.99%)	0.02
	Black, non-Hispanic	2,254 (43.6)	1,237 (22.2)	1.61 (1.55% to 1.67%)	<0.001
	Hispanic	914 (17.7)	1,112 (19.9)	0.92 (0.88% to 0.97%)	0.003
	American Indian/Native	17 (0.3)	10 (0.2)		
	Asian/Pacific Islander	961 (18.6)	2,014 (36.1)	0.60 (0.56% to 0.63%)	<0.001
	Unknown	7 (0.1)	7 (0.1)		
Place of birth	U.S.-born	3,331 (64.4)	2,023 (36.2)	1.83 (1.75% to 1.90%)	<0.001
	Foreign-born	1,825 (35.3)	3,552 (63.6)		
	Unknown	15 (0.3)	6 (0.1)		
Recent arrival in United States[d]	Yes	535 (29.3)	1,225 (34.5)	0.59 (0.55% to 0.63%)	<0.001
	No	1,181 (64.7)	2,111 (59.4)		
	Unknown	109 (6.0)	216 (6.1)		
Site of disease	Pulmonary	3,902 (75.5)	3,835 (68.7)	1.20 (1.14% to1.26%)	<0.001
	Extrapulmonary	788 (15.2)	1,254 (22.5)	0.77 (0.72% to 0.81%)	<0.001

Table 3. (continued) Comparison of demographic and behavioral risk factors and clinical and treatment outcomes of tuberculosis (TB) case-patients who have genetically clustered genotypes with factors and outcomes of patients who had unique genotype patterns[a]

Variable[b]		Clustered (%)	Unclustered (%)	Relative risk (95% CI)	Probability[c]
Site of disease	Pulmonary and extrapulmonary	476 (9.2)	492 (8.8)		NS
	Unknown	5 (0.1)	0		
Sputum smear	Positive	2,270 (43.9)	2,011 (36.0)	1.22 (1.11% to 1.33%)	<0.001
	Negative	1,802 (34.8)	1,943 (34.8)		
	Not done/unknown	1,099 (21.3)	1,627 (29.1)		
Chest radiograph[e]	Cavitary	1,345 (30.7)	1,172 (27.1)	1.09 (1.04% to 1.14%)	<0.001
	Noncavitary	2,639 (60.2)	2,826 (65.3)		
	Normal	146 (3.3)	118 (2.73)		
	Not done/unknown	253 (5.8)	211 (4.9)		
	Total	4,383	4,327		
HIV status[f]	Positive	458 (22.2)	223 (11.8)	1.37 (1.29% to 1.46%)	<0.001
	Negative	978 (47.4)	847 (44.8)		NS
	Indeterminate	0	4 (0.2)		
	Refused	106 (5.1)	138 (7.3)		
	Not offered	252 (12.2)	354 (18.7)		
	Unknown	270 (13.0)	323 (17.1)		
	Total	2,064	1,889		
Homeless within past year	Yes	370 (7.2)	139 (2.5)	1.55 (1.46% to 1.64%)	<0.001
	No	4,724 (91.4)	5,370 (96.2)		
	Unknown	77 (1.5)	72 (1.3)		
Resident of correctional facility at diagnosis	Yes	190 (3.7)	69 (1.2)	1.55 (1.43% to 1.67%)	<0.001
	No	4,966 (96.0)	5,503 (98.6)		
	Unknown	15 (0.3)	9 (0.2)		
Injecting drug use[g]	Yes	312 (6.0)	72 (1.3)	1.73 (1.65% to 1.83%)	<0.001
	No	4,540 (87.8)	5,231 (93.7)		
	Unknown	319 (6.2)	278 (5.0)		
Noninjecting drug use[g]	Yes	460 (8.9)	140 (2.5)	1.65 (1.57% to 1.73%)	<0.001
	No	4,335 (83.8)	5,140 (92.1)		
	Unknown	376 (7.3)	301 (5.4)		
Excessive alcohol use[g]	Yes	948 (18.3)	371 (6.6)	1.61 (1.54% to 1.67%)	<0.001
	No	3,897 (75.4)	4,893 (87.7)		
	Unknown	326 (6.3)	317 (5.7)		
First-line drugs [h]	Yes	622 (12.1)	755 (13.7)	0.93 (0.87% to 0.99%)	0.016
	No	2,718 (53.0)	3,337 (60.5)		
	Not done	1,748 (34.1)	1,356 (24.6)		
	Unknown	45 (0.9	66 (1.2)		
	Total	5,133	5,514		

[a]CI, confidence interval; S.E., standard error.
[b]Only factors that had significant differences are shown.
[c]Probability of chi-square statistic is shown, except for t-test results from analysis of age from each group.
[d]Foreign-born only; arrived in the United States within 2 years.
[e]Excludes cases with extrapulmonary TB only.
[f]California TB cases not included; ages 15–44 years only.
[g]Excessive drug or alcohol use within last year.
[h]First-line drug resistance is resistance to at least one of the following: isoniazid, rifampin, ethambutol, or streptomycin. Second-line drug resistance is resistance to one or more of the following: ethionamide, kanamycin, cycloserine, capreomycin, para-amino salicylic acid, amikacin, rifabutin, ciprofloxacin, ofloxacin, or other drugs. Testing results for one or more of the drugs could have been missing.

Table 4. Odds ratios from best-fit logistic regression analyses of the presence or absence of a specific genetic cluster of *Mycobacterium tuberculosis* on demographic, clinical, behavioral, or treatment outcome variables [a]

Designation[c]	IS6110 copies	Spoligotype[c]	N	Main effect	Odds ratio estimates (95% CI)[b]	Wald p[b]
00003[c]	1	777777777760771	40	Asian/Pacific Islander	3.70 (1.51% to 9.02%)	0.004
				Age	0.98 (0.96% to 0.99%)	0.017
				Foreign-born	12.4 (3.83% to 39.9%)	<0.0001
00129[d]	1	777777777413771	25	Asian/Pacific Islander	73.3 (17.0% to 315.6%)	<0.0001
				Extrapulmonary infection	2.57 (1.10% to 6.03%)	0.03
00129[d]	1	777777774413771	83	Asian/Pacific Islander	282.8 (88.06% to 908.11%)	<0.0001
00129[d]	1	477777777413071	23	Asian/Pacific Islander	6.34 (1.52% to 26.44%)	0.01
				Foreign-born	10.4 (1.55% to 70.12%)	0.02
00129[d]	1	777777777413731	13	Asian/Pacific Islander	13.88 (3.71% to 51.92%)	<0.0001
				Resistance to first-line drugs[d]	3.80 (1.22% to 11.86%)	0.02
00129	1	777776407760601	40	Female	2.73 (1.43% to 5.23%)	0.0025
				Black, non-Hispanic	3.57 (1.47% to 8.68%)	0.005
				Injecting drug use	3.81 (1.81% to 8.03%)	0.0004
00016	2	701776777760601	129	Male	0.58 (0.40% to 0.84%)	0.004
				Black, non-Hispanic	10.88 (5.48% to 21.6%)	0.006
00016[c]	2	777776777760771	82	Hispanic	16.36 (10.15% to 26.37%)	<0.0001
00016	2	037776777760601	30	Age	1.03 (1.01% to 1.05%)	0.006
				Black, non-Hispanic	7.13 (2.36% to 21.53%)	0.0005
				Resident, long-term care facility	3.67 (1.17% to 11.70%)	0.026
00016[d]	2	777776777760601	175	U.S.-born	3.12 (1.85% to 5.26%)	<0.0001
				Excessive alcohol use	0.55 (0.37% to 0.83%)	0.0048
00370	3	700036777760731	13	White, non-Hispanic	5.20 (1.52% to 17.79%)	0.0087
				HIV positive	5.87 (1.69% to 20.41%)	0.005
				Noninjecting drug use	3.74 (1.17% to 12.01%)	0.03
00017[d]	4	700076777760771	25	Hispanic	4.97 (2.16% to 11.44%)	0.0002
00017[d]	4	777776777760771	64	Hispanic	15.7 (9.24% to 26.71%)	<0.0001
01285	4	777776777760771	20	Resident, correctional facility	8.23 (3.08% to 22.01%)	<0.0001
00015	7		28	Black, non-Hispanic	7.04 (1.64% to 30.3%)	0.0087
				Injecting drug use	4.84 (2.11% to 11.09%)	0.0002
				Excessive alcohol use	2.28 (1.02% to 5.13%)	0.05
00768	9		19	Black, non-Hispanic	11.68 (1.54% to 88.87%)	0.02
				Noninjecting drug use	2.77 (1.11% to 6.92%)	0.03
00242[d]	10		95	Male	2.12 (1.27% to 3.56%)	0.004
				Age	0.97 (0.96% to 0.98%)	<0.0001
				U.S.-born	8.44 (2.63% to 27.09%)	0.0003
				Homeless	3.60 (2.16% to 5.98%)	<0.0001
				Noninjecting drug use	0.46 (0.24% to 0.90%)	0.02
00028	11		70	Black, non-Hispanic	17.57 (5.50% to 56.12%)	<0.0001
00159	11		24	Excessive alcohol use	2.76 (1.23% to 6.22%)	0.01
00325	11		20	Age	1.03 (1.01% to 1.06%)	0.01
				Excessive alcohol use	3.08 (1.22% to 7.70%)	0.02
00673	11		25	Asian/Pacific Islander	84.6 (19.85 to 361.9%)	<0.0001
00757	11		16	Age	0.90 (0.85% to 0.94%)	<0.0001
				HIV positive	4.86 (1.60% to 14.79%)	0.005

Table 4. (continued) Odds ratios from best-fit logistic regression analyses of the presence or absence of a specific genetic cluster of *Mycobacterium tuberculosis* on demographic, clinical, behavioral, or treatment outcome variables [a]

Designation[c]	IS6110 copies	Spoligotype[c]	N	Main effect	Odds ratio estimates (95% CI)[b]	Wald p[b]
00019[c]	12		27	Male	3.68 (1.10% to 12.39%)	0.03
				White, non-Hispanic	5.4 (2.35% to 11.08%)	<0.0001
00372	12		20	Homeless	6.09 (2.43% to 15.20%)	0.0001
				Resident, long-term care facility	5.52 (1.535 to 20.0%)	0.009
00035	13		33	Black, non-Hispanic	6.96 (2.3% to 21.0%)	0.0006
				Resistance to second-line drugs[e]	40.59 (16.5% to 99.85%)	<0.0001
00867	14		20	Black, non-Hispanic	11.68 (1.54% to 88.87%)	0.02
				Noninjecting drug use	2.77 (1.11% to 6.92%)	0.03
01284	17		46	Black, non-Hispanic	2.40 (1.22% to 3.57%)	<0.0001
				Pulmonary disease	0.92 (-0.01% to 1.86%)	0.054
00237[c]	21		98	White, non-Hispanic	2.80 (1.81% to 4.33%)	<0.0001
				Excessive alcohol use	2.09 (1.36% to 3.22%)	0.0007
01693	21		29	HIV positive	3.16 (1.39% to 7.18%)	0.006
				Injecting drug use	3.08 (1.26% to 7.56%)	0.014
				Extrapulmonary disease	3.99 (1.69, 9.42)	0.002
00027	22		78	Black, non-Hispanic	1.74 (1.05% to 2.90%)	0.03
				Sputum-smear positive	3.07 (1.75% to 5.39%)	<0.0001

[a]CI, confidence interval.
[b]Only genetic clusters that had ≥20 isolates were included in the analysis; some samples sizes are <20 because of missing data among independent variables (Wald 95% confidence intervals given in parentheses). Only genetic clusters with significant predictors are listed. Age was modeled as a continuous variable.
[c]The National Tuberculosis Genotyping Surveillance Network (NTGSN) designation for the IS*6110* RFLP pattern is represented; spoligotype octal code designations are presented only for those genetic clusters from isolates that had ≤6 copies of IS*6110*. RFLP patterns and spoligotypes are detailed elsewhere (11).
[d]isolates observed in ≥ 4 sites over 5 years.
[e] First-line drug resistance is resistance to at least one of the following: isoniazid, rifampin, ethambutol, or streptomycin. Second-line drug resistance is resistance to one or more of the following: ethionamide, kanamycin, cycloserine, capreomycin, para-amino salicylic acid, amikacin, rifabutin, ciprofloxacin, ofloxacin, or other drugs.

Although an in-depth analysis of epidemiologic links was not possible in this study, we found no evidence of recent transmission between patients with identical genotypes from the different states (data not shown); this lack of transmission was also noted in a smaller study in the United States (27). Since TB transmission is generally considered a local event, these ubiquitous genotypes may be widespread because of social factors (e.g., homelessness or alcohol or drug abuse; Table 4). In addition, these genotypes may represent older, endemic domestic strains that have been in the United States for centuries and have dispersed more widely throughout the United States than the more recently imported strains. Further molecular characterization of these genotypes may show additional differences not detected by RFLP. Nonetheless, the effect of *M. tuberculosis* virulence or host factors on the distribution of these genotypes cannot be ascertained.

The proportion of strains that were classified into clusters of identical genotypes (48%) was comparable with proportions in the Netherlands and Denmark (50%) (2,28), but the proportion was considerably higher than in two other countries (17% in Switzerland [29]; 20% in Norway [23]). The cumulative percentage of clustered strains reached a plateau by the end of the study's second year (Figure 1), a finding consistent with other molecular epidemiologic TB studies (2). Increases in maximum cluster size were anticipated because, as sample sizes increase with time, the number of isolates in each cluster would be expected to increase. In addition, higher proportions of clustered cases were observed for low-band number patterns (Figure 3), which had the maximum cluster size and may

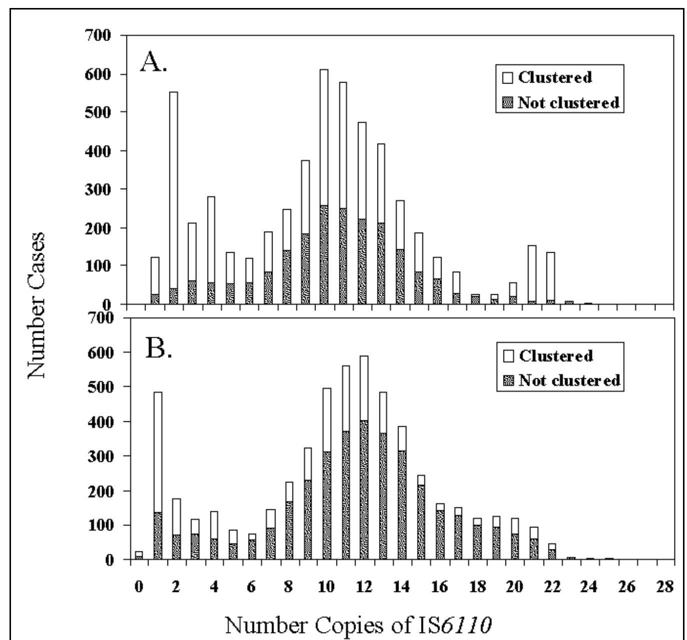

Figure 3. Number of cases with isolates that had unique genotypes ("not clustered") and those in genetic clusters for U.S.-born (A) and foreign-born persons (B) by number of copies of IS*6110*.

indicate that the low-copy IS*6110* patterns are not specific, even with the addition of spoligotyping.

The sensitivity and specificity of IS*6110* RFLP in molecular epidemiologic studies have not been quantified and represent a potential limitation of this study. Although the stability of IS*6110* is relatively high, the half-life of IS*6110* RFLP is estimated to be 3–10 years (29–31) based on typing of serial isolates from individual patients. A study of isolates from patients in confirmed chains of transmission showed little change in IS*6110* patterns (32). Calculation of these rates may be influenced by the duration between time of disease onset and time of sampling and may be proportional to the effectiveness of the TB control program (30). Because genotyping results were not available for 10% of TB cases in this study, estimates of the degree of clustering and the size of clusters are conservative. Some unique isolates might have clustered if some of the missing isolates had been available or if other cases with the same strain were present outside the study area (33).

Sentinel surveillance sites defined by artificial boundaries (i.e., state lines) not entirely representative of TB patients from the United States were included in this study. More than 90% of the isolates from patients from the surveillance areas were genotyped, and these isolates were representative of those culture-positive patients from the sentinel surveillance areas. However, 16% of all TB case-patients reported in the United States were included in these sentinel surveillance sites during the 5-year study period. In addition, the sentinel surveillance population had higher proportions of foreign-born persons than the national average. Because of the propensity of foreign-born persons to have isolates with unique genotypes, the actual rate of clustering may have been underestimated. Nonetheless, sentinel surveillance of TB cases has provided a useful method for documenting genotypes in circulation in the United States and for identifying risk factor correlates of common genotypes.

Annual declines in TB incidence were paralleled by similar declines in the proportion of cases with genotypes in clusters (Figure 2), a finding consistent with the hypothesis that decreased clustering is expected with declining incidence (20). Since effort was similar each year, this association is not likely to be an artifact related to sample size (i.e., as sample size or number of cases becomes smaller, the probability of detecting clusters decreases). These findings underscore the importance of long-term longitudinal molecular studies and the potential usefulness of these methods in evaluating program effectiveness and improving program management.

Acknowledgments

We thank Ida Onorato, Ken Castro, Tom Shinnick, and Thomas Navin for their scientific guidance and logistic support; Elsa Villarino and James Mills for valuable comments on an earlier version of the manuscript; Annie Faye Prescott for excellent editorial assistance; and the health officials at local and state TB control offices that supported the activities of the National Tuberculosis Genotyping and Surveillance Network.

References

1. Foxman B, Riley L. Molecular epidemiology: focus on infection. Am J Epidemiol 2001;153:1135–41.
2. van Soolingen D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med 2001;249:1–26.
3. van Embden JD, van Soolingen D, Small PM, Hermans PW. Genetic markers for the epidemiology of tuberculosis. Res Microbiol 1992;143:385–91.
4. Fletcher HA. Molecular epidemiology of tuberculosis: recent developments and applications. Curr Opin Pulm Med 2001;7:154–9.
5. Castro KG, Jaffe HW. Rationale and methods for the National Tuberculosis Genotyping and Surveillance network. Emerg Infect Dis 2002; 8;1188–91.
6. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192–6.
7. Centers for Disease Control and Prevention. Tuberculosis information management system (TIMS) user's guide. Surveillance appendix I. Atlanta: U.S. Department of Health and Human Services, Public Health Service; 1998. p. 68.
8. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley and Sons, Inc.; 1989.
9. Davis CE, Hyde JE, Bangdiwala SI, Nelson JJ. An example of dependencies among variables in a conditional logistic regression. In: Moolgavkar S, Prentice RL, editors. Modern statistical methods in chronic disease epidemiology. New York: John Wiley and Sons, Inc.; 1986. p. 140–7.
10. SAS/STAT User's Guide. Cary, NC: SAS Institute Inc.; 2000.
11. Cowan LS, Crawford JT. Genotype analysis of *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. Emerg Infect Dis 2002;8;1294-1302.
12. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
13. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 1999;130:971–8.
14. Lockman S, Sheppard JD, Braden CR, Mwasekaga MJ, Woodley CL, Kenyon TA, et al. Molecular and conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. J Clin Microbiol 2001;39:1042–7.
15. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, 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.
16. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996;77:407–13.
17. Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JD, Fine PE. The effect of age and study duration on the relationship between "clustering" of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. Epidemiol Infect 2001;126:43–62.
18. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–6.
19. Friedman CR, Quinn GC, Kreiswirth BN, Perlman DC, Salomon N, Schluger N, et al. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. J Infect Dis 1997;176:478–84.
20. Borgdorff MW, Behr MA, Nagelkerke NJ, Hopewell PC, Small PM. Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. Int J Tuberc Lung Dis 2000;4:287–94.

21. Geng E, Kreiswirth B, Driver C, Li J, Burzynski J, DellaLatta P, et al. Changes in the transmission of tuberculosis in New York City from 1990 to 1999. N Engl J Med 2002;346:1453–8.
22. Tornieporth NG, Ptachewich Y, Poltoratskaia N, Ravi BS, Katapadi M, Berger JJ, et al. Tuberculosis among foreign-born persons in New York City, 1992–1994: implications for tuberculosis control. Int J Tuberc Lung Dis 1997;1:528–35.
23. Dahle UR, Sandven P, Heldal E, Caugant DA. Molecular epidemiology of *Mycobacterium tuberculosis* in Norway. J Clin Microbiol 2001;39:1802–7.
24. Genewein A, Telenti A, Bernasconi C, Mordasini C, Weiss S, Maurer AM, et al. Molecular approach to identifying route of transmission of tuberculosis in the community. Lancet 1993;342:841–4.
25. Park YK, Bai GH, Kim SJ. Restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolated from countries in the western pacific region. J Clin Microbiol 2000;38:191–7.
26. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, et al. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA 1996;275:452–7.
27. Yang Z, Barnes PF, Chaves F, Eisenach KD, Weis SE, Bates JH, et al. Diversity of DNA fingerprints of *Mycobacterium tuberculosis* isolates in the United States. J Clin Microbiol 1998;36:1003–7.
28. Bauer J, Yang Z, Poulsen S, Andersen AB. Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of *M. tuberculosis* infection. J Clin Microbiol 1998;36:305–8.
29. Pfyffer GE, Strassle A, Rose N, Wirth R, Brandli O, Shang H. Transmission of tuberculosis in the metropolitan area of Zurich: a 3 year survey based on DNA fingerprinting. Eur Respir J 1998;11:804–8.
30. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J Infect Dis 1999;180:1238–44.
31. Warren RM, van der Spuy GD, Richardson M, Beyers N, Borgdorff MW, Behr MA, et al. Calculation of the stability of the IS*6110* banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. J Clin Microbiol 2002;40:1705–8.
32. Niemann S, Rusch-Gerdes S, Richter E, Thielen H, Heykes-Uden H, Diel R. Stability of IS*6110* restriction fragment length polymorphism patterns of *Mycobacterium tuberculosis* strains in actual chains of transmission. J Clin Microbiol 2000;38:2563–7.
33. Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol 2002;155:565–71.

Address for correspondence: Barbara Ellis, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Mailstop E79, Atlanta, GA 30333, USA; fax: 404-498-2270; e-mail: bae7@cdc.gov

EMERGING INFECTIOUS DISEASES

Full text free online at www.cdc.gov/eid

The print journal is available at **no charge** to public health professionals

☐ YES, I would like to receive Emerging Infectious Diseases.

Please print your name and business address in the box and return by fax to 404-371-5449 or mail to
EID Editor
CDC/NCID/MS D61
1600 Clifton Road, NE
Atlanta, GA 30333

Moving? Please give us your new address (in the box) and print the number of your old mailing label here______________

EID Online
www.cdc.gov/eid

Quality Assessment of *Mycobacterium tuberculosis* Genotyping in a Large Laboratory Network

Christopher R. Braden,* Jack T. Crawford,* and Barbara A. Schable*

Quality assessment exercises were conducted to evaluate the reproducibility of IS*6110* DNA fingerprinting performed by eight laboratories in the National Tuberculosis Genotyping and Surveillance Network. Three panels, each with 8 to 16 isolates, were typed at all laboratories, resulting in 280 images. When the pattern obtained by the majority for each isolate was used as the standard, exact matches were obtained for 73% of patterns; 90% and 97% of patterns matched within one- and two-band differences, respectively. A second approach involved retyping of randomly selected isolates at the Centers for Disease Control and Prevention. Retyping was done for 8–19 isolates per laboratory (76 total). Paired images matched exactly for 54% of isolates and within one and two band differences, 78% and 93%, respectively. We evaluated reasons for mismatching. We also evaluated the reproducibility of spoligotyping using a test panel of 13 isolates; a discrepancy of 1 in 91 results was noted.

A proposed standard methodology for *Mycobacterium tuberculosis* genotyping was published in 1993 (1). This methodology, restriction fragment length polymorphism (RFLP)–based analysis using IS*6110* as a marker, has been adopted by laboratories worldwide for studying the molecular epidemiology of tuberculosis. Although other methods have been introduced, IS*6110* fingerprinting provides the highest specificity and remains the most effective and consistent genotyping technique for *M. tuberculosis*. Standardization of this technique promises the best results for direct comparison of large numbers of genotype patterns obtained in different laboratories (2–4). Common *M. tuberculosis* genotype patterns from distant populations may be sought to determine geographic mobility of related strains or to identify clonal ancestry in evolutionary genetics. Alternatively, the ability to divide the genotyping workload among several laboratories may allow more complete genotyping for larger host populations.

In 1996, the National Tuberculosis Genotyping and Surveillance Network adopted the standard IS*6110* fingerprinting method for primary genotyping (5). To test the proficiency of laboratories and determine the reproducibility of this genotyping technique in a large network of laboratories, the Centers for Disease Control and Prevention (CDC) instituted quality assessment exercises for the seven genotyping network laboratories. These exercises included sending panels of isolates (from CDC stocks) to all laboratories and retyping (at CDC) a sample of isolates for which an IS*6110* RFLP pattern was previously submitted to the central database. Through these quality assessment exercises, we identified common causes for mismatched patterns and determined the frequency of mismatch occurrences among the laboratories.

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Methods

Genotyping

IS*6110* DNA fingerprinting was performed according to standard methods (1,5). The genotyping network protocol provided standardization of the procedure among laboratories, including use of the same size standards, gel sizes, electrophoresis run conditions, and IS*6110* probes. Size standards were applied in outside and middle lanes (three total) for each gel; standards internal to each lane were not used. Gel electrophoresis equipment and reagent components, such as agarose, were not specified in the protocol and varied among the laboratories. Spoligotyping was performed by the standard procedure (6).

Isolate Panel Quality Assessment Exercise

Three test panels were sent to each of seven genotyping network laboratories. CDC also genotyped the isolates in all panels, so a total of eight laboratories participated in the overall assessment. For two test panels, 16 and 13 isolates selected from CDC stocks were subcultured in 7H9 broth, then onto Lowenstein-Jensen slants, and sent to the seven genotyping network laboratories for genotyping. Occasionally, cultures became contaminated, or a technical mishap occurred; therefore, not every laboratory submitted an image for every isolate. For one panel, DNA was purified from eight isolates at CDC, and aliquots of DNA were sent to all laboratories. With few exceptions, each laboratory ran each panel of isolates on a single gel. The resulting images were digitized and analyzed by using BioImage Whole Band Analyzer software, version 3.4 (BioImage, Ann Arbor, MI) with standardized analysis parameters (5). The match parameter included a possible 2.5% deviation in calculated molecular weights among compared

bands, a parameter setting recommended by BioImage developers, experienced users in genotyping network, and other experts. Digital images of the edited autoradiograms indicating operator determination of band placement were transmitted electronically to the CDC network coordinator for comparison. We conducted these exercises during years 1, 2, and 5 of the 5-year project.

Isolate Retyping Quality Assessment Exercise

We selected recent genotypes submitted by the genotyping network laboratories to the central fingerprint database from a spreadsheet, basing selection on pattern band numbers to standardize the complexity of patterns distribution among laboratories. Isolates were routinely frozen and stored at –70°C at each participating laboratory. Selected isolates were recultured and sent to CDC, where they were retyped according to genotyping network protocol.

Spoligotyping Quality Assessment Exercise

The 13 isolates in the third panel were also spoligotyped at six genotyping network laboratories and CDC. Each laboratory analyzed the resulting patterns and provided a digital result for comparison.

Matching Outcome

For each isolate in the panels, we defined a reference pattern as the pattern that matched exactly by the greatest number of laboratories; in 62% of isolates, at least five of eight results matched exactly. Each isolate pattern was compared to the reference pattern. For isolates retyped at CDC, each isolate pattern from the genotyping network laboratory was compared with the CDC pattern. Neither pattern was assumed to be the correct result. Misalignment between a band in one pattern and the same band in a compared pattern was considered a one-band mismatch. We placed the outcomes in one of the following categories: an exact match (i.e., same number and size of bands), an exact match with the exception of one band (match ±1 band), an exact match with the exception of two bands (match ±2 bands), and no match (three or more bands different). We included only computer-derived comparisons of original, blinded pattern determinations; we did not include any judgments after the computer match in the analysis. Reasons for nonexact matches were categorized as the addition or omission of one or more bands in one pattern compared with the other, mismatch of individual bands in compared patterns, and a shift up or down in one pattern compared with the other.

Results

Isolate Panel Quality Assessment Exercise

Tests by eight laboratories of three panels (8–16 isolates each) resulted in 280 images from 37 isolates. Overall, an exact match was achieved for 73% of all patterns (range by isolate, 33% to 100%; range by panel, 66% to 85%); 90% matched ±1 band (range by isolate, 63% to 100%; range by panel, 85% to 98%); and 97% matched ±2 bands (range by isolate, 86% to 100%; range by panel, 96% to 98%) (Table 1).

No single laboratory achieved exact matches to the reference pattern for all isolates. One laboratory matched ±1 band for all isolates and three laboratories ±2 bands for all isolates. All laboratories matched the reference pattern ±2 bands for at least 90% of isolates (Table 2).

Patterns with a low number of bands (1–6 bands) constituted 20% of patterns; 55% of patterns contained a midrange number of bands (7–15), and 25% of patterns had a high number of bands (16–23). Figure 1 shows the matching results for the three categories of patterns. For low-band number patterns, 100% of images matched exactly. For midband number patterns, 74% matched exactly; and for high–band number patterns, 49% matched exactly. Within limits of ±2 bands, 95% of mid–band number images matched, and 86% of high–band number images matched.

Of the 76 images that did not match exactly the reference pattern for each strain, 41 (54%) showed addition or omission of one or more bands compared to others. Figure 2A shows the normalized, computer-generated lane maps of patterns obtained with one isolate. Although most bands in all of the patterns matched with very small deviations in size, the pattern in lane 2 is missing two bands, and the pattern in lane 7 has one additional band. Figure 2B shows the original IS*6110* RFLP autoradiogram image of the lane with the additional band and two representative images from other laboratories. The extra band is clearly present in the middle image from lane 7 and absent on the others, indicating a true difference in patterns derived from the same isolate. A specific class of discrepancies included the omission of high molecular weight (>10 kb) bands as shown in lane 2 of Figure 2A, which accounted for 10 (13%) of all mismatches. High molecular weight fragments also caused discrepancies in size determinations, especially when they were larger than the size standard (i.e., >15 kb). Figure 3 shows differences in band determination. Two laboratories (lanes 1 and 7) identified a doublet, which was called a single band in the other five laboratories.

Table 1. Percent of restriction fragment length polymorphism images matching reference pattern[a] for all isolates in quality assessment panels

	Panel 1 % (16 isolates, 124 images)	Panel 2 % (8 isolates, 53 images)	Panel 3 % (13 isolates, 103 images)	All panels % (37 isolates, 280 images)
Match	73	85	66	73
Match ±1[b]	91	98	85	90
Match ±2[b]	98	98	96	97

[a]Reference pattern was the pattern that matched exactly in the greatest number of laboratories.
[b]Match ±1, exact match with the exception of one band; match ±2, exact match with the exception of two bands.

Table 2. Number and percent of restriction fragment length polymorphism images from each laboratory matching reference pattern in quality assessment panels

Laboratory	No. of patterns submitted by laboratory	Match[a] (%)	Match ±1[b] (%)	Match ±2[c] (%)
1	36	27 (75)	33 (92)	36 (100)
2	39	29 (74)	35 (90)	38 (97)
3	35	27 (77)	35 (100)	35 (100)
4	38	33 (87)	37 (97)	38 (100)
5	38	33 (87)	35 (92)	36 (95)
6	34	22 (65)	28 (82)	33 (97)
7	27	13 (48)	25 (93)	26 (96)
8	33	20 (61)	25 (76)	30 (91)

[a] Number of patterns submitted by laboratory that matched exactly the reference pattern.
[b] Match ±1, exact match with the exception of one band.
[c] Match ±2, exact match with the exception of two bands.

This type of band misidentification accounted for 31 (76%) of the 41 patterns that did not match the reference patterns because of the omission of one or more bands.

The second most common reason for nonmatches was a misalignment (16%) of individual bands (i.e., images contained a band, but it did not fall within 2.5% of the molecular weight of the band in other images). In additional cases, mismatching was caused by shifts in the pattern (5%), as shown in Figure 3, lane 2. Multiple factors for nonmatching (12%) accounted for the remainder of mismatches. Of images that did not match the reference pattern, 28% matched exactly the image from at least one other laboratory.

Isolate Retyping Quality Assessment Exercise

Seventy-six distinct isolates with genotype images previously submitted to the central database were retyped. Of these, 8% of patterns had 3–6 bands, 67% of patterns had 7–15 bands, and 25% of patterns had 16–23 bands. Overall, 54% of retyped images matched exactly the original submitted image (range by laboratory, 25% to 80%); 77% of image pairs matched ±1 band (range by laboratory, 60% to 100%); and 93% matched ±2 bands (range by laboratory, 67% to 100%). Figure 4 shows the results stratified by pattern band number. Thirty-seven percent of high–band number pattern pairs matched exactly; the proportion improved to 89% with a match of ±2 bands. The presence or absence of bands in one pattern compared with the other accounted for 37% of nonmatching pattern pairs, multiple reasons accounted for 23%, individual band mismatches accounted for 20%, omission of a high molecular band accounted for 11%, and whole pattern shifts accounted for 9%.

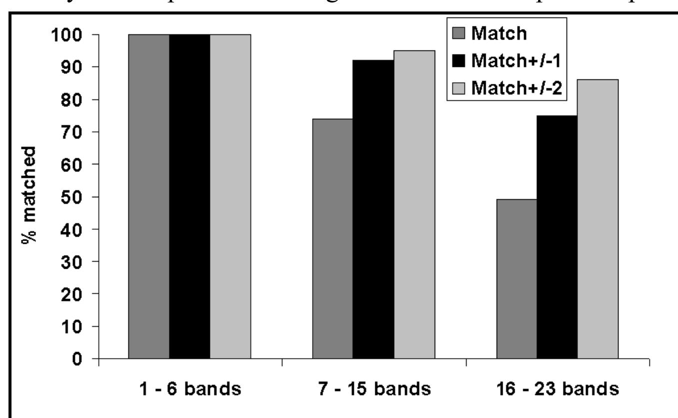

Figure 1. Quality assessment panel match results shown by number of bands in patterns.

Spoligotyping Results

Spoligotyping results were compared for 13 isolates typed at the seven laboratories. Subjective judgments regarding the hybridization patterns were made at the laboratories, and only the final digital results were compared. Identical results were obtained for 90 of the 91 spoligotypes. The one differing result occurred because three consecutive oligonucleotide spacers were absent.

Discussion

We used quality assessment exercises with the laboratories in the genotyping network to evaluate the reproducibility and interlaboratory variability of IS*6110* DNA fingerprinting. The genotyping network project included analyses of large databases of pattern images, with the use of computer-assisted matching algorithms to identify genotype clusters. This method is more complex and difficult than a visual interpretation of a small number of pattern images. The results of the quality assessment exercises suggest that exact computer-identified matches in images produced by the eight laboratories were reproducible for patterns with a small number of bands but not reliably reproducible for complex patterns with large numbers of bands.

Some differences observed were not the result of varying interpretation of the patterns. In Figure 2A, the discrepant band is clearly present in one sample and absent in the others. This discrepancy may have been caused by the presence of two subpopulations of bacteria that emerged upon subculture in the laboratories, a situation often recognized by the presence of faint bands in images (7). However, this discrepancy also occurred in the panel consisting of aliquots of DNA sam-

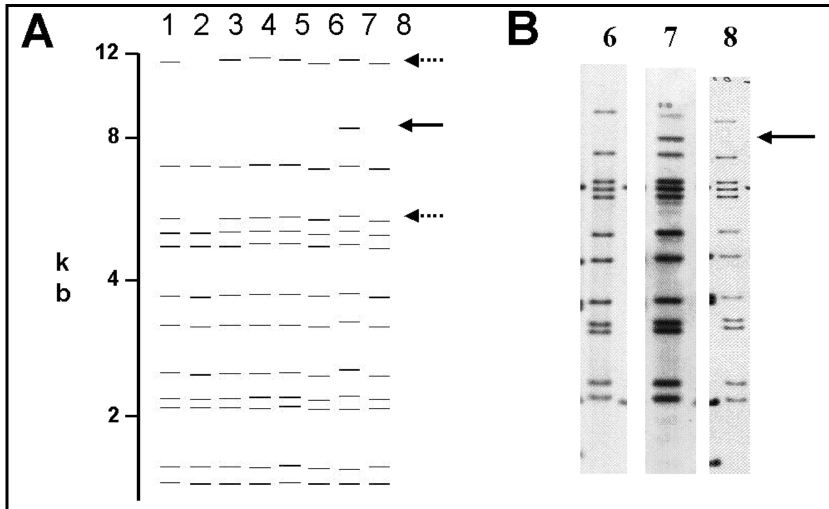

Figure 2. A) Computer-derived IS*6110* restriction fragment length polymorphism patterns from eight laboratories for one isolate. Addition (solid line) or omission (dotted line) of bands is demonstrated. B) Autoradiogram images demonstrating the addition of IS*6110* band in restriction fragment length polymorphism pattern in one subpopulation of *Mycobacterium tuberculosis* isolates used in the quality assessment exercise.

ples provided by CDC. A true difference in the populations of the isolate typed may account for the discrepant results of one laboratory in the spoligotyping exercise; the outlier pattern had three consecutive oligonucleotide spacers missing compared with the others, possibly because of a deletion in the direct repeat sequence of the genome of that *M. tuberculosis* sample. These observations suggest that IS*6110* replication and deletions in the direct repeat sequence may occur within subcultured populations of *M. tuberculosis* during a short period of time. However, none of these observed changes were independently verified by genomic mapping. Absence of high-molecular-weight bands was also common and was likely the result of poor transfer of DNA during the blotting procedure. Absence of high-molecular-weight bands can also result from degraded DNA samples.

Another cause for the differences in patterns relates to variability in band identification, specifically the determination of intense or wide bands as one, two, or even three fragments and discrimination among closely spaced bands (Figure 3). This determination often requires operator judgment and editing in pattern analysis and thus is prone to operator-dependent bias. The exposure times of autoradiograms and the intensity of bands may have influenced this determination and the outcome of the matching procedure. Table 2 shows this type of subjectivity for laboratory 7, where heavy or wide bands tended to be overcalled as multiple bands. The resulting rate of exact matches for laboratory 7 was 48%; the rate jumped to 93% for matches ±1 band.

Some mismatches resulted from small variations in sizing of bands. In many cases, patterns appeared the same on visual inspection, but specific bands fell outside of the 2.5% deviation in calculated molecular weight allowed for matching. Two phenomena are represented. Uneven heating during gel electrophoresis yields variations in mobility in different lanes, which results in miscalculation of the fragment sizes in comparison to the external standards that could be up to four lanes distant. This problem can be reduced by including internal lane standards that allow normalization of individual lanes. However, this approach requires alignment of the images from the two separate probes and can introduce other errors. An unanticipated problem occurred in the computerized matching process when images were slightly misaligned (Figure 3). With two closely spaced bands, the analysis algorithm occasionally matched the upper band in one image to the lower band in the other image, leaving the remaining bands unmatched. This situation could occur despite the fact that the correct band match was within the 2.5% deviation limit. This type of mismatch is readily detected by visual comparison of the patterns.

Advantages and disadvantages exist in either increasing or decreasing the allowable deviation in band size in the matching process. Decreasing the allowable deviation would exacer-

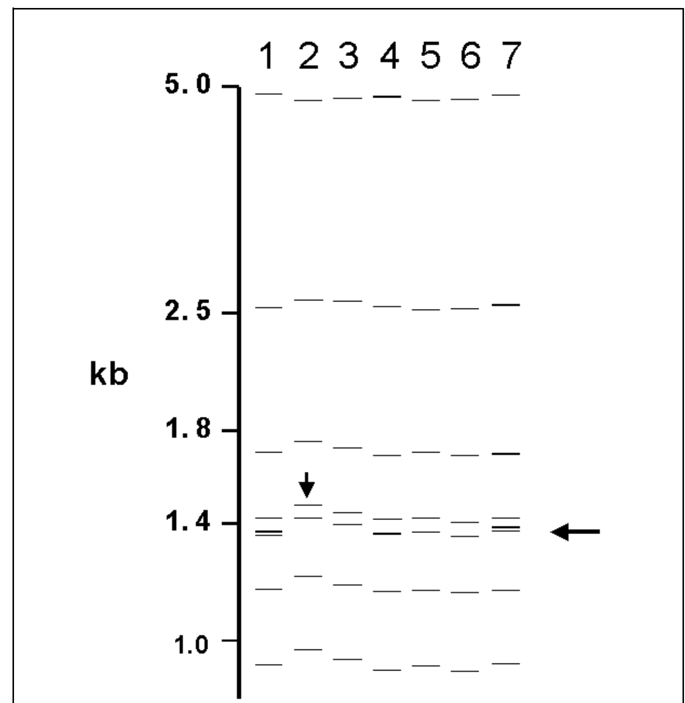

Figure 3. Computer-derived IS*6110* restriction fragment length polymorphism patterns from seven laboratories for one isolate. Misidentified doublet (side arrows) and shifted patterns (vertical arrows) are demonstrated.

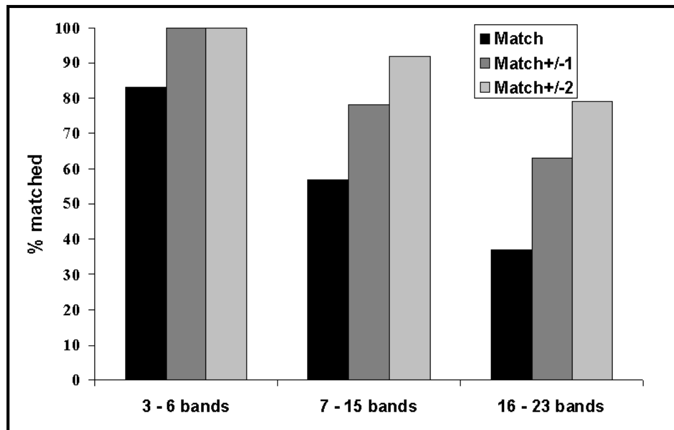

Figure 4. Quality assessment retyping match results by number of bands in the IS*6110* restriction fragment length polymorphism patterns.

bate some of the problems seen in this exercise concerning the difficulties in accurately sizing bands. Increasing the percent deviation allowed may have alleviated some observed mismatches. However, the risk of false matches of different patterns increases with a more liberal allowable deviation in band sizes in the matching process.

The problem of mismatches owing to size variations was magnified for high-molecular-weight fragments. Because of the logarithmic scale in sizing fragments, small differences in migration of large fragments result in large differences in molecular weight size calculation. Some fragments fall outside the range of the size standard (i.e., >15 kb), requiring extrapolation and causing additional inaccuracy.

In a number of instances, whole pattern shifts led to mismatching of many bands, which was typically the result of curvature across the lanes in the gel or distortion of the gel during the blotting process. Occasionally, a shift in the pattern resulted from overloading the sample DNA into the gel. These problems were apparent from observation of the image; such gels should have been rejected and run again. Pattern shifting was a greater problem during routine compilation of the genotyping network database in which images of single lanes, rather than entire gel images, were submitted. Software modifications for computerized matching may help decrease mismatching on the basis of whole pattern shifts.

The reproducibility and interlaboratory variability of IS*6110* fingerprinting in this quality assessment exercise does not necessarily represent the usual methods for pattern analysis and cluster identification for the genotyping network databases. In this exercise, results of computer matching after blind pattern editing were final. In actuality, pattern editing and cluster determination in the genotyping network were iterative processes. During normal analysis, the database manager recognized some of the problems that may occur, such as shifts in mobility; additional analysis and sample rerunning were then required to clarify the relationship of patterns. In some instances, images were reviewed, and easily reconcilable band placement was edited on the basis of definitive information about epidemiologic links among patients. The process is usually referred to as computer-assisted matching, and the database continually changed with updated pattern analysis. Thus, the outcome of the fluid process of pattern analyses and cluster determination was more accurate than the results of this exercise suggest. In addition, prospective cluster analyses and investigations were conducted at each genotyping network laboratory for patients in their respective sentinel surveillance sites; therefore, interlaboratory variability was limited to retrospective analyses of clustering of the combined database. Nonetheless, the amount of variability and nonreproducibility shown during the quality assessment is substantial and should be reflected in the interpretation of genotyping network results. Given this limitation among the genotyping network laboratories, which were experienced and well standardized, the ability to share RFLP DNA fingerprint images among laboratories that do not have strict standardization may be even more limited by interlaboratory variability.

This quality assessment exercise demonstrates the overall difficulty of combining and analyzing DNA fingerprint images from multiple laboratories. Although RFLP methodology has shown great discriminatory capacity and has been the most effective genotyping method for *M. tuberculosis* and many other pathogens, newer DNA sequence-based genotype methods should allow seamless computerization and objective analysis of results that would bypass many of the limitations described in this study. The potential benefits of these new methods are demonstrated by the near perfect reproducibility we obtained with spoligotyping. Because spoligotyping does not possess the discriminatory power needed to generally replace IS*6110* RFLP (8), a combination of spoligotyping and newer variable number tandem repeats assays (9) may provide adequate discrimination for most purposes, with IS*6110* RFLP reserved for resolution of selected sets of clustered isolates.

Acknowledgments

We thank the laboratory investigators in the National Tuberculosis Genotyping and Surveillance Network for their participation in the quality assessment exercises; Charles Woodley for assistance in culturing isolates for quality assessment panels and conducting IS*6110* RFLP; and Thomas Navin, Michael Iademarco, and Thomas Shinnick for their thoughtful reviews of this manuscript.

Dr. Braden is a medical epidemiologist in the Division of Bacterial and Mycotic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention. His research interests include surveillance and molecular epidemiology of infectious diseases.

References

1. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
2. Braden CR. Current concepts in *Mycobacterium tuberculosis* DNA fingerprinting. Infectious Diseases in Clinical Practice 1997;6:89–95.
3. Casper C, Singh SP, Rane S, Daley CL, Schecter GS, Riley LW, et al. The transcontinental transmission of tuberculosis: a molecular epidemiological assessment. Am J Public Health 1996;86:551–3.

4. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, et al. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA 1996;275:452–7.
5. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192–6.
6. Kamerbeek J, Schouls L, van Agterveld M, Kolk A, van Agterveld D, van Soolingen D, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
7. van Soolingen D, Arbeit RD. Dealing with variation in molecular typing of *Mycobacterium tuberculosis*: low-intensity bands and other challenges. J Med Microbiol 2001;50:749–51.
8. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
9. Supply P, Lesjean S, Savine K, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for the study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 2001;39:3563–71.

Address for correspondence: Christopher R. Braden, Centers for Disease Control and Prevention, Mailstop A38, 1600 Clifton Road, N.E., Atlanta, GA 30333, USA; fax: 404-639-2205; e-mail: cbraden@cdc.gov

ScholarOne™

Web-Based Solutions for Scholarly Publishers

EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends

Emerging Infectious Diseases to launch Web-based manuscript submission and peer review in 2003

Emerging Infectious Diseases will be launching **Manuscript Central**, a Web-based system for manuscript submission and peer review. Manuscript Central is operated by **ScholarOne**, a software company that specializes in scholarly publishing applications. The system allows authors, reviewers, editors, and editorial office staff direct access to journal operations through the Web.

Digital workflow expedites peer-review and administrative processes

- Allows manuscripts and reviews to be transmitted electronically between authors, the journal office, editors, and reviewers around the world
- Enables authors to submit their manuscripts as Microsoft Word, RTF, PostScript, HTML, and PDF files
- Accepts a wide variety of graphic file types and multimedia formats
- Makes manuscript files accessible to reviewers and editors through the journal's Manuscript Central site

Digital workflow eliminates mailing costs and has many other benefits

- Encourages global participation
- Enables authors to check the status of their manuscripts or update their contact information
- Allows editors direct access to an extensive database from which to select reviewers—by name, areas of expertise, key words

Built-in security ensures that only editors and reviewers assigned to a particular manuscript can view that manuscript. Reviewers, whose identities are kept confidential, have access to the manuscript only until they have completed and submitted their reviews.

Emerging Infectious Diseases, now in its 8th year of publication, welcomes submissions from infectious disease specialists in academia, industry, clinical practice, and public health, as well as from specialists in economics, demography, sociology, and other disciplines. For information on the types of articles published in Emerging Infectious Diseases, see http://www.cdc.gov/ncidod/eid/instruct.htm.

More detailed information on electronic submission and peer review will be published in upcoming issues of Emerging Infectious Diseases and on the Web at www.cdc.gov/eid.

Identifying the Sources of Tuberculosis in Young Children: A Multistate Investigation

Sumi J. Sun,* Diane E. Bennett,† Jennifer Flood,* Ann M. Loeffler,* Steve Kammerer,† and Barbara A. Ellis†

To better understand the molecular epidemiology of tuberculosis (TB) transmission for culture-confirmed patients <5 years of age, data were analyzed from a population-based study conducted in seven U.S. sites from 1996 to 2000. *Mycobacterium tuberculosis* isolates were genotyped with IS*6110*-based restriction fragment length polymorphism analysis and spoligotyping. Case-patient data were obtained from the Centers for Disease Control and Prevention's national tuberculosis registry and health department records. Routine public health investigations conducted by local health departments identified suspected source patients for 57 (51%) of 111 culture-confirmed patients <5 years of age. For 8 (15%) of 52 culture-confirmed patients <5 years of age and their suspected source patients with complete genotyping results, genotypes suggested infection with different TB strains. Potential differences between sources for patients <5 years of age and source patients that transmitted TB to adolescent and adult patients were identified.

The occurrence of tuberculosis (TB) in children is an indicator of ongoing *Mycobacterium tuberculosis* transmission and of deficiencies in current public health efforts. In the United States, strategies to prevent childhood TB include identifying and promptly initiating treatment for adults with active TB to interrupt transmission (1–4). Since children have an increased risk for developing severe disease within weeks to months of infection, they are high priorities when identified as contacts to infectious patients (5,6).

For newly diagnosed TB in children, source-case investigations are conducted to ascertain the source of infection and to prevent ongoing transmission from infectious persons. Despite efforts by TB-control programs, suboptimal numbers of source patients are identified for children (7–12). Pinpointing the source of TB infection may be particularly challenging when numerous exposures exist, including contact with persons who reside outside the United States (13). Failure to find the true source patient may have treatment implications; decisions about the treatment regimen for children often hinge on the drug-susceptibility results of the suspected source patient because cultures from young children are often not available or attempts are not made to obtain these cultures (14).

The use of molecular analysis with conventional epidemiology has increased our understanding of TB transmission (15,16). In outbreaks and population-based studies, genotyping has been instrumental in identifying previously unsuspected connections among TB patients (17). Genotyping has also been used to evaluate epidemiologic links established through contact investigations. One report found that more than one quarter of index patients and their contacts who had TB and shared a household were infected with different TB strains, indicating that transmission did not occur between the household contacts (18).

In 1996, the Centers for Disease Control and Prevention (CDC) established the National Tuberculosis Genotyping and Surveillance Network (genotyping network) to conduct population-based genotyping in seven U.S. sentinel surveillance sites (19). During a 5-year period, the network collected information on culture-confirmed patients and their contacts with TB who were identified through routine public health investigations. Study sites also attempted to collect and genotype at least one *M. tuberculosis* isolate from each reported culture-confirmed case in the surveillance area.

To better understand the molecular epidemiology of TB transmission among young children (patients <5 years of age), data collected by the genotyping network were analyzed to report the frequency that suspected source patients were identified for young children, to examine the frequency and characteristics of source patients for young children, and to determine the proportion of isolates from young children and their identified source patients with discordant genotypes. We also investigated potential differences in the characteristics of source patients who transmitted TB to young children as compared to source patients who transmitted to adolescent and adult patients.

Methods

Collection of Epidemiologic Data

A detailed description of study participants, population, and methodology is reported elsewhere (20). In brief, health department records were reviewed for all culture-confirmed patients who met the surveillance case definition (21) and

*California Department of Health Services, Berkeley, California, USA; and †Centers for Disease Control and Prevention, Atlanta, Georgia, USA

were reported from the seven sites (Arkansas, California [six counties], Maryland, Massachusetts, Michigan, New Jersey, and Texas [four counties]) from January 1996 through December 2000. Contacts (of culture-confirmed patients in the sentinel areas) with active TB were identified through routine public health investigations, as defined by local contact and source-case investigation policies and practices at each study site. Source case investigations were undertaken for all patients <5 years of age. Two sites also routinely performed source case investigations on children ≥5 years of age. Information about epidemiologically related patients identified from public health investigations was gathered with a standardized data collection form that included the direction of transmission (i.e., whether the contact was a source patient or secondary patient in relation to the index patient or whether the direction of transmission was unknown), the relationship between patients (shared a household, nonhousehold friends or relatives, co-worker, or common source), and the exposure setting (correctional, school or day-care center, workplace, emergency shelter, group quarters, hospital, nursing home, other long-term care facility, or other setting). Data were entered into Epi Info version 6d (22) databases and routinely sent to CDC. State TB registry numbers for patients in the multisite genotyping network database were matched against the CDC's national TB surveillance registry to obtain sociodemographic, behavioral, clinical, treatment, and drug-susceptibility information, which is routinely reported for all TB patients on the Report of Verified Patients of Tuberculosis (23). Project activities described in this paper were determined by CDC's institutional review board to be exempt from full committee review since genotyping of isolates was considered a public health surveillance activity and all other data used in the analysis of this paper were previously collected.

DNA Fingerprinting

Genotyping of *M. tuberculosis* isolates was conducted in accordance with standardized study protocols (20). IS*6110*-based restriction fragment length polymorphism (RFLP) analysis was performed on all available isolates. Because low-copy numbers of IS*6110* reduce test specificity, isolates containing six or fewer IS*6110* copies were further analyzed by spacer oligonucleotide typing (spoligotyping) (24). Patients were determined to have concordant genotypes if their isolates contained seven or more IS*6110* bands with identical patterns or six or fewer IS*6110* bands with identical patterns and matching spoligotypes.

Study Case Definitions

Our investigation focused on culture-confirmed patients <5 years of age; TB in young children represents recent transmission, and source patient investigations are routinely conducted for this group. A source patient was defined as a confirmed TB patient who was identified by chart abstraction as the likely source of infection for another reported TB patient. A secondary patient was defined as a confirmed TB patient who was infected by an identifiable source. Epidemiologically related source patients and secondary patients identified through routine public health investigations were considered suspected patient pairs. Because some source patients transmitted TB to more than one secondary patient, the number of suspected patient pairs does not equal the number of source patients. A secondary patient, however, could have only one designated source patient.

Genotypes for isolates from suspected patient pairs were compared, and patient pairs were categorized as 1) confirmed patient pairs, if isolates had concordant genotypes, 2) refuted patient pairs, if isolates had discordant genotypes, and 3) undetermined patient pairs, if genotypes were unavailable for the patient pair.

Data Analysis

Data in the multisite genotyping network database were analyzed with SAS version 8.0 (25) and Epi Info version 6d (22) software packages. Patients were excluded from analysis when records were not available for review or lacked complete information from public health investigations, including three patients <5 years of age from one site, for whom source patients were not identified but who were entered into the database as the source for an adult case. Because young children are not typically considered to be infectious (26) and records for these patients were not available for further examination at the time of this analysis, information was determined to be incomplete for these patients.

Univariate analysis was conducted to examine factors associated with the identification of source patients for young children and to investigate associations between key variables and the identification of refuted patient pairs. Differences in proportions were assessed with the chi-square statistic or 2-tailed Fisher exact test. Relative risks (RR) and 95% confidence intervals of point estimates were generated where appropriate. Differences in the means of continuous data were tested with the Wilcoxon rank-sum test when sample sizes were small. Unless otherwise noted, p values <0.05 were interpreted as statistically significant differences for all statistical tests.

Genotypes of isolates from young children without a known source patient were matched against the genotyping network project database to find previously unidentified adult TB patient(s) whose genotype matched the child's. Since the sentinel study sites represented geographically dispersed states that did not necessarily share a common border, genotype matches were limited to patients from the same site.

To better describe the unique characteristics of patients who transmit TB to young children, source patients (in confirmed patient pairs) who transmitted TB to young children were compared with those who transmitted to adolescents or adult patients. Since source patients who infected children 5 years of age or older may be very similar to source patients who infected children newly born to 4 years of age, two different comparison groups were identified 1) source patients for

all secondary patients ≥5 years of age and 2) source patients for secondary patients ≥15 years of age (excluding source patients that transmitted to children 5–14 years of age).

Results

Characteristics of Children with TB

From 1996 to 2000, a total of 15,035 TB patients were reported from the seven sentinel surveillance sites; 11,923 (79%) were culture confirmed, and isolates from 10,752 (90%) culture-confirmed patients were genotyped. Of all patients in the study, 518 (3%) patients were <5 years of age. Culture was attempted in 270 (52%) patients <5 years of age, and 122 (45%) of these patients were culture confirmed. Isolates from 114 (93%) culture-confirmed children <5 years of age were genotyped.

Texas and California sites reported 73 (60%) of the 122 culture-confirmed patients <5 years of age; the Michigan and New Jersey sites reported 18 patients each, and the remaining three sites reported ≤6 patients each. Most (65%) of the study patients were <2 years of age, and 49% were girls. Forty-three percent were black, non-Hispanic; 37% Hispanic; 15% Asian; 4% white, non-Hispanic, and 2% Native-American or Alaskan Native. Of the 11 foreign-born patients <5 years of age, 4 were from Mexico, 2 were from Kenya, and 5 were from other countries. Two thirds of the young children had pulmonary TB disease, 15% had extrapulmonary disease, and 20% had both pulmonary and extrapulmonary TB.

With some notable exceptions, culture-confirmed patients <5 years of age had demographic or clinical characteristics similar to those of the 396 young children from the surveillance area who were either culture-negative or did not have a specimen collected for culture. The culture-confirmed group was more likely to be ≤1 year old (RR=1.94, p=<0.001); whereas white, non-Hispanic children (RR=0.41, p=0.02) and those treated only by private providers (RR=0.6, p=0.002) were underrepresented in the sample of culture-confirmed patients.

Suspected Source Patients for Young Children

Results of routine investigations used in identifying source patients for culture-confirmed children <5 years of age are presented in Figure 1. Health department records were unavailable or lacked sufficient information about investigations of 11 patients; these records were excluded. At least one epidemiologically related case was identified for 66 (59%) young children with culture-confirmed TB; 57 (86%) patients had a source patient designated, but a source could not be determined for the remaining 9 patients, although an epidemiologically related case was identified. For five of the nine patients, multiple epidemiologically related patients (ranging from 2–11 related patients) were identified.

To examine factors associated with the identification of source patients for culture-confirmed children <5 years of age, we compared young children with a suspected source patient to patients with an unknown source of infection (Table 1). Children ≤2 years of age were more likely to have a source patient identified from routine public health investigations; however, source patients were less frequently found for foreign-born children. No other statistically significant differences were found. Drug-susceptibility patterns for isolates from young children with any drug resistance are detailed in Table 2.

Table 3 lists characteristics of the 53 source patients identified from public health investigations. In 41 (72%) of 57 suspected patient pairs involving young children, the source patient lived in the child's household. Of the 16 nonhousehold sources, 3 were babysitters, 4 were neighbors or visitors, 2 were relatives, and 1 attended the same church as the child's family; the specific relationship was unknown for 6 patient pairs. Eight (15%) of the source patients resulted in disease in more than one young child (including culture-negative children and patients outside of the study population).

Molecular Fingerprint Data

Of the 57 culture-confirmed patients <5 years of age for whom a source patient was identified, 91% (52) had genotyping results for both the young child and the suspected source patient (Figure 2). Forty-four (85%) of 52 suspected patient pairs had concordant genotypes, and 8 (15%) of 52 had discordant genotypes. Young children in refuted patient pairs were more likely to be older than those in confirmed patient pairs (Table 4). No association between gender, ethnicity, or foreign-born status of patients and the identification of refuted patient pairs was found. Nearly three quarters (37 of 52) of suspected patient pairs lived in the same household; however,

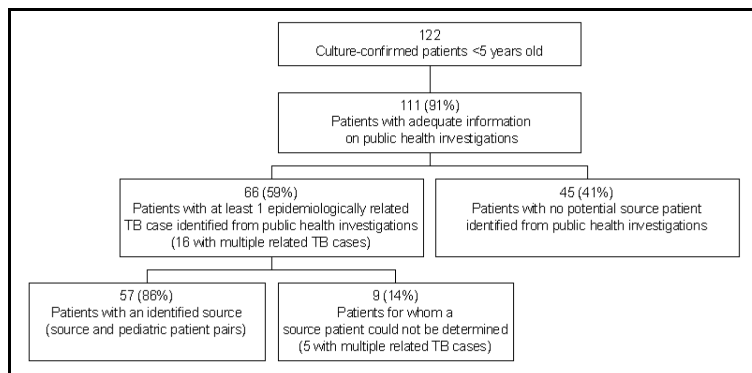

Figure 1. Results of public health investigations for culture-confirmed tuberculosis patients <5 years of age, 1996–2000.

Table 1. Factors associated with identifying source patients for culture-confirmed tuberculosis in children <5 years of age[a]

Characteristics	Suspected sources identified n=57 (%)		No suspected source identified n=45 (%)		Relative risk (95% CI)	p value
Age ≤2 yrs[b]	44	(77)	20	(45)	1.96 (1.23 to 3.12)	0.001
Female	30	(53)	21	(47)		NS
Race or ethnicity						NS
Black, non-Hispanic	26	(46)	19	(42)		
Hispanic	23	(40)	12	(27)		
Asian	5	(9)	11	(24)		
White, non-Hispanic	1	(2)	3	(7)		
Native American or Alaskan Native	2	(4)	0	(0)		
Foreign-born[c]	1	(2)	7	(16)	0.21 (0.03 to 1.31)	0.02
Type of disease						NS
Pulmonary only	40	(70)	30	(67)		
Extrapulmonary only	5	(9)	8	(18)		
Pulmonary and extrapulmonary	12	(21)	7	(16)		
Provider type[d]						NS
Health department	17	(31)	10	(23)		
Private provider	18	(33)	21	(49)		
Both	20	(36)	12	(28)		
Directly observed therapy[e]	46	(85)	28	(68)		NS
Drug-resistant isolate[f]	6	(11)	8	(16)		NS

[a]NS, not significant; CI, confidence interval.
[b]Age at start of treatment. Excludes one child whose date of treatment was unknown.
[c]Excludes one child whose birthplace was unknown.
[d]Excludes four children whose provider type was unknown.
[e]Compared to patients on self-administered therapy.
[f]Drug resistance on initial testing of isolate; resistance to at least one of the following: isoniazid, rifampin, ethambutol, pyrazinamide, streptomycin, and ethionamide. Testing results for one or more drugs could have been unknown or not done. Excludes two children for whom drug-susceptibility testing was not done.

5 (14%) of these patient pairs had discordant genotypes. Suspected patient pairs with differing drug-susceptibilities were not associated with discordant genotypes; all three patient pairs with differing drug resistance patterns had concordant genotypes (Table 5).

For the nine young patients who had at least one epidemiologically related patient identified by public health investigations but for whom the source patient could not be determined, genotyping patterns from the isolates of the epidemiologically related cases and the young child were identical, almost without exception. The only two discordant genotypes were in young children with a single related case, not among the five young children with multiple related patients.

Genotyping also identified patients in the local surveillance site who had the same genotype as young children without an identified source patient. Isolates were genotyped from 40 of 45 patients <5 years of age without a known source patient. Of these genotyped isolates, 23 (58%) matched the strain from at least 1 adult pulmonary TB case in the local surveillance site. For most young children (13 [57%] of 23) without an identified source patient, at least 5 adult pulmonary TB patients with genotypes matching the child's were identified. We found a wide range in the number of adult patients (2–128) with genotypes matching the genotypes of these young children.

Confirmed Source Patients for Children, Adolescents, and Adults

To better characterize the unique attributes of patients who transmit TB to young children, characteristics of their source patients (in confirmed patient pairs) were compared with those for adults and adolescents. No significant differences were found when the comparison group for this analysis consisted of all sources to secondary patients ≥5 years of age or when the comparison group was limited to sources to secondary patients ≥15 years of age. The results of the latter comparison are presented. More than 60% (354 of 584) of the suspected patient pairs in which the secondary patient was not a child were genotyped, and 240 (68%) of these patient pairs had concordant genotypes (Figure 2). The likelihood of identifying patient pairs with discordant genotypes was more than two times higher among suspected patient pairs involving secondary patients ≥15 years of age than for those involving young

Table 2. Drug-resistent patterns for isolates from culture-confirmed patients <5 years of age with and without a suspected source patient identified[a]

Suspected source patient identified	Source patient not identified
Ethionamide	Isoniazid
Streptomycin	Streptomycin
Streptomycin	Streptomycin
Isonazid, streptomycin	Pyrazinamide
Isonazid, streptomycin	Pyrazinamide
Isonazid, rifampin, ethambutol streptomycin	Isoniazid, streptomycin
	Isoniazid, rifampin
	Isoniazid, ethambutol, streptomycin

[a]Through routine public health investigations.

children (32% vs. 15% discordant genotypes) (RR=2.09, p=0.01).

Univariate associations between source patient characteristics and transmission to young children were assessed (Table 6). Although the mean age for sources to secondary patients <5 years of age was slightly lower than the mean age of sources to the comparison group, these differences were not significant (p=0.06; Wilcoxon test). For this population, confirmed source patients to young children were more likely to be foreign-born (p=0.02), Hispanic (p<0.001), a household member (p<0.001), and not receiving directly observed therapy (p<0.01) as compared with sources for adolescents and adults.

Discussion

Despite the continued decline in the number of TB patients in the United States, ongoing TB transmission persists in many communities. For public health agencies, TB in young children signals recent transmission and missed opportunities for TB prevention. In this investigation, molecular tools were used in conjunction with information from conventional public health investigations to better understand issues related to the identification of source patients for young children.

In this multisite study, 57 (51%) of 111 culture-confirmed patients <5 years of age had a source patient identified by routine investigations. Although this finding is comparable to the frequency of source patient identification reported for other subpopulations of children with TB (8,10–11), the finding may be lower than anticipated for a sample of young children predominantly born in the United States. Children ≤2 years of age and those born in the United States were more likely to have a source patient found than children without these characteristics. These results corroborated findings from a study of children <5 years of age with TB in California, which demonstrated that the source of infection is more likely to be identified for children who were found in a contact investigation, born in the United States, <1 year of age, or black (9).

Children <5 years of age with an unknown source of infection composed a substantial proportion of the study population (41%), a finding that underscores shortcomings in identifying all contacts of infectious patients. While molecular data alone are not enough to prove recent transmission, the presence of infectious TB patients in the community who share the same strain with a young child without a known source suggests the possibility of casual transmission. Other impediments in identifying source patients may include barriers in completing contact investigations, delays in evaluation, and problems in identifying source patients who reside outside the health department's jurisdiction (27). Eighty-four percent of young children without a source patient in this study were born in the United States; this observation is likely to underestimate the contribution of the global TB epidemic, because TB surveillance systems in the United States do not routinely monitor the birthplace or travel history of parents or guardians, factors pre-

Table 3. Demographic, clinical, and risk characteristics of 53 source patients with tuberculosis (TB) identified from public health investigations[a]

Source patient characteristics	No. (%)
Age group, yrs	
15–24	11 (21)
25–44	28 (53)
45–64	10 (19)
65+	4 (8)
Female	24 (45)
Race or ethnicity	
Black, non-Hispanic	22 (42)
Hispanic	24 (45)
Asian	5 (9)
Native American or Alaskan Native	2 (4)
Foreign-born[b]	27 (51)
Bacteriologic results, sputum	
Smear positive/culture positive	42 (79)
Smear positive/culture negative	1 (2)
Smear negative/culture positive	8 (15)
Smear not done/culture not done	2 (4)
Cavitary chest radiograph[c]	33 (63)
Provider type	
Health department	31 (58)
Private provider	11 (11)
Both	11 (11)
Directly observed therapy[d]	47 (71)
Previous diagnosis of TB	5 (9)
Drug-resistant isolate[e]	5 (9)

[a]Three source patients were identified as the source of infection for more than one culture-confirmed patient who was <5 years of age in the sentinel study population; two source patients transmitted to two children, and one transmitted to three children.
[b]Country of origin was Mexico for 14 (26%) of the foreign-born patients.
[c]Results unknown for one patient.
[d]Compared to patients on self-administered therapy.
[e]Drug resistance on initial testing of isolate; resistance to at least one of the following: isoniazid, rifampin, ethambutol, pyrazinamide, streptomycin, and ethionamide. Testing results for one or more drugs could have been unknown or not done. Excludes one source patient for whom drug susceptibility testing was not done.

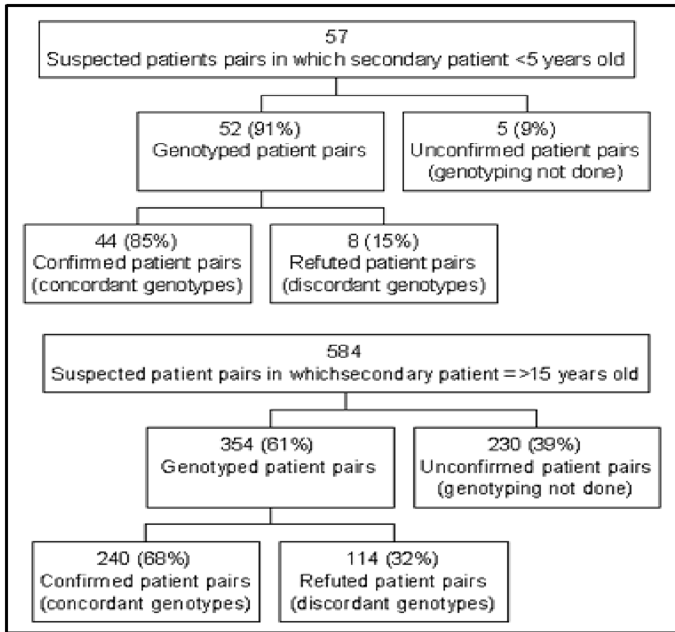

Figure 2. Comparison of genotypes for isolates from suspected source and secondary patient pairs identified through public health investigations, 1996–2000. Children <5 years of age with culture-confirmed TB and their sources are compared with patients $\geq$15 years of age with culture-confirmed TB and their sources.

viously identified as significant predictors for pediatric TB (28,29).

Of particular concern is the finding that 16 (14%) of 111 young children with culture-confirmed TB had more than one epidemiologically related TB source identified. This finding indicates that a substantial number of children have multiple TB exposures that need to be carefully assessed. For most, the source of infection was ascertained and later confirmed by genotyping analysis. When multiple epidemiologically related patients existed, but none was identified as the source patient, genotyping analysis did not provide added benefit since the related patients were more likely to have the same genotype.

Clinicians and TB-control programs often rely on the drug-susceptibility results of the suspected source patient to guide the treatment of the child since specimens for culture are not frequently collected from children (14). Previous studies by Steiner et al. reported 93% to 96% concordance of drug-susceptibility patterns of TB isolates from children <15 years of age and their source (30,31), comparable to the 93% drug-susceptibility concordance among suspected patient pairs in our study population. All suspected patient pairs with discordant drug-susceptibility results were among patient pairs with con-

Table 4. Characteristics of refuted and confirmed patients pairs[a]

	Refuted patient pairs (n=8) (%)	Confirmed patient pairs (n=44) (%)
Characteristics of young children		
Mean age, months[b]	16	13
Female	4 (50)	23 (52)
Race or ethnicity		
Black, non-Hispanic	5 (63)	20 (45)
Hispanic	3 (38)	16 (36)
Asian	0 (0)	5 (11)
White, non-Hispanic	0 (0)	1 (2)
Native American/Alaskan Native	0 (0)	2 (5)
Foreign-born	1 (13)	0 (0)
Source patient characteristics		
Mean age, yrs	25	31
Female	6 (75)	18 (41)
Race or ethnicity		
Black, non-Hispanic	5 (63)	20 (45)
Hispanic	3 (38)	17 (39)
Asian	0 (0)	5 (11)
Native American or Alaskan Native	0 (0)	2 (5)
Foreign-born	3 (38)	20 (45)
Case-patient characteristics		
Shared household	5 (63)	32 (73)
Discordant drug susceptibilities[c]	0 (0)	3 (7)
Different race or ethnicity	0 (0)	1 (2)

[a]For tuberculosis patients <5 years of age and their suspected source patients. Refuted patient pairs are suspected patient pairs with discordant genotypes; confirmed patient pairs are suspected patient pairs with concordant genotypes.
[b]Wilcoxon rank-sum test: p=0.03.
[c]Excludes two patient pairs in which the children had drug-resistant *Mycobacterium tuberculosis* strains (streptomycin and ethambutol resistance, respectively), but susceptibility results were not done for the identified source patient.

Table 5. Comparison of drug-susceptibility and genotyping results for isolates of suspected patient pairs[a] with any drug resistance

Source-patient isolate	Secondary patient isolate (children <5 years of age)	Drug-susceptibility comparison (patient pairs)	Genotype comparison (patient pairs)
Isoniazid	Isoniazid, streptomycin	Discordant	Concordant
Isoniazid, rifampin	Isoniazid, rifampin, ethambutol, streptomycin	Discordant	Concordant
Streptomycin	Pan-susceptible	Discordant	Concordant
Streptomycin	Streptomycin	Concordant	Concordant
Isoniazid, streptomycin	Isoniazid, streptomycin	Concordant	Concordant
Pan-susceptible[b]	Ethionamide	Undetermined[c]	Concordant
Not done	Streptomycin	Undetermined	Undetermined

[a]Tuberculosis patients <5 years of age and their suspected source patients.
[b]Isolate from source patient was not tested for ethionamide resistance.
[c]Undetermined, results unknown for one or both patients.

Table 6. Characteristics of source patients in confirmed patient pairs[a]

Source patient characteristics	Confirmed sources for secondary patients <5 yrs of age (n=44) (%)	Confirmed sources for secondary patients ≥15 ys of age (n=240) (%)
Mean age, yrs	31	38
Female	18 (41)	103 (43)
Race or ethnicity[b]		
Black, non-Hispanic	20 (45)	148 (62)
Hispanic	17 (39)	23 (10)
Asian	5 (11)	19 (8)
White, non-Hispanic	0	48 (20)
Native American or Alaskan Native	2 (5)	2 (1)
Foreign-born[c]	20 (45)	56 (23)
Bacteriologic results, sputum[d]		
Smear positive/culture positive	34 (81)	200 (85)
Smear negative/culture positive	8 (19)	33 (14)
Smear negative/culture negative	0	1 (<1)
Cavitary chest radiograph[e]	27 (64)	125 (53)
Provider type[f]		
Health department	25 (57)	141 (60)
Private provider	9 (20)	43 (18)
Both	10 (23)	52 (22)
Directly observed therapy [b,g]	30 (68)	198 (84)
Previous diagnosis of TB[h]	6 (14)	27 (11)
Shared household with secondary case-patient [b,i]	32 (91)	116 (50)

[a]Confirmed patient pairs include source patients who transmitted TB to young children and source patients who transmitted TB to adolescent and adult patients.
[b]Chi-square statistic, $p<0.05$.
[c]Country of origin unknown for one patient.
[d]Excludes eight patients in whom either the culture or smear was not done.
[e]Chest radiograph results unknown for two patients.
[f]Provider type unknown for four patients.
[g]Compared to patients on self-administered therapy only. Directly observed therapy status unknown for six patients.
[h]History of TB unknown for two patients.
[i]Relationship to secondary case unknown for 19 patients.

cordant genotypes, indicating the value of drug-susceptibility results in young children, even when genotyping results are known to the local health department.

The high frequency (85%) of concordant genotypes among young children and their source patients represents good news for public health agencies; when a potential source of infection was identified in this population, it was most often accurate. However, for as many as 15%, the true source was not identified and presumably could have contributed to the further spread of disease in the community. Because young children may have more limited opportunities for exposure than older children, we anticipated that the frequency of confirmed patient pairs would be associated with young age. We also speculated that foreign-born children, especially those from high TB-prevalence areas, might have an increased risk of being involved in a refuted patient-pair. These children might have had multiple opportunities for exposure to active TB before entering the United States, which may increase the possibility that the source of infection could have been someone other than the suspected source patient. However, this potential association could not be assessed because our sample of foreign-born children with culture-confirmed TB was small.

The increased likelihood of concordant genotypes among suspected patient pairs involving young children as compared with suspected patient pairs that did not include children (85% vs. 68%) may be explained by a number of factors, including the greater number of casual contacts with whom adults interact, biases in the case-finding practices for these groups, and potential for coincidental reactivation of a latent TB infection in older patient pairs. Source patients who transmitted TB to young children were more likely to be Hispanic, foreign-born, a household member, and not receiving directly observed therapy as compared to sources for adolescents or adults. The latter may indicate nonadherence of source patients to drug treatment and corroborates an observation by Kimerling et al. (32). However, additional data are needed to determine the confounding factors, including site-to-site variance, that may affect which source patients receive directly observed therapy, as discussed in the limitations that follow.

A key limitation in this study was the inability to assess the effect of potential confounding factors, such as differences in case-finding methods (i.e., if patients were identified through contact investigation, source patient investigations, or screening activities) on the outcome of interest (i.e., identification of source patients or confirmed patient pairs). These data represented the sites' routine public health practices and policies, since uniform policies for public health investigations were not instituted, and potential systematic variances across sites were not ascertained by the project. In addition, analysis of epidemiologic investigations for infectious patients in the community who shared the same TB strain as the young child but were not identified from routine public health investigations was outside the scope of this paper. A follow-up investigation to find epidemiologic connections among patients currently linked by genotyping results alone may provide important data regarding potential missed opportunities in this group. Finally, the predictive value of a discordant genotype result is not yet known. Although study protocols instituted quality-control measures across genotyping laboratories, a subset of isolates from suspected patient pairs who were determined to have discordant genotypes might include TB strains that are potentially the same. Thus, the proportion of discordance observed in this study may represent an overestimate of the actual frequency of suspected patient pairs with discordant genotypes.

This study highlights the challenges in identifying the sources of infection for children under 5 years of age with culture-confirmed TB and potential weakness in our current TB-control and prevention practices in this population. Although contact and source patient investigations are central to any TB-control strategy, the usefulness of these activities in identifying

the true source of infection in young children has not been previously evaluated for a large population of children by using molecular methods. While indicating a high degree of concordance between genotypes from young children and their identified sources, genotyping analysis also refuted some source patients and pointed to other potential sources in the community who were previously unsuspected. Further assessment of shortcomings in current methods to prevent transmission to children and to identify their source of infection is warranted to ultimately eliminate TB in young children in the United States.

Acknowledgments

The authors thank Chris Braden and Scott McNabb for scientific guidance and for overseeing the genotyping network project, Marisa Moore for supplying data from the Centers for Disease Control and Prevention's national TB registry, and all participants of the genotyping network for their efforts to ensure the success of the project.

This work was supported by the Centers for Disease Control and Prevention through funds for the genotyping network.

Ms. Sun is a research scientist in the Surveillance and Epidemiology Section of the California Tuberculosis Control Branch.

References

1. American Thoracic Society. Control of tuberculosis in the United States. Am Rev Respir Dis 1992;146:1623–33.
2. Perry S, Starke JR. Adherence to prescribed treatment and public health aspects of tuberculosis in children. Seminars in Pediatric Infectious Diseases 1993;4:291–8.
3. Centers for Disease Control and Prevention. Screening for tuberculosis and tuberculosis infection in high-risk populations: recommendations of the Advisory Council for the Elimination of Tuberculosis. MMWR Morb Mortal Wkly Rep 1995;44:19–34.
4. Centers for Disease Control and Prevention. Essential components of a tuberculosis prevention and control program: recommendations of the Advisory Council for the Elimination of Tuberculosis. MMWR Morb Mortal Wkly Rep 1995;44:1–16.
5. Fujiwara PI, editor. Clinical policies and protocols. New York: New York City Department of Health, Bureau of Tuberculosis Control; 1999. p. 93–7.
6. Marshman FC, editor. Prevention and control of pediatric tuberculosis in New York City: recommendation for the expert advisory panel. New York: New York City Department of Health; 1995. p. 18–9.
7. Driver CR, Luallen JJ, Good WE, Valway SE, Frieden TR, Onorato IM. Tuberculosis in children younger than five years of age: New York City. Pediatr Infect Dis J 1995;14:112–7.
8. Gessner BD. Incidence rates, clinical features, and case identification of pediatric tuberculosis in Alaska. Int J Tuberc Lung Dis 1998;2:378–83.
9. Lobato MN, Mohle-Boetani JC, Royce SE. Missed opportunities for preventing tuberculosis among children younger than five years of age. Pediatrics 2000;106:E75.
10. Lobato MN, Loeffler AM, Furst K, Cole B, Hopewell PC. Detection of *Mycobacterium tuberculosis* in gastric aspirates collected from children: hospitalization is not necessary. Pediatrics 1998;102:E40.
11. Moss W. Tuberculosis in the home: contact history and childhood tuberculosis in central Harlem. Clin Pediatr (Phila) 1998;37:753–5.
12. Watchi R, Kahlstrom E, Vachon LA, Barnes PF. Pediatric tuberculosis: clinical presentation and contact investigation at an urban medical center. Respiration 1998;65:192–4.
13. Sullam PM, Slutkin G, Hopewell PC. The benefits of evaluating close associates of child tuberculin reactors from a high prevalence group. Am J Public Health 1986;76:1109–11.
14. Al-Dossary F, Ong L, Correa A, Starke JR. Treatment of childhood tuberculosis with a six month directly observed regimen of only two weeks daily therapy. Pediatr Infect Dis J 2002;21:91–2.
15. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 1999;130:971–8.
16. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL, et al. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet 1999;353:444–9.
17. Chin DP, Crane CM, Diul MY, Sun SJ, Agraz R, Taylor S, et al. Spread *of Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. JAMA 2000;283:2968–74.
18. Behr MA, Hopewell PC, Paz EA, Kawamura LM, Schecter GF, Small PM. Predictive value of contact investigation for identifying recent transmission of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1998;158:465–9.
19. Castro KG, Jaffe HW. Rationale and methods for the National Tuberculosis Genotyping and Surveillance network. Emerg Infect Dis 2002;8; 1188–91.
20. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192–6.
21. Centers for Disease Control and Prevention. Case definitions for public health surveillance. MMWR Morb Mortal Wkly Rep 1990;39:39–40.
22. Epi Info: a word processing, database, and statistics program [computer program]. Version 6. Atlanta: Centers for Disease Control and Prevention; 1994.
23. Centers for Disease Control and Prevention. Tuberculosis information management system (TIMS) user's guide. Surveillance appendix 1. Atlanta: U.S. Department of Health and Human Services, Public Health Service; 1999.
24. Bauer J, Andersen AB, Kremer K, Miorner H. Usefulness of spoligotyping to discriminate IS*6110* low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J Clin Microbiol 1999;37:2602–6.
25. SAS/STAT user's guide. Cary (NC): SAS Institute Inc.; 2000.
26. Scientific Committee of the International Union Against Tuberculosis and Lung Disease. Tuberculosis in children: guidelines for diagnosis, prevention, and treatment. Bull Int Union Tuberc Lung Dis 1991;66:114–9.
27. Besser RE, Pakiz B, Schulte JM, Alvarado S, Zell ER, Kenyon TA, et al. Risk factors for positive mantoux tuberculin skin tests in children in San Diego, California: evidence for boosting and possible foodborne transmission. Pediatrics 108:305–10.
28. Kenyon TA, Driver C, Haas E, Valway SE, Moser KS, Onorato IM. Immigration and tuberculosis among children on the United States-Mexico border, County of San Diego, California. Pediatrics 1999;104:E8.
29. Saiman L, San Gabriel P, Schulte J, Vargas MP, Kenyon T, Onorato I. Risk factors for latent tuberculosis infection among children in New York City. Pediatrics 2001;107:999–1003.
30. Steiner M, Zimmerman R, Park BH, Shirali SR, Schmidt IT. Primary tuberculosis in children: II. Correlation of susceptibility patterns of *Mycobacterium tuberculosis* isolated from children with those isolated from source cases as an index of drug resistant infection in a community. Am Rev Respir Dis 1968;98:201–9.
31. Steiner P, Rao M, Mitchell M, Steiner M. Primary drug-resistant tuberculosis in children. Correlation of drug-susceptibility patterns of matched patient and source case strains of *Mycobacterium tuberculosis.* Am J Dis Child 1985;139:780–2.
32. Kimerling ME, Vaughn ES, Dunlap NE. Childhood tuberculosis in Alabama: epidemiology of disease and indicators of program effectiveness, 1983 to 1993. Pediatr Infect Dis J 1995;14:678–84.

Address for correspondence: Sumi J. Sun, 2151 Berkeley Way, Room 608, Berkeley, CA 94704, USA; fax: 510-849-5269; e-mail: ssun1@dhs.ca.gov

DNA Fingerprinting of *Mycobacterium tuberculosis* Isolates from Epidemiologically Linked Case Pairs

Diane E. Bennett,* Ida M. Onorato,* Barbara A. Ellis,* Jack T. Crawford,* Barbara Schable,* Robert Byers,* J. Steve Kammerer,* and Christopher R. Braden*

DNA fingerprinting was used to evaluate epidemiologically linked case pairs found during routine tuberculosis (TB) contact investigations in seven sentinel sites from 1996 to 2000. Transmission was confirmed when the DNA fingerprints of source and secondary cases matched. Of 538 case pairs identified, 156 (29%) did not have matching fingerprints. Case pairs from the same household were no more likely to have confirmed transmission than those linked elsewhere. Case pairs with unconfirmed transmission were more likely to include a smear-negative source case (odds ratio [OR] 2.0) or a foreign-born secondary case (OR 3.4) and less likely to include a secondary case <15 years old (OR 0.3). Our study suggests that contact investigations should focus not only on the household but also on all settings frequented by an index case. Foreign-born persons with TB may have been infected previously in high-prevalence countries; screening and preventive measures recommended by the Institute of Medicine could prevent TB reactivation in these cases.

Investigating persons who have had close contact with tuberculosis (TB) cases is an essential element of public health programs to control and eliminate TB (1,2). These contact investigations are done primarily to discover persons who may require treatment for latent TB infection and also to find and treat additional persons with TB. While not usually highly contagious, TB is generally transmitted to persons who have shared indoor air space frequently or for a long period of time with a person who is infectious (3). Factors that may influence transmission include prolonged hours of contact during the infectious period, close proximity to the person with TB, and lack of ventilation and ultraviolet light in a shared environment. Generally, close contacts who live with a person identified with active TB or who habitually spend time indoors in close proximity to this person are investigated first. If no evidence of TB transmission is found in these close contacts, the investigation ceases. If transmission has occurred, the investigation may be extended. The "stone-in-the-pond" principle, a technique in which concentric circles of contact persons around the case are sequentially investigated, is practiced in many countries (4).

If one or more additional persons with TB are identified among the contacts of a person with TB, the person is labeled as the index case for the purpose of the investigation; those subsequently identified are classified as source cases, secondary cases, or unlinked cases. An active TB case found in a contact investigation may be classified as the source of infection to the index case, a secondary case infected by the index case, or a case who neither infected nor was infected by the index case (with a strain of TB unrelated to the strain of the index case). Information about the start and duration of symptoms for the index and the contact cases, and the start and duration of contact between them, facilitates categorization. Categorizing a contact with active TB as a source case, a secondary case, or an unlinked case, based on epidemiologic evidence, helps to direct further investigation. If the source case is known to have drug-resistant TB, establishing epidemiologic links may also aid in choosing an appropriate drug regimen for the contact before cultures and sensitivity test results are available (5).

The chief priority of TB-control programs is to identify and treat active cases of TB before transmission can occur (1,3). If an index case is identified and treated soon after symptoms begin, the time during which TB could be transmitted can be minimized, and active secondary cases are unlikely to be found in contact investigations. If contact investigations are carried out soon after the index case is identified, the time is minimized during which infected persons could progress to disease before receiving treatment for latent TB infection to prevent progression. In a low-prevalence country, a well-resourced and active TB-control program would not expect to find a high proportion of active TB cases among contacts in investigations.

Systematic evaluations of contact investigations are infrequent. In an Australian study, of 1,142 close contacts of 231 cases diagnosed in 1991, a mean of 4.9 contacts per case were identified, but only 3 (0.3%) of the contacts had active TB (6). However, the authors stated that the screening of these contacts was inadequate so TB may have been underdiagnosed. A 1996–1997 U.S. study of contact investigations of 1,080 pul-

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

monary, smear-positive TB patients found a median of four close contacts per patient (7). Thirty-six percent of contacts were tuberculin skin-test positive, and 2% had active TB. A systematic review of health department records for all contacts of 349 patients with culture-positive TB in five study areas in 1996 revealed that 13% had not identified contacts (8). Although 3,824 contacts were identified, only 2,095 (55%) completed screening; of these, 1% had active TB.

DNA fingerprinting has been used to support contact investigations of large clusters of cases in institutional settings and to suggest possible connections among cases without obvious epidemiologic links. This technique is also used, though rarely, to evaluate epidemiologic links found in contact investigations. In San Francisco, culture-positive TB cases previously identified as contacts to active TB cases were evaluated along with their index cases (9); a median of four contacts was investigated for each of the 1,308 culture-positive index cases reported from 1991 to 1996. Of 11,211 contacts evaluated, 108 (1%) had active TB. Of 94 pairs of index and contact cases with active TB, 66 had positive cultures; of these, 54 had restriction fragment length polymorphism results for both strains. Transmission was confirmed (that is, the same strain was identified in both cases) in 38 (70%) of the 54 pairs.

Between 1996 and 2000 in the United States, the National Tuberculosis Genotyping and Surveillance Network collected information on contacts with culture-confirmed TB identified during the course of contact investigations and medical record reviews in seven sentinel areas (the states of Arkansas, Massachusetts, Maryland, Michigan, and New Jersey and selected sites in California and Texas). These data, combined with the results of DNA fingerprinting, were analyzed to evaluate the proportion of epidemiologically linked cases in which transmission was confirmed by matching fingerprints and to investigate the characteristics of case pairs with unconfirmed transmission (unmatched fingerprints).

Methods

Epidemiologic Information from Contact Investigations

Details regarding the study population are presented elsewhere (10,11). Information captured from routine contact investigations and medical record reviews (done before fingerprinting data were available) were entered into an Epi Info version 6.03 database (Centers for Disease Control and Prevention [CDC], Atlanta, GA) and sent quarterly to CDC. Participating sites entered information on any index cases reported as a definite TB case to the state TB surveillance system from January 1996 to December 2000 and on any contact to an index case reported as a definite TB case between January 1990 and December 2000.

Because the study was directed at an evaluation of information acquired during routine public health procedures, information acquired in cluster investigations done because of DNA fingerprint clustering was not included. Information from epidemiologic cluster investigations conducted because more than one case of TB was noted in a congregate setting, before fingerprinting had been performed, was included.

Participating sites recorded the nature of the relationship between the index case and each contact, the setting in which the two persons interacted, and the direction of transmission (i.e., whether the contact was identified as the source case in relation to the index case, a secondary case in relation to the index case, or whether the direction of transmission remained unidentified). In our analysis, we included only case pairs in which the direction of transmission was specified.

Demographic and Clinical Information

Demographic and clinical information on source and secondary cases were noted by matching state case numbers to the national TB surveillance system database at CDC. Data reported included gender, race/ethnicity, age at diagnosis, country of origin, previous episodes of TB, sputum smear status at diagnosis, chest x-ray results, drug susceptibility results, HIV status, and occupation.

Laboratory Methods

In this study, a source case was defined as a person with active TB identified in a routine contact investigation as the probable source of transmission to another person with active TB. A secondary case was a person with active TB identified in a routine contact investigation as having acquired TB from one source case. An epidemiologically linked case pair consisted of two persons identified respectively as linked source and secondary case in the course of a contact investigation. A secondary case could not be linked to more than one source case. However, one or more secondary cases could be linked to a single source case.

Transmission was considered confirmed if matching DNA fingerprint patterns were found in both isolates from a case pair. We attempted DNA fingerprinting for all available isolates from participating sites during the period of the study. The methods we used, including a definition for matching fingerprint patterns, are described in a related article (10).

Statistical Methods

Univariate associations between demographic, setting, and clinical variables, and the dependent variable (TB transmission unconfirmed by DNA fingerprinting) were examined by using the Cochran-Mantel-Haenszel test for unequal odds using SAS version 8.0 (SAS Institute, Inc., Cary, NC). Associations were also examined by multiple logistic regression analysis in SPSS version 6.0 (SPSS, Inc., Chicago, IL). The correlation matrix produced was examined for potential colinearity between variables. Goodness-of-fit analysis was performed by using the SPSS Hosmer-Lemeshow goodness-of-fit test (12). To choose the best-fitting model, we used the likelihood ratio test to compare models. A p value ≤ 0.05 was considered statistically significant.

Results

Source and Secondary Cases in Epidemiologically Linked Case Pairs

Contact investigations in the sentinel sites identified 538 epidemiologically linked case pairs in which a direction of transmission was specified, both cases were culture positive, and fingerprints were available for both isolates (Table 1). These pairs included 397 source cases, of which 324 (82%) were linked to only 1 secondary case. Of the remaining 73 source cases, 48 were linked to 2 secondary cases; the rest were linked to between 3 and 11 secondary cases.

Factors Associated with Transmission Unconfirmed by DNA Fingerprinting

Transmission was not confirmed (source and secondary cases had different fingerprints) in 156 (29%) of the 538 epidemiologically linked case pairs. This proportion was unchanged for the 260 case pairs in which both persons lived in the same household; transmission was unconfirmed in 80 (31%) of these pairs.

Table 1. Characteristics of source and secondary cases in epidemiologically defined tuberculosis case pairs from seven states, 1996–2000

Characteristic	Tuberculosis case pairs (%) (n=538)	
	Source cases (%) (n=397)	Secondary cases (%) (n=538)
Gender		
Men	62	60
Women	38	40
Race/ethnicity		
White	16	16
Black or African-American	51	56
Hispanic	17	15
Asian or Pacific Islander	16	13
Age		
<15 yrs	1	14
15–44 yrs	61	54
44–65 yrs	28	26
>65 yrs	9	7
History of previous tuberculosis	7	3
Foreign-born	35	25
Pulmonary disease	95	91
Negative sputum smear	25	47
Cavitary disease	47	22
Homeless	9	7
Alcoholic	27	22
Injection drug user	6	5
HIV seropositive	11	9

The results of univariate analysis comparing case pairs with unconfirmed transmission to case pairs with confirmed transmission are shown in Table 2. Case pairs with unconfirmed transmission were as likely to be from the same household as those with confirmed transmission, but were more likely to be identified through a shared workplace (odds ratio [OR] 2.3). In univariate analysis, case pairs with unconfirmed transmission were more likely to include a smear-negative source case (OR 2.4), a foreign-born secondary case (OR 4.7), and a case pair in which both persons were Asian or Pacific Islanders (OR 3.5) than case pairs with confirmed transmission. Case pairs with unconfirmed transmission were less likely to include a secondary case <15 yrs of age (OR 0.4) and a case pair in which both persons were black or African-American (OR 0.5) than pairs with confirmed transmission.

All levels of any factor substantially associated (positively or negatively) with the unconfirmed transmission in the univariate analysis were entered into a multiple logistic regression model. The HIV status of secondary cases and cavitary disease in source cases was also included because of potential biologic importance. No interactions were seen, and interaction terms were removed. In the full model without interaction terms, case pairs with unconfirmed transmission were significantly more likely to include a smear-negative source case (OR 2.0) or a foreign-born secondary case (OR 3.4), and significantly less likely to include a secondary case <15 yrs of age (OR 0.3) than case pairs with confirmed transmission. The association for workplace setting seen on univariate analysis was not significant in this model (OR 2.7, p=0.08). We did not find a high level of colinearity between pairs of variables. Confidence intervals and p values for factors significantly associated with unconfirmed transmission in this model are shown in Table 3.

The DNA fingerprints of 34 nonmatching case pairs differed by fewer than three bands. When we removed these case pairs from the analysis, we found the same variables were still substantially associated with unconfirmed transmission, although the ORs differed slightly (results not shown), so these pairs were retained.

Discussion

Our study found that 29% of TB case pairs identified in contact investigations in seven sites during the course of 5 years were not confirmed as linked by using DNA fingerprinting. We used a restricted definition for considering a case pair to be epidemiologically linked: direction of transmission had to be identified as a result of the epidemiologic investigation. Despite the differences in methodology between our research and a California study (9), which used a less restrictive definition (an epidemiologic case pair was an index case and contact case), both studies found a similar percentage of unconfirmed case pairs.

The fact that case pairs in which transmission was unconfirmed were no less likely to be found in the same household than case pairs in which transmission was confirmed suggests

Table 2. Univariate analysis of factors associated with transmission unconfirmed by DNA fingerprinting in tuberculosis case pairs identified in contact investigations in seven sites, 1996–2000

Variable	Transmission unconfirmed (%) (n=156)	Transmission confirmed (%) (n=382)	Odds ratio (95% confidence intervals)	p value[a]
Relationship between persons in case pairs				
Shared household	80 (51)	180 (47)	1.2 (0.8 to 1.7)	NS
Friends/social	55 (35)	156 (41)	0.8 (0.5 to 1.2)	NS
Shared workplace	11 (7)	12 (3)	2.3 (1.1 to 5.4)	0.04
Other	10 (6)	34 (9)	1.0 (0.6 to 1.5)	NS
Characteristics of source cases				
Sputum smear-negative	56 (36)	73 (19)	2.4 (1.6 to 3.6)	<0.0001
Cavitary disease	64 (41)	197 (52)	0.7 (0.5 to 1.0)	NS
Characteristics of secondary cases				
Foreign-born	73 (47)	60 (16)	4.7 (3.1 to 7.2)	<0.0001
Previous tuberculosis	5 (3.0)	12 (3.0)	1.1 (0.4 to 3.7)	NS
<15 yrs	8 (5)	65 (17)	0.4 (0.2 to 0.8)	0.007
≥65 yrs	16 (11)	21 (6)	1.8 (0.9 to 3.6)	NS
HIV positive	13 (8)	33 (9)	1.0 (0.5 to 1.9)	NS
Characteristics of case pairs				
Both white	14 (9)	42 (11)	0.5 (0.4 to 1.5)	NS
Both black	62 (40)	221 (58)	0.5 (0.3 to 0.7)	0.005
Both Hispanic	21 (14)	46 (12)	1.1 (0.7 to 2.0)	NS
Both Asian or Pacific Islander	35 (22)	29 (8)	3.5 (2.1 to 6.0)	<0.0001
Racial discrepancy	24 (15)	40 (11)	1.2 (0.9 to 1.4)	NS
Age difference >10 yrs	88 (56)	198 (52)	1.2 (0.8 to 1.7)	NS

[a]NS, not significant.

that apparent household transmission should not be taken at face value. Marks et al., in a review of U.S. contact investigations (7), report that 33% of contact investigations focused only on the household of the index case. Veen, however, suggests that a modern day stone-in-the-pond approach should focus first on the places and groups frequented by the person classified as the index case, which may not only be the household (4). Onorato notes that as TB in the United States becomes less prevalent, forms and settings of transmission previously considered unusual will appear relatively more common (13). Many public health departments have designed contact investigation formats to focus on places frequented by persons newly diagnosed with TB and the persons with whom they have had contact during the infectious period, rather than focusing only on the household (S. Sharnprapai, pers. comm.). Our findings suggest that this strategy may be useful, even when an apparent source case has been located in the same household.

Unconfirmed transmission between an epidemiologically linked case pair has two interpretations: either the secondary case actually has a reactivation of a latent, remotely acquired TB infection or the transmission was recent, but the source case has not been identified. The progression of a latent, remotely acquired infection to active TB implies that an opportunity to identify infection and apply appropriate preventive measures may have been missed. Failure to correctly identify the source of a recent transmission suggests that contact-tracing procedures may have been inadequate or that an unusual setting or mode of transmission may have been associated with the infection. The significant associations found in our study suggest that either explanation could apply to many of the unconfirmed transmissions we identified.

Few secondary cases in the study reported previous active TB, and prevalence of previous TB was equal among the confirmed and unconfirmed transmission groups. However, previous TB infection could still be an important factor among some foreign-born persons classified as secondary cases in case pairs in which transmission was unconfirmed. Foreign-born persons from countries with high TB prevalence are likely to have had multiple opportunities for exposure to TB before arrival in the United States (3), diminishing the likelihood that any one identified exposure, such as the source case identified in the contact investigation, is the source of transmission. Screening new arrivals from countries with a high prevalence of TB and scheduling additional screening of persons from such countries were both given a high priority in the

Table 3. Logistic regression model[a] showing factors associated with transmission unconfirmed by DNA fingerprinting in tuberculosis case pairs identified in contact investigations in seven sites, 1996–2000

Factor	Adjusted odds ratio (95% confidence intervals)	p value
Smear-negative source case	2.0 (1.2 to 3.1)	0.001
Foreign-born secondary case	3.4 (2.0 to 6.0)	<0.0001
Secondary case <15 yrs	0.3 (0.2 to 0.8)	0.01
Secondary case ≥65 yrs	1.7 (0.8 to 3.6)	0.2
Workplace setting	2.7 (0.9 to 5.7)	0.08

[a]The model also included in the multiple logistic regression analysis all levels of the following factors that were not significantly associated with unconfirmed transmission: cavitary disease in source case, race/ethnicity of case pair, and HIV status of secondary case. In the goodness-of-fit test, full model's p value was 0.885; and the p value for a reduced model containing only the variables with significant associations was 0.725. Thus, both models fit adequately. For the full model, –2 x log likelihood was 564.044; for the reduced model, –2 x log likelihood was 571.469. The difference was distributed as a chi-square variable with 1 degree of freedom. The full model gave a better fit than the reduced model and was retained (p=0.0064, chi square >7.425).

Institute of Medicine's recent recommendations for eliminating TB in the United States (3). The Institute noted that such screening is currently inadequate and warrants expansion, as well as follow-up to ensure that foreign-born persons with positive tuberculin skin tests receive treatment. Much of the active TB among the 133 foreign-born persons classified as secondary cases in this study, as well as some of the TB among the 139 foreign-born persons classified as source cases, might have been prevented if such screening and follow-up were more widely implemented.

Another major finding in this survey is the significant positive association between a smear-negative source case and unconfirmed transmission. This finding suggests that identifying a smear-negative source case for an index case should not preclude ongoing investigation of other possible sources.

On the other hand, the study does not suggest that transmission from smear-negative cases does not occur. Contact investigations of smear-negative cases generally have a low priority in most public health departments. Given that contact investigations were not likely to have been conducted for many smear-negative cases in the network, the fact that 19% of confirmed case pairs had a smear-negative source case suggests that transmission from smear-negative source cases is not negligible. This proportion is similar to that found in a DNA fingerprinting study of two types of epidemiologic clusters, those with smear-negative source cases and those with smear-positive index cases (14). Twenty percent of case pairs with confirmed transmission had a smear-negative index case.

Eighty-nine percent of case pairs with a secondary case <15 years of age were confirmed by genotyping. The shorter incubation of TB in children than in adults and the more limited social circles of children increase the likelihood that children with TB were infected within the circle of present contacts located in a contact investigation. These same factors make TB in children preventable. If TB in the 55 adult source cases who transmitted the infection to these children had been diagnosed and treated promptly and contact tracing had resulted in timely and complete treatment for latent TB infection, latent TB infection in most of these children would have been prevented or would not have progressed to TB.

We hypothesized that secondary cases ≥65 years of age might have had a greater risk for unconfirmed transmission, given longer lives which included the era of relatively high TB rates before the availability of TB drugs. However, case pairs with unconfirmed transmission were no more likely to include an elderly person as a secondary case than pairs with confirmed transmission. TB in an elderly person should not be assumed to be a reactivation, and contact investigations should attempt to identify a possible source case.

Nearly half the workplace case pairs did not have transmission confirmed by DNA fingerprinting, although the number of case pairs was small. When active TB is diagnosed in a person in a workplace, screening is often not confined to close contacts. Casual contacts may be screened because of anxiety on the part of management and staff. Including nonclose contacts in the screening increases the likelihood that if an additional case is found, transmission did not occur from the index case.

Our hypothesis that secondary cases with HIV are more likely to be part of a confirmed case pair than other secondary cases, since their compromised immune status predisposes them to a quicker progression to active TB once infected, was not borne out in this study. Persons with HIV are also more likely to reactivate previous TB infections than persons not dually infected, which may explain the relatively equal proportions of confirmed and unconfirmed transmissions from the identified source among HIV-positive persons. HIV may also be underreported in the TB surveillance system, and the numbers reported may not represent the real numbers of persons with HIV.

Although DNA fingerprinting provided helpful research information in our study, we think the tool should not be considered a routine contact investigation tool. Contact investigations should ideally be undertaken, and often will be completed, before the isolates from the index case are available. Since early identification of TB cases and prevention of further cases is the highest priority, DNA fingerprinting methods will not be relevant to most contact investigations in a TB control program. If TB cases are identified early and appropriate preventive measures taken for latent TB infection seen in contacts, transmission of TB and progression of latent TB infection to TB in infected contacts should be rare. As noted earlier, other studies report a secondary case rate of between <1% and 4% in contact investigations; as programs progress towards TB elimination, this rate may be further reduced. DNA fingerprinting is most likely to be relevant when a case is suspected of being part of a larger outbreak, for which fingerprints could confirm extensive transmission. This survey suggests that DNA fingerprinting could also be used when an apparent source for an index case is sputum smear-negative, to evaluate whether the smear-negative case could be the true source case, or whether further investigation is needed.

Acknowledgments

We thank the participants in the National Tuberculosis Genotyping and Surveillance Network for their concerted efforts in making this study possible.

Dr. Bennett is the coordinator of Antiretroviral Drug Resistance Monitoring, Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention. Her research interests include the application of molecular methods to the epidemiology of HIV and tuberculosis.

References

1. Advisory Council for the Elimination of Tuberculosis. Essential components of a tuberculosis prevention and control program. MMWR Morb Mortal Wkly Rep 1995;44(RR-11):1–18.
2. Advisory Council for the Elimination of TB. Tuberculosis elimination revisited: obstacles, opportunities, and a renewed commitment. MMWR Recomm Rep 2002;48(RR-9):1–13.
3. Institute of Medicine. Ending neglect: the elimination of tuberculosis in the United States. 1st ed. Washington: National Academy Press; 2000.
4. Veen J. Microepidemics of tuberculosis: the stone-in-the-pond theory. Tuber Lung Dis 1992;73:73–6.
5. Villarino M, Dooley S, Geiter L, Castro K, Snider D. Management of persons exposed to multidrug-resistant tuberculosis. MMWR Morb Mortal Wkly Rep 1992;41(RR-11):61–71.
6. MacIntyre CR, Plant AJ. Impact of policy and practice on the effectiveness of contact screening for tuberculosis. Prev Med 1998;27:830–7.
7. Marks S, Taylor Z, Qualls N, Shrestha-Kuwahara R, Wilce M, Nguyen C. Outcomes of contact investigations of infectious tuberculosis patients. Am J Respir Crit Care Med 2000;162:2033–8.
8. Reichler M, Reeves R, Bur S, Thompson V, Mangura B, Onorato I. Evaluation of investigations conducted to detect and prevent transmission of tuberculosis. JAMA 2002;287:991–5.
9. Behr M, Hopewell P, Paz A, Kawamura M, Schecter G, Small P. Predictive value of contact investigation for identifying recent transmission of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1998;158:465–9.
10. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192-6.
11. Ellis BA, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S, et al. Molecular epidemiology of tuberculosis in a sentinel surveilance population. Emerg Infect Dis 2002;8;1197-1209.
12. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.
13. Onorato I. TB outbreaks in the United States. Int J Tuberc Lung Dis 2000;4:5121–6.
14. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leone A, Daley CL, et al. Transmission of *Mycobacterium tuberculosis* from acid-fast bacilli smear-negative patients. Lancet 1999;353:444–9.

Address for correspondence: Diane Bennett, Division of HIV/AIDS Prevention-SE, Centers for Disease Control and Prevention, Mailstop E46, 1600 Clifton Road, Atlanta, GA 30333, USA; fax: 404-639-8959; e-mail: dbennett@cdc.gov

The Ellison Medical Foundation

Senior Scholar Award in Global Infectious Disease
Request for Letters of Intent
Deadline: February 28, 2003

The Ellison Medical Foundation, established by Lawrence J. Ellison, is announcing an opportunity to submit letters of intent for the Senior Scholar Award in Global Infectious Disease (GID). The GID Program, now in its 3rd year, focuses on the most basic research on molecular and cellular mechanisms of disease with special attention afforded to tuberculosis, malaria, and parasitoses, which account for much of the world's morbidity, and are grossly neglected in federally funded research in the United States. **Areas of interest include, but are not limited to, the following:**

- Fundamental studies on exotic microbes and diseases*
- Therapeutic role of probiotics
- Implications of disease eradication
- Comparative immunology
- Threats from newly explored habitats
- Parasite molecular mimicry
- Plasmid and phage determinants of virulence
- New concepts for antivirals and antiparasitic drugs
- Inappropriate use of antibiotics and microbicides
- Our microbiome: natural microflora and pathogen ecology and evolution
- Diet, nutrition and immunity
- Zoonoses: wildlife and human disease
- Host factors, human genomics and disease susceptibility
- Signaling and gene flow between parasites & hosts
- Fever and other symptomatology
- Phylogeny and ultimate origins of viruses
- Nosocomial infection and sanitary precaution

*Model systems and familiar diseases should contain applicability to the global disease burden

Letters of Intent must be received before close of business February 28, 2003 **at The Ellison Medical Foundation address below.** For guidelines on submitting a letter of intent, see **http://www.ellisonfoundation.org**. **Applicants invited to submit a full application will be notified and provided with application forms in June 2003. Up to 10 Senior Scholars will be selected**. Each award will be made for up to $150,000 per year direct cost, with full indirect cost at the institution's NIH-negotiated rate added to that, for up to 4 years.

Contact: Stephanie L. James, Ph.D., Deputy Director
The Ellison Medical Foundation 4710 Bethesda Avenue, Suite 204
Bethesda, MD 20814-5226 Phone: 301-657-1830; Fax: 301-657-1828

Molecular Epidemiology of Multidrug-Resistant Tuberculosis, New York City, 1995–1997

Sonal S. Munsiff,*† Trina Bassoff,* Beth Nivin,* Jiehui Li,* Anu Sharma,* Pablo Bifani,‡ Barun Mathema,‡ Jeffrey Driscoll,§ and Barry N. Kreiswirth‡

From January 1, 1995, to December 31, 1997, we reviewed records of all New York City patients who had multidrug-resistant tuberculosis (MDRTB); we performed insertion sequence (IS) *6110*-based DNA genotyping on the isolates. Secondary genotyping was performed for low IS*6110* copy band strains. Patients with identical DNA pattern strains were considered clustered. From 1995 through 1997, MDRTB was diagnosed in 241 patients; 217 (90%) had no prior treatment history, and 166 (68.9%) were born in the United States or Puerto Rico. Compared with non-MDRTB patients, MDRTB patients were more likely to be born in the United States, have HIV infection, and work in health care. Genotyping results were available for 234 patients; 153 (65.4%) were clustered, 126 (82.3%) of them in eight clusters of $\geq$4 patients. Epidemiologic links were identified for 30 (12.8%) patients; most had been exposed to patients diagnosed before the study period. These strains were likely transmitted in the early 1990s when MDRTB outbreaks and tuberculosis transmission were widespread in New York.

Widespread transmission of multidrug-resistant *Mycobacterium tuberculosis* (MDRTB) strains occurred during the epidemic of the 1980s and early 1990s in New York City. Outbreaks were identified in many New York City hospitals and subsequently in New York State correctional facilities. Many of these outbreaks were associated with one strain (known as the "W" strain of TB) that was resistant to isoniazid, rifampin, ethambutol, and streptomycin and usually to kanamycin (1–5). However, other multidrug-resistant (MDR) strains were associated with outbreaks and nosocomial transmission during these years (6–8). Previous molecular epidemiology surveys in New York City showed that MDRTB was associated with clustered *M. tuberculosis* strains, which suggests recent transmission of the organism (9–11). The incidence of tuberculosis (TB) and MDRTB has been decreasing rapidly in New York City since 1992, when an enhanced Tuberculosis Control Program was implemented. The number of TB cases decreased 21.5% by 1994 (from 3,811 in 1992 to 2,995 in 1994), and MDRTB cases decreased 60% (from 441 to 176) (12,13). Since 1994, no outbreaks of MDRTB have been documented in the city.

To better understand the epidemiology of MDRTB, the New York City Tuberculosis Control Program began DNA genotyping of MDRTB strains from new cases in 1995. The objectives were to provide descriptive molecular epidemiology of MDRTB cases in the city during 1995–1997 and to identify predominant MDR strains present during these years, as well as the extent and risk factors for clustering among these cases.

Methods

Patient Selection

All patients with MDRTB (*M. tuberculosis* isolate resistant to at least isoniazid and rifampin) confirmed as TB cases from January 1, 1995, to December 31, 1997, in New York City were included. Demographic and clinical data were obtained from the New York City Tuberculosis Case Registry. The Registry's data were obtained from patient interviews and medical record reviews at the treatment or residential facilities by trained case managers using standardized data collection instruments and from contact investigations for each pulmonary case.

Susceptibility results were reviewed for the following TB treatment drugs: isoniazid, rifampin, pyrazinamide, ethambutol, streptomycin, and rifabutin (first-line drugs) and fluoroquinolone (usually ciprofloxacin or oflaxacin), kanamycin or amikacin, capreomycin, ethionamide, para-aminosalicylic acid, and cycloserine (second-line drugs). Susceptibility tests were done by Bactec radiometric method (Becton Dickinson and Co., Sparks, NY) for first-line drugs, except rifabutin (14), for most isolates and with standard proportion method with Middlebrook 7H10 media for both first- and second-line drugs for all isolates (15). Most of these tests were conducted at two reference laboratories, the New York City Department of Health and the New York State Department of Health, Wadsworth Center.

*New York City Department of Health, New York, New York, USA; †Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ‡Public Health Research Institute TB Center, Newark, New Jersey, USA; and §Wadsworth Center, Albany, New York, USA

As part of routine surveillance, we reviewed the clinical histories of all pulmonary TB patients who had a negative acid-fast bacilli smear and only one positive *M. tuberculosis* isolate from a respiratory source. This review was to determine the accuracy of the culture result and to rule out laboratory error. If laboratory error was suspected for other types of specimens, clinical and laboratory data for patients were reviewed. Laboratory error was defined as a false-positive *M. tuberculosis* culture result that was caused by specimen mislabeling or laboratory cross-contamination, as evaluated by a described method (10). These patients were not counted as having verified cases of TB and were excluded from the analysis.

Definitions

Patients were defined as having had prior treatment for TB if 1) drug-susceptible *M. tuberculosis* isolates were identified before the drug-resistant isolates that qualified the patients for this study; 2) they had documentation of previous TB disease or treatment; or 3) they had received >30 days of treatment with anti-TB drugs before collection of the specimen that grew MDR *M. tuberculosis*.

Patients were considered HIV seropositive when a positive HIV antibody test result was documented in the medical record or when AIDS was diagnosed before the TB diagnosis. The MDRTB diagnosis date was defined as the collection date of the first specimen from which an MDR *M. tuberculosis* isolate was cultured. Homelessness was defined as being in a public or private shelter or having no address at the time of the MDRTB diagnosis. Information about injection drug use within the 12 months before diagnosis was elicited from direct patient interviews and medical record reviews.

Epidemiologic Investigations

Trained case managers obtained information about suspected and confirmed nosocomial and community exposure from patient interviews, contact investigations, and medical record reviews at the treatment or residential facilities. Probable nosocomial transmission was considered if the newly infected patient was in the same section of an institution as another patient who had an identical *M. tuberculosis* strain and was infectious (i.e., the patient had a positive culture from a respiratory site) at least 30 days before disease onset in the newly infected patient.

Community transmission was considered probable if either of the following occurred: 1) A patient was exposed to another patient who had the identical *M. tuberculosis* strain and was infectious (i.e., had a positive culture from a respiratory site) at least 30 days before disease onset in the subsequent patient. The exposure would have occurred in a home, single-room occupancy hotel, homeless shelter, or another noninstitutional setting. 2) The patient named another patient as a contact whose *M. tuberculosis* isolate had the same DNA pattern or who had MDRTB, but DNA genotyping result was not available.

Transmission could have been from a patient whose condition was diagnosed before the study period. If evidence of nosocomial or community transmission was found, patients had an epidemiologic link. The source patient was not considered to have an epidemiologic link.

During 1995 through 1996, nosocomial transmission was suspected at a hospital where the same MDR strain (i.e., identical insertion sequence [IS] *6110* band patterns) was found in six patients. Hospital floor, ward, and bed information and computerized outpatient clinic records from 1990 to 1996 were analyzed for temporal and spatial overlap among these patients. Medical records were reviewed for patient breaches of isolation protocol during hospitalization. Additional social and demographic information was collected through questionnaires. Specifically, patients were asked with whom and where they spent considerable time, and names of additional social contacts were requested. Patients were asked where and how they thought they had been exposed to TB.

IS*6110* DNA Genotyping and Other Molecular Studies

From 1995 through 1997, one *M. tuberculosis* isolate from each patient with MDRTB in New York City was sent to the Public Health Research Institute Tuberculosis Center, where DNA fingerprint analysis, based on IS*6110* Southern blot hybridization pattern, was performed by using a standardized protocol (16). The Southern hybridization patterns were compared on a Sun Sparc5 Workstation (Sun MicroSystems, Santa Clara, CA), using Bio Image Whole Band Analyzer software version 3.4 (Bio Image, Ann Arbor, MI). Previously described methods were used to classify isolates (17). IS*6110* banding patterns, which were similar to a parent strain but differed by one or two hybridization bands, were denoted by the addition of a number to the cluster letter (e.g., W, W1, P, or P1).

Secondary genotyping was performed by using spacer oligonucleotide typing (spoligotyping) and DNA sequencing of target gene regions that confer drug resistance. Spoligotyping and DNA sequencing of target gene regions used previously described methods (18–21).

If *M. tuberculosis* isolates had identical IS*6110* band patterns, they were considered clustered. However, identical IS*6110* patterns with less than six bands were not considered clustered, unless secondary DNA analysis confirmed a match, as noted in the results.

Data Analysis

To examine how MDR patients differed from non-MDR patients, study subjects were compared to persons who had culture-positive TB diagnosed during the same period but were not included in this study. Descriptive analysis was performed for all study patients according to drug resistance patterns, DNA patterns, prior TB treatment, social and demographic variables, and evidence of nosocomial and community transmission. The Wilcoxon rank-sum test was used to compare medians of continuous variables, and the Pearson chi-square test was used to compare categorical data. Unconditional logistic regression was used to assess crude odds ratios and 95% confidence intervals for the association between

potential risk factors and clustering. Statistical Analysis System Software (Release 8.01, SAS Institute, Inc., Cary, NC) was used for all data analyses. Statistical significance was set at a two-sided 5% level.

Results

From 1995 through 1997, a total of 6,228 cases of TB were confirmed in New York City. Cultures from 5,136 (82.4%) persons were positive for *M. tuberculosis.* Of these, susceptibility results were available for 4,955 (96.5%); 241 (4.9%) persons had MDRTB. Findings of MDR for 11 additional isolates resulted from laboratory error (10 sputum and 1 urogenital); they were excluded from further analyses. The 241 patients made up 4.4% (106 of 2,445), 3.9% (81 of 2,053), and 3.1% (54 of 1,730) of all verified patients who had TB from 1995, 1996, and 1997, respectively. Table 1 presents a comparison of the demographic characteristics of these patients to those of culture-positive non-MDRTB patients from the same time period in New York City for whom drug susceptibility results were available. Compared with patients with culture-positive non-MDRTB during the same period, MDR patients were more likely to be born in the United States, have HIV infection, and be health-care workers, homeless, and injection drug users. MDR patients were more likely to have respiratory specimens positive for acid-fast bacilli and were less likely to be Asian. By further stratification, none of Asian MDRTB patients were born in the United States, and 68.7% of U.S.-born MDRTB patients were HIV infected.

Strains were resistant to a median number of 6 drugs (range 2–10). Eight (3.3%) patients had strains of *M. tuberculosis* that were resistant to isoniazid and rifampin only, and 146 (60.6%) had isolates that were also resistant to one or more second-line anti-TB drugs. Most of these strains were also resistant to rifabutin. Twenty-four (10%) patients had received prior treatment for TB. Compared with patients who had no prior treatment, patients who had received such treatment were significantly older (median age 46 years vs. 41 years, p=0.010) and had less drug resistance (median 5 drugs versus 6, p=0.042). Patients with prior treatment were less likely to be born in the United States (45.8% vs. 71.4%, p=0.001) and were less likely to be HIV infected (33.3% vs. 55.3%, p=0.041). Patients who had received prior treatment did not differ from those who had no prior treatment according to gender, race or ethnicity, occupation, and histories of alcohol or drug abuse and homelessness.

DNA Genotyping Analysis

Of 241 MDR patients, 234 (97%) had IS*6110* fingerprint patterns. Ninety-two different patterns were identified (band range 2–22). Thirty-six (15.4%) of 234 isolates had patterns with five or fewer IS*6110* bands. Five were in one cluster, the C strain, and all had the same spoligotype (700036777760731). Two were clustered as a four-band strain with the same spoligotypes, and three other strains had unique genotypes. Twenty-six strains had an identical two-band IS*6110* pattern designated as H; 25 of the 26 were resistant to pyrazinamide. All 17 with available results had identical spoligotypes (777776777760601); 18 of the 20 strains that were tested had identical *pncA* genotype (Nt70; G deletion). One pyrazinamide-susceptible strain had the wild-type *pncA* genotype, and one resistant strain had a different *pncA* genotype (139; GTG>GCG, Val>Ala). On the basis of the results, 18 of the 26 H strains were considered clustered. Thus, 25 of the 36 isolates with low IS*6110* copies were considered clustered.

Of 234 patients with DNA results, 153 (65.4%) were grouped into 19 clusters: 6 clusters with 2 cases each; 5 clusters with 3 cases each; and 8 different clusters with 4, 5, 6, 7, 13, 14, 18, and 59 cases each. The eight clusters had 126 (52.2%) of 241 MDRTB patients from the study period. Table 2 shows the distribution of these eight strains during 1995

Table 1. Comparison of social and demographic characteristics of multidrug-resistant (MDR) tuberculosis patients and non-multidrug-resistant tuberculosis patients, New York City, 1995–1997

	MDR (n=241)		Non-MDR (n=4,714)		
Characteristic	No.	%	No.	%	p value
Yr of diagnosis					
1995	106	44.0	1,816	38.5	0.0898
1996	81	33.6	1,586	33.6	0.9912
1997	54	22.4	1,312	27.8	0.0660
Median age (range), yrs	41 (5–85)		41 (0–100)		0.3071
Male sex	147	61.0	3,020	64.1	0.3333
U.S.-born	166	68.9	2,483	52.7	<0.001
HIV serostatus					
Positive	128	53.1	1,478	31.4	<0.001
Negative	75	31.2	1,759	37.3	0.0521
Unknown	38	15.8	1,477	31.3	<0.001
Race/ethnicity					
Asian	24	10.0	817	17.3	0.0029
Hispanic	75	31.1	1,266	26.9	0.1461
Black non-Hispanic	105	43.6	2,089	44.3	0.8200
White non-Hispanic	37	15.4	542	11.5	0.0692
Health-care worker	15	6.2	109	2.3	0.0001
Homeless	28	11.6	344	7.3	0.0130
Injection drug user	33	13.7	275	5.8	<0.001
Disease site					
Pulmonary only	176	73.0	3,397	72.1	0.7440
Extra-pulmonary only	35	14.5	894	19.0	0.0848
Pulmonary + extra-pulmonary	30	12.5	423	9.0	0.0679
Specimen AFB smear-positive[a]	141	68.5	2,129	55.7	0.0003
Cavitary lesion(s)[b]	47	22.8	740	19.4	0.2247

[a]Respiratory specimen during the 30 days after initial specimen for bacteriologic test was taken. Excludes those who had only extrapulmonary disease.
[b]Excludes those with extrapulmonary disease only.

Table 2. Social and demographic characteristics of patients in predominantly multidrug-resistant tuberculosis clusters, New York City, 1995–1997

	Clustered strain (n=234)									Unique
Characteristics	W	W1	H	AB	P	AU	C	P1	Other	RFLP[a]
No. of patients	59	7	18	14	13	6	5	4	27	81
No. of bands	18	19	2	11	11	10	3	11	4–18	2–22
Known epidemiologic links										
Nosocomial	6	1	0	0	0	0	0	0	0	0
Community	4	1	2	7	1	2	0	0	1	5
Age (median, in yrs)	41	41	37	42	43	40	37	37	41	42
Male	41	3	7	7	7	4	4	3	17	49
Race/ethnicity										
Asian	2	1	1	0	1	0	1	0	0	16
Hispanic	18	2	10	1	3	1	3	2	9	24
Black, non-Hispanic	25	1	6	12	9	1	1	2	15	31
White, non-Hispanic	14	3	1	1	0	4	0	0	3	10
U.S.-born	46	4	15	13	10	5	3	3	21	42
HIV positive	42	1	14	10	4	4	3	1	17	29
History of—										
Homelessness	6	0	1	8	3	0	2	0	2	5
Alcohol abuse	11	1	1	3	4	0	0	0	5	9
Injection drug use	9	0	2	6	0	1	0	1	5	8
Prior tuberculosis treatment	0	0	1	0	1	1	2	0	5	12
Health-care worker	4	3	1	0	1	0	1	0	2	2
Borough of residence										
Manhattan	23	1	8	4	1	1	3	0	10	19
Bronx	17	0	4	0	0	0	1	0	2	8
Brooklyn	7	2	5	10	9	5	1	4	11	30
Queens	12	4	1	0	3	0	0	0	4	23
Staten Island	0	0	0	0	0	0	0	0	0	1

[a]RFLP, restriction fragment length polymorphism.

through 1997 with social, demographic, and epidemiologic link information. Figure 1 shows the geographic distribution, and Figure 2 shows the IS*6110* patterns of these eight strains.

Epidemiologic links were identified for 30 (12.8%) of the 234 patients with genotyping results; most had been exposed to patients diagnosed before the study period. Twenty-five (19.8%) of 153 patients clustered by DNA genotyping were epidemiologically linked; 18 (72%) had probable community transmission, and 7 (28%) had probable nosocomial transmission. All nosocomial links were to patients whose conditions were diagnosed before the study period. Seven community transmission links were to patients from the study period, and 11 were to patients whose diseases were diagnosed before the study period. Epidemiologic links of community transmission were identified for 5 (6.2%) of 81 nonclustered patients; all were links to persons whose conditions were diagnosed before the study period. Of the 23 community links, 3 were to household members, 4 to nonhousehold relatives, and 7 to friends. One was linked to another case in a single-room occupancy hotel; seven were linked in a crack den, and one had an unknown exposure site.

Table 3 shows a comparison of patients clustered by DNA analysis to those nonclustered according to various demographic and clinical characteristics. Factors significantly associated with clustering were HIV infection and birth in the United States. There was no difference in proportion clustered by year. Patients with histories of prior treatment and Asian patients were significantly less likely (odds ratio [OR] = 0.40, 95% confidence interval [CI] = 0.17 to 0.98 and OR=0.18, 95% CI-0.06 to 0.53, respectively) to be in a cluster. Patients in clusters were 3 times more likely to have epidemiologic links than those not in clusters. In a subanalysis that included only non–U.S.-born patients who had a known date of entry to the United States, clustering was significantly associated

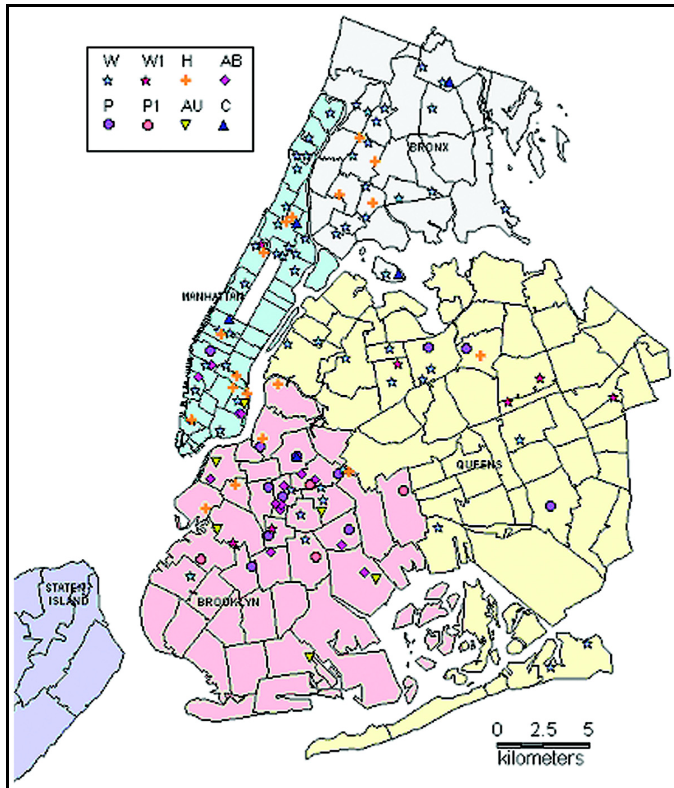

Figure 1. Geographic distribution of patients in major multidrug-resistant tuberculosis clusters, New York City, 1995–1997.

(OR=1.09, 95% CI=1.02, 1.16; p=0.01) with longer time of residence in the United States.

Epidemiology of Predominant MDR Strains

Fourteen patients in this study had an 11-band strain (AB). Six of these patients were diagnosed at a single medical facility in Brooklyn, New York. At the time of diagnosis, five of these persons reported a home address in the same health district as the medical facility. Although two patients were hospitalized at the medical facility when transmission could have occurred, hospital inpatient and outpatient records showed that nosocomial transmission was unlikely because of the room locations and documented adherence to isolation protocol.

Our study showed the following characteristics for patients in the AB cluster: 92.9% were born in the United States, 71.4% were infected with HIV, 85.7% were non-Hispanic black, 42.8% used injection drugs, and 100% had no prior treatment for TB. These patients reported home addresses from only two of five boroughs in New York City, 10 (71.4%) in Brooklyn and 4 (28.6%) in Manhattan. However, 57% were homeless. Five patients agreed to additional interviews; six patients had died, and three patients could not be located. On the basis of the additional interviews and available data from initial interviews, 7 of these 14 patients had community transmission links. Two of these links were found through standard contact investigations, and five were disclosed by the additional patient interviews. Three patients had close contacts with two patients who had the AB strain in 1992; four frequented the same crack den in the neighborhood of the medical facility before their TB diagnosis. The remaining seven patients had no history of contact with persons who had the AB strain.

The largest cluster was from the W strain—59 patients representing almost 25% of the 241 MDRTB patients in the 3 years. This strain caused a well-documented multi-institutional outbreak in New York City from 1990 through 1993 (1–5). Strain W1, which was isolated in seven patients, is a variant of the W strain. It has an additional IS*6110* copy and is part of the W strain outbreak (4,5). Forty percent (12 of 30) of the epidemiologic links in this cohort were to patients with these two strains. Seven (46.7%) of the 15 health-care workers had either the W strain (4 cases) or the W1 strain (3 cases). However, epi-

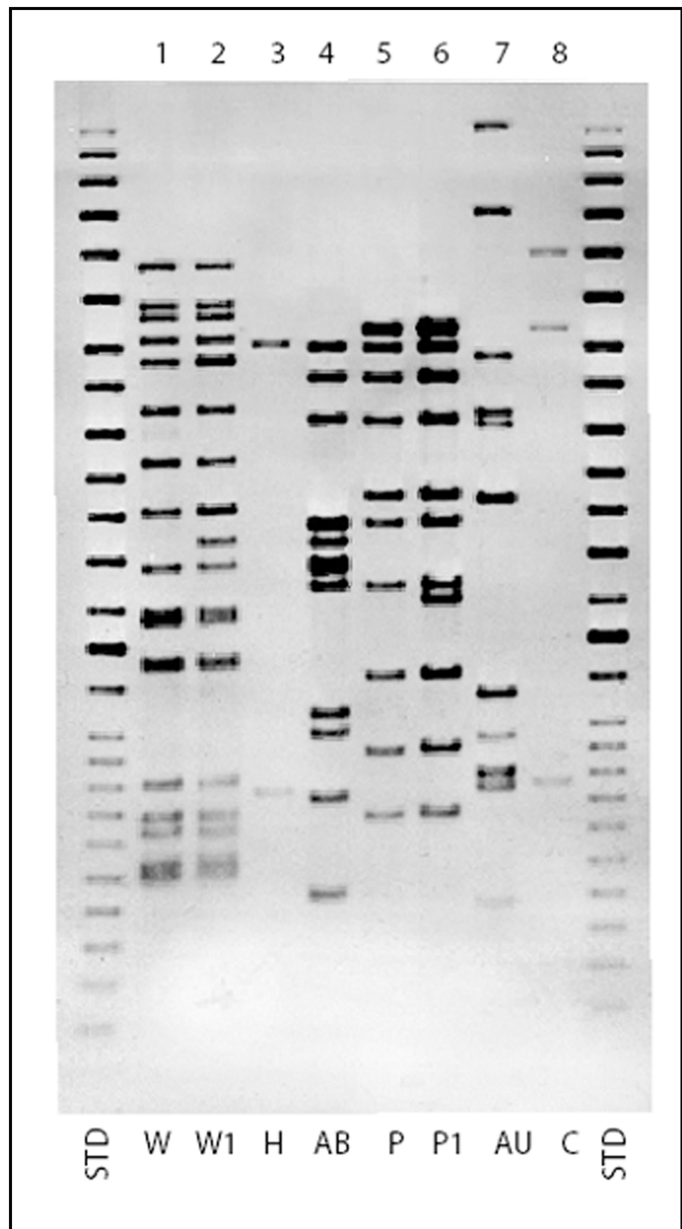

Figure 2. Insertion sequence (IS) *6110* Southern blot hybridization patterns for major multidrug-resistant *Mycobacterium tuberculosis* strains, New York City, 1995–1997. STD, standard.

Table 3. Risk factors associated with clustering of multidrug-resistant tuberculosis cases, New York City, 1995–1997 (n=234)[a]

	Clustered (n=153)	Nonclustered (n=81)	Crude	
Characteristic	No. (%)	No. (%)	OR	95% CI
Median age in yr (range)	41 (5–85)	42 (22–77)	0.99	0.98, 1.02
Male sex	93 (60.8)	49 (60.5)	1.01	0.58, 1.76
U.S.-born	120 (79.0)	42 (51.9)	3.48	1.94, 6.25
Median years of residence in United States[b]	12 (0–47)	6.5 (0–24)	1.09	1.02, 1.16
HIV serostatus				
Positive	96 (62.8)	29 (35.8)	2.81	1.52, 5.22
Negative	40 (26.1)	34 (42.0)	1.00	
Unknown	17 (11.1)	18 (22.2)	0.80	0.36, 1.80
Race/ethnicity				
Asian	6 (3.9)	16 (19.8)	0.18	0.06, 0.53
Hispanic	49 (32.0)	24 (29.6)	1.00	
Black non-Hispanic	72 (47.1)	31 (38.3)	1.14	0.60, 2.17
White non-Hispanic	26 (17.0)	10 (12.4)	1.27	0.53, 3.06
Health-care worker	12 (7.8)	2 (2.5)	3.36	0.73, 15.40
Homeless	22 (14.4)	5 (6.2)	2.55	0.92, 7.02
Injection drug use[c]	24 (15.7)	8 (9.9)	1.70	0.73, 3.97
Prior treatment history	10 (6.5)	12 (14.8)	0.40	0.17, 0.98
Having epidemiologic link[d]				
Nosocomial	7 (4.6)	0 (0)	2.97	1.02, 9.26
Community	18 (11.8)	5 (11.1)		
No link	128 (83.7)	76 (88.9)	1.00	
Year of diagnosis				
1995	69 (45.1)	32 (39.5)	1.00	
1996	54 (35.3)	27 (33.3)	0.93	0.47, 1.81
1997	30 (19.6)	22 (27.2)	0.63	0.30, 1.34

[a]OR, odds ratio; CI, confidence interval.
[b]Excludes non-U.S.-born patients.
[c]Injection drug use within 12 months before diagnosis.
[d]Compared epidemiologic link with no epidemiologic link.

demiologic links for nosocomial transmission were found for only two of the seven. Patients with this strain were identified from four of the city's five boroughs. The epidemiology of these clusters has been described in greater detail after the institutional outbreaks (22).

The only difference between the P and P1 strains is that the P1 strain has an additional band. Both strains have been nosocomially transmitted in one institution in New York City (7). Nine of the 13 patients with the P strain and all 4 with the P1 strain were living in the same borough as the institution where this outbreak was identified. However, epidemiologic links were identified for only one patient. Patients in these clusters were much less likely to be HIV infected than the other clustered patients (29% vs. 67%, p=0.002).

The H strain, the other major cluster, was also associated with a nosocomial outbreak in an institution in New York City (8). During the study period, patients with this strain were identified from all the city's boroughs. Two patients with this strain had epidemiologic links.

Discussion

During the 3-year period, 241 (3.9%) of all 6,228 TB cases in New York City and 241 (4.9%) of all 4,995 *M. tuberculosis* culture-positive patients with susceptibility had MDR strains. MDRTB patients were more likely to have acid-fast bacilli visible on microscopic examination of respiratory specimens and thus were more infectious. MDRTB was more common in patients who were born in the United States, HIV infected, non-Asian, or health-care workers. The finding of greater prevalence of HIV infection in MDRTB patients compared with non-MDRTB patients is likely due to several reasons. The initial outbreaks during which these strains were transmit-

ted mostly involved HIV-infected persons (1–8). A large number of HIV-infected patients were likely infected in those outbreaks. HIV-infected patients progress from infection to disease at a much higher rate than non–HIV-infected persons.

Most patients in this study had primary MDRTB caused by a few strains. The proportion of patients clustered in this cohort is much higher than in previously reported New York City patients. Few demographic and clinical characteristics were associated with clustering. According to previous citywide surveys of all patients who had cultures that were positive for TB, the proportions of clustered patients who had TB were 37% in 1991 and 32.4% in 1994 (10,11). Another investigation from one hospital in the city found similar results for patients during 1989–1991 (9). A more recent survey from 1997 in New York City found that, for persons born in the United States, the proportion clustered had not decreased (23). In all these surveys, MDRTB was associated with clustering in multivariate analyses. The higher proportion of MDRTB clustering seen in this study cannot be explained by the exclusion of low-band patients in previous citywide surveys. When we exclude low-band patients from our analysis, we still have a similar proportion of clustering (128 [64.6%] of 198 isolates with less than five bands). Our proportion of MDRTB clustering is also higher than that reported from other U.S. cities and other industrialized countries, where approximately 18% to 49% of clustering has been observed (24–28). However, few patients in these studies had MDRTB. The inclusion of MDRTB patients only in this study may have contributed to this difference. A study conducted during 1995 to 1997 by Moro et al. in Italy showed 74.2% clustering among MDRTB patients, compared with 39.3% among non-MDR cases (29). Our study reiterates that a few, highly resistant strains were transmitted widely in New York City during the late 1980s and early 1990s.

Strains W, W1, P, P1, and H were transmitted in the early 1990s during the period of MDRTB outbreaks in New York City because five of the eight major clustered strains were associated with hospital outbreaks during that time (1–8). Few patients in this cohort had epidemiologic links, but most of these links were to patients whose diseases were diagnosed before the study period. Most health-care workers (10 of 14 with DNA results) had one of the known outbreak strains, but only 2 could be linked to facilities where nosocomial transmission occurred.

In addition to the nosocomially transmitted strains, we identified a large cluster that may have been transmitted in a community of persons who were HIV infected, homeless, and drug users. Before this study period, at least 14 additional MDRTB patients with this strain had been identified and confirmed by genotyping from 1989 through 1994. Six of these patients were from the same borough, and four were from the same health district as many of the patients in 1995–1997. This strain was transmitted over many years among drug users who were frequenting crack dens in the same neighborhood. Since many of these venues were closed in the late 1990s, this social group was disrupted, and transmission was interrupted. The AB strain has been found in only two new patients during 1998–2001, one in 1998 and the other in 2001. The patient from 2001 had epidemiologic links to a patient from 1995.

Five patients had the C strain, which has three IS*6110* copies. This *M. tuberculosis* strain is the most common in the city. Most of the C strains in the city share the same spoligotype and pTBN12-based RFLP pattern and are clonal (30, New York City Department of Health and Public Health Research Institute, unpub. data). Most of the C strains have been drug-susceptible; however, we identified C strains with varying drug-resistant patterns, occasionally in clusters (30, New York City Department of Health and Public Health Research Institute, unpub. data). The MDR strains in this period appear to be a recent cluster, or each may have acquired drug resistance separately.

MDRTB continues to decline in New York City at a rapid rate, with only 38, 31, and 25 new cases identified in 1998, 1999, and 2000, respectively (31). However, most of the major strains found in this investigation continued to be identified in new MDRTB patients in New York City from 1998 through 2001 (New York City Department of Health and Public Health Research Institute, unpub. data). Most nonclustered patients had primary drug-resistant TB. The improved Tuberculosis Control Program, which was implemented in 1992 with aggressive case management and direct observation of anti-TB therapy for most patients, quickly curtailed the development of newly acquired drug resistance. Since primary and acquired drug resistance and MDRTB, in particular, were prevalent before 1995 (32,33), many MDRTB strains likely were disseminated in the community because most patients in this cohort with unique strains had no histories of prior treatment.

In this study, we may have underestimated the number of cases that had nosocomial and community epidemiologic links. We did not use medical record reviews of hospitalizations before the diagnosis of MDRTB for all the patients to identify potential nosocomial exposures. Many patients died before identification of MDRTB; therefore, interviews could not be conducted to identify potential nosocomial and community exposures before diagnosis of TB. The outbreaks associated with the W and W1 strains were well investigated and publicized, and staff were aware of the locations of the outbreak hospitals. This fact may have allowed for easier identification of epidemiologic links in these patients. In the AB community outbreak cluster, most epidemiologic links were identified from the detailed interviews with the few patients who were still alive. Traditional contact investigations did not identify these links in this subpopulation. This observation underscores that other methods, such as ongoing surveillance for unusual patterns of disease and unusual patient characteristics, should also be used to identify possible transmission in the community. Prospective DNA typing of all isolates can also supplement traditional contact investigation methods.

The molecular analysis of the MDRTB strains in New York City during these years demonstrated that the improved Tuberculosis Control Program has reduced dramatically the transmission of these strains. These investigations have also established important baseline data for the study of the epidemiology of MDRTB over the next decades.

Acknowledgments

We acknowledge the work of the following Multi-Drug Resistant Tuberculosis Coordinators in the Tuberculosis Control Program for the case management and epidemiologic assessment of patients: Tripti Bhattacharjee, Sharif Choudhury, Anatole Hounnou, Cliff Johnson, Angel La Paz, Dileep Sarecha, and Iris Winter. In addition, we thank Cindy Driver, Paula Fujiwara, and Thomas Frieden for their thoughtful review of the manuscript.

The Public Health Research Institute TB Center and the Wadsworth Center received funds for this project from the Centers for Disease Control and Prevention, National Tuberculosis Genotyping and Surveillance Network.

Dr. Munsiff has been the director of the New York City Tuberculosis (TB) Control Program since December 2000, and she has been a medical officer in the Division of TB Elimination, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention since November 2001. Her research interests include the epidemiology and clinical aspects of TB, particularly as manifested in HIV-infected persons, epidemiology and treatment of drug-resistant TB, and program evaluation.

References

1. Centers for Disease Control and Prevention. Nosocomial transmission of multidrug-resistant tuberculosis among HIV-infected persons—Florida and New York, 1988-1991. MMWR Morb Mortal Wkly Rep 1991;40:589–91.
2. Valway SE, Richards SB, Kovacovich J, Greifinger RB, Crawford JT, Dooley SW. Outbreak of multidrug-resistant tuberculosis in a New York State Prison. Am J Epidemiol 1994;140:113–22.
3. Coronado VG, Beck-Sague CM, Hutton MD, Davis BJ, Nicholas P, Villareal C, et al. Transmission of multidrug-resistant *Mycobacterium tuberculosis* among persons with human immunodeficiency virus infection in an urban hospital: epidemiological and restriction fragment length polymorphism analysis. J Infect Dis 1993;168:1052–5.
4. Frieden TR, Sherman LF, Maw KL, Fujiwara PI, Crawford JT, Nivin B, et al. A multi-institutional outbreak of highly drug resistant tuberculosis. JAMA 1996;276:1229–35.
5. Nivin B, Nicholas P, Gayer M, Frieden TR, Fujiwara P. A continuing outbreak of multidrug-resistant tuberculosis, with transmission in a hospital nursery. Clin Infect Dis 1998;26:303–7.
6. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, et al. Exogenous reinfection with multidrug-resistant *Mycobaterium tuberculosis* in patients with advanced HIV infection. N Engl J Med 1993;328:1137–44.
7. Shafer RW, Small PM, Larkin C, Singh SP, Kelly P, Sierra MF, et al. Temporal trends and transmission patterns during the emergence of multidrug-resistant tuberculosis in New York City: a molecular epidemiological assessment. J Infect Dis 1995;171:170–6.
8. Pearson ML, Jereb JA, Frieden TR, Crawford JT, Davis BJ, Dooley SW, et al. Nosocomial transmission of multidrug-resistant *Mycobacterium tuberculosis*: a risk to patients and health-care workers. Ann Intern Med 1992;117:191–6.
9. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bloom BR, et al. Transmission of tuberculosis in New York City—an analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–6.
10. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996;77:407–13.
11. Sachdev PS, Fujiwara PI, Cook S, Kreiswirth BN, Frieden TR. Epidemiology of tuberculosis in NYC, April 1994: Determinants of recent transmission. In: Abstracts of the American Thoracic Society: San Francisco, California; 1997 May 16–21; Abstract A224. New York, New York: American Thoracic Society, 1997.
12. New York City Tuberculosis Control Program. 1994 Annual Summary. New York: New York City Department of Health;1995.
13. Frieden TR, Fujiwara PI, Washko RM, Hamburg MA. Tuberculosis in New York City—turning the tide. N Engl J Med 1995;333:229–33.
14. Hawkins JE, Wallace RJ Jr, Brown BA. Antibacterial susceptibility tests: mycobacteria. In: Balows A, Hausler WJ Jr, Herrmann K, Isenberg HD, Shadomy HJ, editors. Manual of clinical microbiology. 5th ed. Washington: American Society for Microbiology; 1991. p. 1138–52.
15. Kent PT, Kubica GP. Public health mycobacteriology: A guide for the level III laboratory. Atlanta: Department of Health and Human Services; 1985.
16. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenbach KD, Gicquel P, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
17. Kreiswirth BN, Moss A. Genotyping multidrug-resistant *M. tuberculosis* in New York City. In: Rom WN, Garay SM, editors. Tuberculosis. Boston: Little, Brown, and Company, Inc; 1996. p.199–209.
18. Groenen PM, Bunschchoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain identification by a novel typing method. Mol Microbiol 1993;10:1057–65.
19. Goyal M, Saunders MA, van Embden JDA, Young DB, Shaw RJ. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J Clin Microbiol 1997;35:647–51.
20. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 Update. Tuber Lung Dis 1998;79:3–29.
21. Dale JW, Brittan D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, et el. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendation for standardized nomenclature. Int J Tuberc Lung Dis 2001;5:216–19.
22. Munsiff SS, Nivin B, Sacajiu G, Bassoff T, Mathema B, Bifani P, et el. Evolution of a highly drug resistant strain of tuberculosis in New York City. In: Abstracts of the American Society of Microbiology Tuberculosis: past, present and future; New York, New York, 2000 June 20–24; Abstract 90. Washington: American Society of Microbiology; 2000.
23. Sachdev PS, Bassoff T, Kreiswirth B, Cook S, Munsiff SS, Fujiwara PI. Molecular epidemiology of tuberculosis in NYC: an ongoing survey, April 1997. Resistant strain of tuberculosis in New York City. In: Abstracts of the American Thoracic Society: San Diego, California; 1999 April 23–28; Abstract A904. New York: American Thoracic Society; 1999.
24. Bauer J, Yang Z, Poulsen S, Andersen AB. Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of *M. tuberculosis* infection. J Clin Microbiol 1998;36:305–8.
25. Dahle UR, Sandven P, Heldal E, Caugant DA. Molecular epidemiology of *Mycobacterium tuberculosis* in Norway. J Clin Microbiol 2001;39:1802–7.
26. Gutierrez MC, Vincent V, Aubert D, Bizet J, Gaillot O, Lebrun L, et al. Molecular fingerprinting of *Mycobacterium tuberculosis* and risk factors for tuberculosis transmission in Paris, France, and surrounding area. J Clin Microbiol 1998;36:486–92.

27. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991-1997. Ann Intern Med 1999;130:971–8.
28. van Soolingen D, Borgdorff MW, de Haas PEW, Sebek MMGG, Veen J, Dessens M, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993-1997. J Infect Dis 1999;180:726–36.
29. Moro ML, Salamina G, Gori A, Penati V, Sacchetti R, Mezzetti F, et al. Two-year population-based molecular epidemiological study of tuberculosis transmission in the Metropolitan area of Milan, Italy. Eur J Clin Microbiol Infect Dis 2002;21:114–22.
30. Friedman CR, Quinn GC, Kreiswirth BN, Perlman DC, Berger J, Riley LW, et al. Widespread dissemination of a drug-susceptible strain of *Mycobacterium tuberculosis*. J Infect Dis 1997;176:478–84.
31. New York City Tuberculosis Control Program. 2000 annual summary. New York: New York City Department of Health; 2001.
32. Frieden TR, Sterling T, Pablos-Mendez A, Kilburn JO, Cauthen GM, Dooley SW, et al. The emergence of drug-resistant tuberculosis in New York City. N Engl J Med 1993;328:521–6.
33. Fujiwara PI, Cook SV, Rutherford CM, Crawford JT, Glickman SE, Kreiswirth BN, et al. A continuing survey of drug-resistant tuberculosis, New York City, April 1994. Arch Intern Med 1997;157:531–6.

Address for correspondence: Sonal S. Munsiff, New York City Department of Health and Mental Hygiene, 125 Worth St., Room 216, CN74, New York, NY 10013, USA; fax: 212-788-9836; e-mail: smunsiff@health.nyc.gov

Eleanor Roosevelt died of tuberculosis November 7, 1962

Genotyping Analyses of Tuberculosis Cases in U.S.- and Foreign-Born Massachusetts Residents

Sharon Sharnprapai,* Ann C. Miller,* Robert Suruki,* Edward Corkren,* Sue Etkind,* Jeffrey Driscoll,† Michael McGarry,† and Edward Nardell*‡

We used molecular genotyping to further understand the epidemiology and transmission patterns of tuberculosis (TB) in Massachusetts. The study population included 983 TB patients whose cases were verified by the Massachusetts Department of Public Health between July 1, 1996, and December 31, 2000, and for whom genotyping results and information on country of origin were available. Two hundred seventy-two (28%) of TB patients were in genetic clusters, and isolates from U.S-born were twice as likely to cluster as those of foreign-born (odds ratio [OR] 2.29, 95% confidence interval [CI] 1.69 to 3.12). Our results suggest that restriction fragment length polymorphism analysis has limited capacity to differentiate TB strains when the isolate contains six or fewer copies of IS*6110*, even with spoligotyping. Clusters of TB patients with more than six copies of IS*6110* were more likely to have epidemiologic connections than were clusters of TB patients with isolates with few copies of IS*6110* (OR 8.01, 95%; CI 3.45 to 18.93).

The incidence of tuberculosis (TB) in the United States is closely linked to the global TB epidemic (1). In 2000, 46% of all reported TB cases in the United States occurred among persons not born in the United States (foreign-born), and 20 states reported that >50% of TB cases occurred among the foreign-born (2). In Massachusetts, 202 (71%) of 285 cases reported were among foreign-born persons (from 41 different countries). Being born outside the United States is the primary risk factor for being reported with TB in Massachusetts (3).

The distribution of places of birth among TB patients reported in Massachusetts has changed greatly over the past 3 decades, reflecting changes in populations immigrating to Massachusetts. As late as 1970, 80% of foreign immigrants in Massachusetts were from Europe or Canada; only 5% of the immigrants were from Asia, and less than 3% were from Central and South America combined and Africa (4). Since 1970, the proportion of immigrants to Massachusetts from Europe has declined, and the proportion of those from Asia, the Caribbean Islands, Africa, and South and Central America has risen. Immigrants from Asia increased sharply, from 3% to 16%. Between 1996 and 2000, the proportion of foreign-born TB patients reported in Massachusetts rose from 61% to72%. This increase was seen primarily among Asians, Africans, and immigrants from Central and South America.

Understanding the factors that contribute to the incidence of TB is critical for TB elimination. Molecular fingerprinting data can be used to further an understanding of the epidemiology and transmission patterns of TB. In this article, we describe the epidemiology of TB patients in Massachusetts and results of using genotyping to evaluate the extent to which genetic clustering of *Mycobacterium tuberculosis* differs between foreign-born and U.S.-born TB patients.

*Massachusetts Department of Public Health, Boston, Massachusetts, USA; †New York State Department of Health, Wadsworth Center, Albany, New York, USA; and ‡Harvard Medical School, Boston, Massachusetts, USA

Methods

In 1996, the Massachusetts Department of Public Health, Division of Tuberculosis Prevention and Control (TB Division) became part of the Centers for Disease Control and Prevention (CDC)'s National Tuberculosis Genotyping and Surveillance Network. The TB Division attempted to locate and submit at least one isolate for every culture-confirmed TB case-patient reported from July 1, 1996, through December 31, 2000, to the Northeast Regional Genotyping Laboratory, New York State Department of Health, Wadsworth Center, Albany, New York. DNA genotyping by using IS*6110* restriction fragment length polymorphism (RFLP) and the spoligotyping technique (spacer oligotyping) was performed by the Wadsworth Center. RFLP analysis was performed by using the standard method (5,6) with the molecular weight standards provided by CDC. Spoligotyping was performed with a commercially available kit, in accordance with the manufacturer's instructions (Isogen Bioscience BV, Maarseen, the Netherlands).

Specimen Collection for DNA Fingerprinting Analysis

The following procedures were used to identify patients with positive *Mycobacterium tuberculosis* cultures and obtain isolates for RFLP analysis. In 1996, a survey of hospitals and private physicians was conducted to ascertain where specimens were being sent for mycobacterial culture. This survey allowed the TB Division to determine which laboratories

inside and outside of the state were processing clinical specimens for Massachusetts residents. In addition, a letter was sent to directors of all laboratories in Massachusetts that are licensed under the Clinical Laboratory Improvement Act (CLIA) to perform mycobacteriology services and to other laboratories that were identified through the survey, asking for their cooperation with the TB genotyping network project. Most (71%) hospitals and physicians sent specimens to the Massachusetts State Laboratory Institute (MSLI) for culture identification, susceptibility testing, or both. The TB Division and the Mycobacteriology Laboratory, MSLI, share a joint database where all bacteriology reports, including drug susceptibility information, are automatically linked to suspected and confirmed cases of TB. For *M. tuberculosis* specimens that were processed elsewhere, the epidemiologists on the TB genotyping network project identified laboratories by attending routine TB case and cohort reviews conducted monthly by the state TB nurses and the Boston Public Health Commission TB Program. Laboratories were then contacted and arrangements were made for shipment of specimens to the MSLI and the Wadsworth Center.

Cluster Investigation

RFLP analysis by using IS*6110* is a powerful tool for discerning one strain of *M. tuberculosis* from another when there are many copies of IS*6110*. However, for strains of *M. tuberculosis* with low copy numbers of IS*6110*, RFLP analysis has less discriminating power, and therefore a secondary typing method is used to help differentiate strains (7,8). For the TB genotyping network project, isolates were considered to be clonally related (i.e., were the same strain of TB) if they had identical IS6*110* patterns containing seven or more bands or they had identical IS6*110* patterns containing six or fewer bands with identical spoligotyping. A cluster was defined as containing two or more patients with clonally related TB strains.

In 1998, CDC funded the Cluster Investigation Study to evaluate epidemiologic links between clustered cases in a more formal manner. Cluster investigations consisted of standardized medical record reviews wherever a patient was seen for TB, and standardized interviews with the patient (or a proxy) if the patient was unable to participate. All patients were eligible for interview, unless strong epidemiologic links were found between all members of the cluster. In that situation, interviews were considered unnecessary. Written informed consent was obtained from all subjects, and interpreters were used as needed. Information collected through medical record reviews and patient interviews included the estimated period of infectivity, demographics, employment history, and social connections and activities during the 2 years before diagnosis. Each patient in a genetic cluster was examined to determine the following: 1) the period of infectivity (by reviewing date of diagnosis, disease type, smear status, chest radiology results, and date treatment started), 2) name of contacts identified, and 3) how and where the patient spent his or her time during the period of infectivity. If a patient identified another patient in the same cluster, or if patients were found to be in the same place at the same time when one was infectious, the likelihood of transmission was classified as "definite." Transmission was "possible" if patients were thought to be at the same place, at the same time up to 2 years before diagnosis, or if patients identified the same contact as being the source of TB. A final category, "unlikely," was designated when no common place or other epidemiologic connection was identified or when patients had arrived so recently in the country that transmission was unlikely to have occurred. Further details about the formal cluster investigation study are provided elsewhere (9). Data were analyzed by using Epi Info version 6.03 (10). The study was reviewed and approved by the, Human Research Review Committee, Massachusetts Department of Public Health.

Results

Epidemiology of TB in Massachusetts and Genotypes

From July 1, 1996, to December 31, 2000, a total of 1,281 cases were reported and verified as TB by the TB Division, of which 1,032 (81%) were confirmed with positive culture for *M. tuberculosis*. Genotype results were obtained for 984 (95%) of the culture confirmed cases. For the remaining 48 cases, genotype results were not obtained for a variety of reasons, including inability to obtain *M. tuberculosis* isolates from private laboratories and too little growth on culture. Of the 984 TB patients for whom DNA fingerprinting results were obtained, epidemiologic analyses were conducted for 983 patients whose country of origin was known. The greatest risk for developing TB in Massachusetts was being born outside the United States.

Six hundred eighty four (70%) of the TB patients were foreign-born (from 78 different countries). Most (295; 43%) foreign-born patients were from Asia, followed by the Caribbean region (118;17%) and Africa (116;17%). Countries with the highest number of cases included: Vietnam: 87 cases (13%); Haiti, 83 (12%); China, 59 (9%), India, 54 (8%); Cambodia, 31 (5%), and the Dominican Republic, 30 (4%). Analyses of intervals between arrival into the United States and diagnosis of TB indicated that 176 (26%) patients were diagnosed with TB within 1 year of arrival and 353 (52%) were diagnosed with TB within 5 years of arrival (Table 1).

Foreign-born patients were likely to be younger than U.S.-born TB patients (Table 2). Three hundred twenty-seven (48%) of the foreign-born patients were ages 25–44, as compared to 75 (25%) of U.S.-born patients; 103 (15%) of foreign-born patients were >65 years, as compared with 108 (36%) of U.S.-born patients. Foreign-born patients were also more likely to have extrapulmonary disease: 232 (34%) of foreign-born patients had extrapulmonary TB compared with 61 (20%) of U.S.-born patients. TB patients born in the United States were more likely to have been homeless within the year before diagnosis, and drug use and excessive alcohol use were higher

Table 1. Tuberculosis (TB) cases in foreign-born persons by number of years in the United States, by geographic region

Time from arrival in the United States to TB diagnosis	Asia (%)	Caribbean (%)	Africa (%)	Europe (%)	South America (%)	Central America (%)	Former Soviet Union (%)	Other[a](%)	Total (%)
<1 year	64 (22)	24 (21)	47 (41)	10 (29)	16 (34)	9 (22)	6 (27)	0	176 (26)
1–5 years	62 (21)	25 (21)	51 (44)	2 (6)	17 (36)	15 (37)	5 (23)	0	177 (26)
6–10 years	66 (22)	18 (15)	14 (12)	2 (6)	7 (15)	10 (24)	5 (23)	1 (12)	123 (18)
>10 years	103(35)	51 (43)	4 (3)	23 (62)	7 (15)	7 (17)	6 (27)	7 (88)	208 (30)
Total (n=684)	295 (43)	118 (17)	116 (17)	37 (5)	47 (7)	41 (6)	22 (3)	8 (1)	684 (100)

[a]Other, 7 patients from Canada and 1 patient from Australia.

among U.S.-born patients than among foreign-born TB patients. Definition of drug use (injecting drug use and noninjecting drug use), homelessness, and excessive alcohol use are based on CDC criteria as contained in the instruction for the completion of the CDC TB cases reporting forms (11).

Distribution of Genotypes

Analyses of RFLP distribution indicated that 208 (21%)of 983 isolates contained six or fewer copies of IS*6110*. Sixty-seven (22%) of the isolates from 299 U.S.-born TB patients contained few copies of IS*6110*, as did 141 (21%) of the 684 isolates from foreign-born TB patients. However, isolates from foreign-born patients differed substantially by geographic region and country of birth (Table 3).One hundred one (34%) of isolates from Asian patients contained few copies of IS*6110* compared with 2 (4%) of isolates from persons born in South America. In addition, 42 (48%) of isolates from Vietnam contained few copies of IS*6110* compared with 7 (12%) from China.

Genetic Clustering of TB Cases by Genotyping

Of isolates from 983 TB patients, 711 (72%) had DNA fingerprints unique among Massachusetts isolates. The remaining 272 (27.7%) were in 82 genetic clusters. However, 171 (22%) of the 775 isolates containing more than six copies of IS6*110* were in genetic clusters as compared to 100 (48%) of the 208 isolates containing few copies of IS*6110*. Of the 208 isolates, 158 (76%) clustered by IS*6110* alone; 100 (48%) of the isolates remained clustered even with the addition of spoligotyping data to further differentiate the TB strain. The genetic clusters were relatively small in size; 52 (63%) of 82 clusters contained only 2 people, 25 clusters (30%) contained 3–5 people, and the largest cluster contained 16 people. Among the 299 U.S.-born TB patients, 119 (40%) patients had isolates in genetic clusters; 180 (60%) of those had isolates with a unique fingerprint. These figures compare with 153 (22%) of the 684 foreign-born TB patients who had isolates in genetic clusters and 531 (78%) who had unique fingerprints. U.S.-born TB patients were more likely to cluster than foreign-born TB patients (odds ratio [OR] 2.29, 95% confidence interval [CI] 1.69 to 3.12). Foreign-born patients who had lived longer in the United States were more likely to have isolates that clustered than were recent arrivals (chi square for trend 6.31, $p<0.05$). Overall, 29 (16%) of those diagnosed with TB within 1 year of arrival had isolates that clustered with others as compared to 38 (22 %) among those diagnosed from 1 to 5 years of arrival and 26% among those diagnosed >5 years after arrival (Table 4). Stratified analyses by age group (<25, 25–44, 45–64, >65) indicated that clustering was associated with increased time spent in the United States for all age groups; however, the association was strongest among those 25–44 years of age ($p<0.05$).

Likelihood of Epidemiologic Link among Clustered TB Cases

Although the TB genotyping network was started in 1996, cluster investigation did not formally begin until 1998. Therefore, of the 272 patients found in 82 clusters overall, only 161 patients in 52 clusters were investigated for epidemiologic connections as part of the formal Cluster Investigation Study. Information regarding epidemiologic connections was obtained for 152 (94%) of 161 patients. Epidemiologic connections were established for 68 (45%) of the 152 clustered TB cases, but none were found for 84 (55%) of the clustered TB cases. Epidemiologic connections were more likely to be identified for clusters containing only U.S.-born TB patients than clusters containing some or all foreign-born TB patients (62% vs. 42% and 33%, respectively; chi square for trend, $p<0.05$). In addition, clustered TB patients with many copies of IS*6110* were more likely to have epidemiologic connections than clusters with few copies of IS*6110* (OR 8.01; 95% CI 3.45 to 18.93). Of the 90 clustered TB isolates with many copies of IS*6110*, 57 (63%) had epidemiologic connections identified, compared with the 11 (18%) epidemiologic connections that were identified among the 62 clustered TB case-patients with few copies of IS*6110*. Among the U.S.-born patients, 26 (79%) of the 33 patients with many copies of IS*6110* had definite or possible epidemiologic connections, whereas none of the 9 patients with few copies of IS*6110* had connections (Table 5).

Of the 152 clustered TB patients, 42 (28%) were in clusters containing only U.S.-born patients, 67 (44%) were in clusters with mixed U.S.-born and foreign-born patients, and 43 (28%) were in clusters containing only foreign-born patients. Analysis of the 67 TB patients in mixed clusters containing both U.S.-born and foreign-born persons indicate that 38 (57%) of the TB patients were foreign-born, and 29 (43%) were U.S.-born. Epidemiologic connections were established

Table 2. Demographic characteristics of U.S.-born and foreign-born tuberculosis (TB) case-patients[a]

Demographics	Foreign-born (%) n=684	U.S.-born (%) n=299	Odds ratio and 95% confidence interval (CI)	p value
Sex				
Male	380 (56)	183 (61)	0.80 (95% CI 0.60, 1.06)	p=0.11
Female	304 (44)	117 (39)		
Age group[b]				
<1–24	106 (16)	23 (8)	1.0	
25–44	327 (48)	75 (25)	1.06 (95% CI 0.61,1.83)	p=0.83
45–64	148 (22)	93 (31)	2.90 (95% CI 1.67,5.04)	p<0.01
>65	103 (15)	108 (36)	4.83 (95% CI 2.77,8.47)	p<0.01
Site of disease[b]				
Pulmonary	452 (66)	238 (80)	0.50 (95 %CI 0.36,0.70)	p<0.01
Extrapulmonary	232 (34)	61 (20)		
HIV status				
Positive	59 (9)	31 (10)	0.75 (95% CI 0.44,1.30)	p= 0.28
Negative	177 (26)	70 (23)		
Unknown	449 (66)	198 (66)		
Homeless in past year[b]				
Yes	16 (2)	38 (13)	0.16 (95% CI 0.09,0.31)	p<0.01
No	666 (97)	258 (86		
Unknown	2 (<1)	3 (1)		
Drug use in past year[b,c]				
Yes	6 (<1)	27 (9)	0.08 (95% 0.03,0.21)	p<0.01
No	611 (89)	221 (74)		
Unknown	67 (10)	51 (17)		
Excessive alcohol use in past year[b]				
Yes	34 (5)	74 (25)	0.14 (95% CI 0.09,0.22)	p<0.01
No	577 (84)	176 (59)		
Unknown	73 (11)	49 (16)		

[a]Definitions of homeless, dug use and alcohol use are based on criteria established by the Centers for Disease Control and Prevention.
[b] Significant difference observed between U.S.-born and foreign-born at p<0.01.
[c]Includes both injecting and noninjecting drug users.

for 28 (42%) of the 67 TB patients in mixed clusters, and the 17 resulting relationships were analyzed to determine the direction of TB transmission between the cluster members. Results indicate that TB was transmitted from foreign-born to U.S.-born persons in 6 (35%) relationships, foreign-born to foreign-born persons in five (29%) relationships, U.S.-born to U.S.-born persons in three (18%) relationships and U.S.-born to foreign-born persons in three (18%) relationships. However, three of the six foreign-born to U.S.-born relationships involved children of foreign-born parents born in the United States. Epidemiologic relationships were established for 26 (62%) of the 42 TB patients in clusters containing only U.S.-born persons, resulting in 20 relationships. Of the 43 TB patients in clusters containing only foreign-born persons, epidemiologic connections were established for 14 patients (33%), resulting in eight relationships. Overall, of the 45 relationships established through the 68 clustered TB patients with epidemiologic connections, possible TB transmission between U.S.-born persons occurred in 23 (51%) relationships, from foreign-born to foreign-born persons in 13 (29%) relationships, from foreign-born to U.S.-born in 6 (13%) relationships and from U.S.-born to foreign-born in 3 (7%) relationships. In addition, of the 38 foreign-born TB patients in mixed U.S.-born and foreign-born clusters, 10 (26%) TB was diagnosed within 1 year of arrival, in 7 (18%), TB was diagnosed from 1–5 years of arrival, and among 21 (55%), TB was diagnosed > 5 years after the person arrived in the United States. However, TB patients in mixed clusters were no more likely than patients in clusters containing only foreign-born persons to be diagnosed with TB within 1 year, from 1–5 years, or >5 years of arrival (chi square for trend 0.038, p=0.85).

Discussion

The greatest risk of developing TB in Massachusetts is being foreign-born. This finding is consistent with the results found by Mitnick et al., indicating that the foreign-born were

Table 3: *Mycobacterium tuberculosis* IS*6110* copy numbers in genotypes by geographic region

Geographic region	No. of isolates in foreign-born (%) (n=684)	Containing > 6 copies of IS*6110* (%)	Containing ≤6 copies of IS*6110* (%)
Asia	295	194 (66)	101 (34)
China	59	52 (88)	7 (12)
India	54	34 (63)	20 (37)
Vietnam	87	45 (52)	42 (48)
Other	95	63 (66)	32 (34)
Caribbean	118	111 (94)	7 (6)
Dominican Rep	30	30 (100)	0
Haiti	83	76 (92)	7 (8)
Other	5	5 (100)	0
Africa	116	97 (84)	19 (16)
Europe	59	55 (93)	4 (7)
South America	47	45 (96)	2 (4)
Central America	41	36 (88)	5 (12)
Other [a]	8	5 (57)	*3* (43)

[a]Other, 7 patients from Canada and 1 patient from Australia.

7.5 times more likely to have TB than U.S.-born residents of this state (3). An analysis of time from arrival to TB diagnosis indicated that among 26%,TB was diagnosed within 1 year of arrival and among another 26%, it was diagnosed from 1 to 5 years of arrival. This increased risk soon after arrival is particularly true for persons arriving from Africa and South America, among whom TB was diagnosed within 1 year of their arrival for 41% and 35%, respectively. In Massachusetts, the TB Division is notified of refugees and immigrants with a class A or B TB condition identified through the overseas screening process. Together with the Massachusetts Refugee and Immigrant Health Program, the TB Division works to ensure that those refugees and immigrants are evaluated for active TB soon after their arrival in the United States. However, most foreign-born persons moving into Massachusetts are not refugees or immigrants but students or tourists, and therefore the TB Division has little or no information that would allow targeted TB screening.

Only 28% of Massachusetts TB patients had *M. tuberculosis* isolates that were clonally related. Most TB cases were likely the result of reactivation of old infection or recent infection that occurred in the person's country of origin, rather than new infection acquired in this state.. U.S.-born patients were twice as likely to cluster as foreign-born TB patients, suggesting that transmission may be occurring more in the U.S.-born population. U.S.-born TB patients were significantly more likely than foreign-born patients to have a communicable form of TB disease, which may be one more explanation for the increase in clustering among U.S.-born patients. TB transmission between foreign-born and U.S.-born cluster members was established in 9 (20%) of the clustered TB patients with epidemiologic connections; however, to fully examine the extent that U.S.-born and foreign-born TB patients transmit TB in Massachusetts requires a longitudinal investigation of contacts, which was beyond the scope of this investigation. In addition, among those not born in the United States, increased time spent in the United States and clustering appeared to be related. Thus, TB that developed soon after the arrival of the foreign-born appeared to have been acquired abroad, and more of the later onset cases in foreign-born persons appeared to be due to infection acquired in Massachusetts.

The comparison between genotype clustering and epidemiologic connection provides evidence that the ability of DNA fingerprints to differentiate TB strains is limited when there are few copies of IS*6110.* Only 37% of the isolates in clusters containing few copies of IS*6110* had their TB strain differentiated further by spoligotyping. Examination of clustered TB patients with no epidemiologic links indicated that two thirds had few copies of IS*6110.* Epidemiologic connections were more often discovered when the clusters involved U.S.-born TB patients. Despite the use of interpreters, we may have been less successful in obtaining epidemiologic relationship information from foreign-born patients than from U.S.-born patients because of language and cultural barriers. However, even in the clusters of the U.S.-born patients, in which language was not an issue, epidemiologic connections could not be found in clusters with few RFLP bands. This suggests that the use of RFLP analysis, even with spoligotyping, may not be powerful enough to identify true clustering among isolates with few copies of IS*6110*.

The drawbacks to the RFLP technique include the following: it is labor-intensive, requires culture growth, is difficult to reproduce, and can require laborious secondary typing techniques (7,8,12). Other genotyping techniques, such as mycobacterial interspersed repetitive units–variable number of tandem repeats, are being considered that may offer advan-

Table 4. Molecular clustering of tuberculosis (TB) cases among foreign-born persons by time to TB diagnosis after arrival in the United States

Time to TB diagnosis	Cluster (%)	Nonclustered (%)	Odds ratio and 95% confidence interval (CI)	Chi square for trend [a]
<1 year of arrival	29 (16))	148 (84)	1.0	6.31 p =.012
1–5 years of arrival	38 (22)	139 (78)	1.40 (95% CI 0.79,2.47)	
>5 years of arrival	86 (26)	244 (74)	1.80 (95% CI 1.10, 2.95)	
Total	153 (22)	531 (78)		

[a]Statistically significant trend for overall link based on country of origin was observed at $p<0.05$

Table 5. Epidemiologic connection among clustered tuberculosis (TB) cases in cluster investigation study, 1998–2000

	Overall (%)[a]			Many (>6 copies of IS*6110*) (%)			Few (≤6 copies of IS*6110*) (%)		
Cluster contains	No. Cases	Definite/possible	Unlikely	No. Cases	Definite/possible	Unlikely	No. Cases	Definite/possible	Unlikely
U.S.-born only	42	26 (62)	16 (38)	33	26(79)	7(21)	9	0	9 (100)
U.S.-born and foreign-born	67	28 (42)	39 (58)	30	19 (63)	11 (37)	37	9 (24)	28 (76)
Foreign-born only	43	14 (33)	29 (67)	27	12 (44)	15 (56)	16	2 (13)	14 (87)
Total	152	68 (45)	84 (55)	90	57 (63)	33 (37)	62	11 (18)	51 (82)

[a]%, epidemiologic link.

tages, including rapid turnaround time for results, reproducibility, and high sensitivity and specificity for *M. tuberculosis*. However, those methods may have less discriminating power than RFLP (7,12). Analyses of distribution and clustering of RFLP patterns may provide information regarding the ability of RFLP and other possible DNA fingerprinting methods to differentiate TB strains within various communities. For example, our analysis suggests that the ability of DNA fingerprinting to differentiate TB strains in the Asian community may be limited because one third of the isolates contained few copies of IS*6110,* and the secondary fingerprinting technique had less discriminatory power (Table 3).

Some limitations of the study must be noted. First, in RFLP analysis, the usual turnaround time between specimen collection and availability of result is lengthy (7,8). In some years, our turnaround time averaged 8 months. This lag time hindered the program's ability to locate clustered patients for interview and affected the patients' ability to recall contacts, and thus could have contributed to the relatively low percentage of completed interviews (65%). Of 56 patients eligible for interviews, 41% were lost to follow-up or had moved out of state.

Other limitations include the lack of specificity to differentiate TB strains with few copies of IS6110 (7) and incomplete sampling (13). An overestimation of genetic clustering, particularly among isolates with few copies of IS*6110,* may have occurred. On the other hand, clustered TB patients may have been underestimated because possible clonal relationships of isolates from our study population may have existed with patients reported outside of Massachusetts or outside the study time frame. In addition, a certain number of isolates in every population are unable to be given RFLP types.

Conclusions

Molecular fingerprint data were useful in describing the epidemiology of TB in Massachusetts. Using this information, the TB Division can estimate TB patients that resulted from transmission in this state and design appropriate interventions. However, the capacity of DNA fingerprinting data to differentiate TB stains may vary by community of interest, and RFLP analysis, even with secondary typing, may not identify true clusters when isolates have few copies of IS*6110.* This situation has implications for genotyping techniques that have less discriminatory power than RFLP analysis. DNA fingerprinting should therefore be used in conjunction with effective surveillance and appropriate epidemiologic investigation.

Acknowledgments

The authors thank Paul Elvin, Alissa Scharf, and the Mycobacteriology Laboratory, Massachusetts State Laboratory; Denise O'Connor, John Bernardo, and the Tuberculosis (TB) Division nurses of the Boston Public Health Commission; Janice Boutotte and the Division nurses of the Massachusetts Department of Public Health; Muriel Day, JoAnn Dopp, and Harry Taber; Al DeMaria, Barbara Ellis, and Jack Crawford for manuscript review; and Christopher R. Braden for project direction.

This research was supported in part by the National Tuberculosis Genotyping and Surveillance Network cooperative agreement U52/CCU100156, Centers for Disease Control and Prevention.

Ms. Sharnprapai is the director of TB surveillance and epidemiology for the Bureau of Communicable Disease, Division of Tuberculosis Prevention and Control, Massachusetts Department of Public Health.

References

1. Institute of Medicine. Ending neglect. Washington: National Academy Press; 2001:149–58.
2. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2000. Atlanta: U.S. Dept. of Health and Human Services; 2001. p.24.
3. Mitnick C, Furin J, Henry C, Ross J. Tuberculosis among the foreign-born in Massachusetts 1982–1994: a reflection of social and economic disadvantages. Int J Tuberc Lung Dis 1998; 2: S32–S40.
4. Sum AM, Fogg WN, Palma S, Fogg N, Kroshko J, Suozzo, P et al. The changing workforce: immigrants and the new economy in Massachusetts: final report. Boston: Center for Labor Market Studies, Northeastern University; 1999.
5. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993; 31:406–9.
6. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Bunschoten A, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997; 35:907–4.
7. van Soolingen D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Clin Med 2001;249:1–26.

8. Mazars E, Lesjean S, Banuls AL, Gilbert M, Vincent V, Gicquel B, et al. High resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proc Natl Acad Sci USA 2001; 98:1901–6.
9. Miller AC, Sharnprapai S, Suruki R, Corkren E, Nardell EA, Driscoll JR, et al. Impact of genotyping of *Mycobacterium tuberculosis* on public health practice in Massachusetts. Emerg Infect Dis 2002;8;1285–9.
10. Dean AG, Dean JA, Coulombier D, Brendel KA, Smith DC, Burton AH, et al. Epi-Info, version 6: a word processing, database, and statistics program for epidemiology on microcomputers. Atlanta:Centers for Disease Control and Prevention; 1994.
11. Centers for Disease Control and Prevention. SURVS-TB RVCT instructions version 2.0. Atlanta: U.S. Department of Health and Human Services; 1994.
12. Supply P, Lesjean S, Evgueni S, Kremer K, van Soolinger D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 2001; 39:3563–71.
13. Murray M. Sampling bias in the molecular epidemiology of tuberculosis. Emerg Infect Dis 2002; 8:363–9.

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated.

Address for correspondence: Sharon Sharnprapai, Massachusetts Dept. of Public Health, Division of Tuberculosis Prevention and Control, 305 South Street, Boston, MA 02130, USA; fax: 617-983-6990; e-mail: Sharon.Sharnprapai@state.ma.us

Antique poster from the collection of Joel and Jahan Montague. Joel Montague is Board Chairman of Partners for Development, a nonprofit organization. Jahan Montague is associate professor of nephrology, University of Massachusetts Medical School in Worcester.

Transmission of *Mycobacterium tuberculosis* in a Rural Community, Arkansas, 1945–2000

Jennifer A. Dillaha,*† Zhenhua Yang,*† Kashef Ijaz,‡ Kathleen D. Eisenach,*† M. Donald Cave,*† Frank J. Wilson,‡ William W. Stead,‡ and Joseph H. Bates*‡

A cluster of tuberculosis cases in a rural community in Arkansas persisted from 1991 to 1999. The cluster had 13 members, 11 linked epidemiologically. Old records identified 24 additional patients for 40 linked cases during a 54-year period. Residents of this neighborhood represent a population at high risk who should be considered for tuberculin testing and treatment for latent tuberculosis infection.

Research based on IS*6110*-based DNA fingerprinting provides insights into the transmission patterns of *Mycobacterium tuberculosis*. Population-based studies from San Francisco (1) and New York City (2) showed a high incidence of DNA fingerprint clustering that reflected recent transmission. Further studies in urban populations found that clusters occurred within geographic areas, demographic groups, or social networks (3); a high proportion of recently transmitted disease was propagated among a population of difficult-to-serve patients (3). Traditional contact investigations did not reliably identify patients who were infected with the same *M. tuberculosis* strain (4,5).

In contrast to studies in urban areas, few investigations of transmission patterns in rural populations have been undertaken. A study conducted in Arkansas suggests that DNA fingerprint clustering in stable rural populations may result from remote rather than recent transmission (6). However, studies have been limited by short time periods (2–3 years), which made it difficult to assess the contribution of remote transmission to clustering. In addition, some unique strains may result from recent rather than remote transmission, and they would form clusters over a sufficient time span.

The Study

In this study, DNA fingerprints that were based on IS*6110* genotyping were used. This method of strain typing provides stable restriction fragment length polymorphism patterns that are useful in epidemiologic investigations to determine transmission between persons. A cluster is defined as two or more isolates with fingerprints that matched 100% by software analysis using a BioImage Whole Band Analyzer (Millipore, Ann Arbor, MI). Clustering refers to a set of isolates that have the same fingerprint pattern or genotype.

*University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA; †Medical Research Service, Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, USA; and ‡Arkansas Department of Health, Little Rock, Arkansas, USA

During our investigation of nursing home–related tuberculosis (TB) in Arkansas, we used a database of all DNA fingerprints obtained from Arkansas isolates in 1991–1999. A large, persistent cluster in a rural county (county Q, population 21,000) was identified. This cluster was an extension of one described previously (6). The strain variation in county Q was studied to determine whether this cluster could represent an endemic strain. For this purpose, DNA fingerprints obtained from Arkansas isolates during 1993–1999 were examined. (DNA fingerprints were available for 85% of all culture-confirmed cases in county Q during that period.) Public health and medical records identified epidemiologic links among members of the cluster and other TB case-patients in the county.

Sixteen cases of TB were reported from county Q during 1993–1999. Of the 13 culture-positive patients, we obtained DNA fingerprints for 11. Four distinct patterns were observed. The most common pattern was in isolates from eight patients in the county Q cluster. Isolates from three patients from that county did not match any other isolates from county Q, but each patient's isolate matched at least one from another area in Arkansas. DNA fingerprints were not obtained for 2 of the 13 culture-proven cases. One was epidemiologically linked to a case that had one of the unique fingerprints. The second (patient 35, Figure) was epidemiologically linked to the county Q cluster.

Three persons also had negative cultures during this period. They included two siblings identified as contacts to their father, who had recently immigrated from Africa, and a resident of another town in county Q who had no epidemiologic links.

The county Q cluster had a 15-band fingerprint and spanned 9 years (1991–1999). The cluster contained 13 members; 11 had epidemiologic links. Nine of the 11 linked members resided in the same area (neighborhood X) of a rural community (town Y, population 3,500). They frequented the same neighborhood bars associated with extensive transmission in the past. Three of the nine persons (patients 15, 19, and 23; Figure) had culture-confirmed TB that was associated with

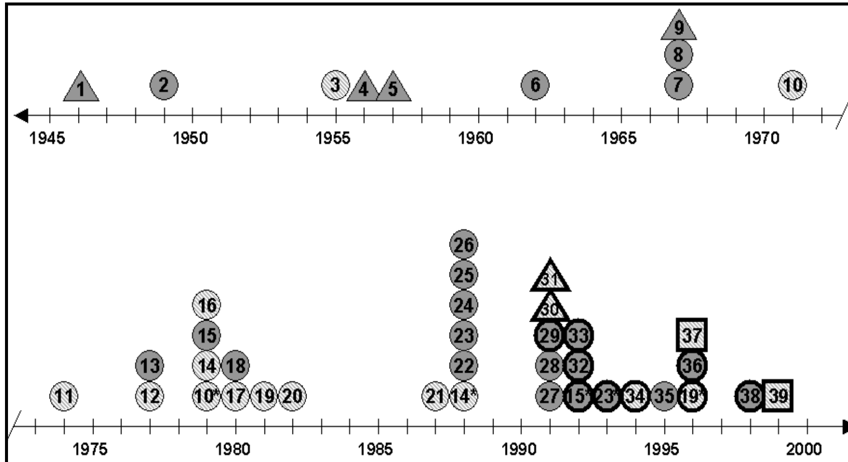

Figure. Timeline of tuberculosis cases in a large, epidemiologically linked group of persons living in rural Arkansas. Circle, epidemiologically linked case-patient, resided in neighborhood X; triangle, epidemiologically linked patient who did not reside in neighborhood X; square, nonepidemiologically linked case-patient; gray, resided in village Z; shaded, did not reside in village Z; bold circle, bold triangle, or bold square, DNA fingerprint available; asterisk, reactivation of prior disease.

bar-related outbreaks in 1979 and 1988. Compliance with appropriately prescribed therapy was poor, according to health-care providers. One of the three persons (patient 19, Figure) required incarceration to complete his treatment. All three persons were HIV negative and probably had reactivation of prior disease. Their initial isolates were not available for comparison.

Eight of the nine cluster members from neighborhood X were men; the woman (patient 32, Figure) was the girlfriend of a bar owner. She had diabetes mellitus, renal failure, HIV infection, and smear-negative pulmonary TB. Five of the eight men had smear-positive sputum; all were HIV negative. Two of the men (patients 19 and 34, Figure) had been inmates in the county jail; one (patient 34, Figure) also had been in a state prison.

Six of the cluster members from neighborhood X were originally from a farming community (village Z, population 200) 15 miles from town Y. Two of these six patients (patients 23 and 36, Figure) are brothers, and family members of five of the six had TB in the past.

Two of the linked cluster members did not reside in town Y. They were medical personnel infected during an autopsy on a linked patient (patient 29, Figure), who died of unrecognized, disseminated TB (7). This patient had returned to Arkansas recently after living in California for many years. His medical records indicated that infection likely occurred after his return to town Y. The two medical personnel were not usual cases; induced sputum grew *M. tuberculosis* and verified that transmission occurred during the autopsy. Symptoms did not develop in either person.

Two cluster members had no epidemiologic links. The first, a 78-year-old nursing home resident (patient 37, Figure), resided in neighborhood X before entering the nursing home in town Y in March 1995. He had a negative two-step tuberculin skin test on admission. TB was diagnosed in August 1996 by a culture from a shoulder aspirate. His chest x-ray showed an abnormality, and his tuberculin skin-test result was positive at the time of diagnosis. During the source case investigation, it was learned that the tuberculin skin tests of four other residents converted after they were admitted. No source case was identified.

The second unlinked cluster member was a 55-year-old woman (patient 39, Figure) who was employed at the county Q jail; she resided in a different community. She was a nurse and had received annual tuberculin skin tests through October 1997 that were negative. After she had worked at the jail for a year, cavitary TB was diagnosed in December 1998. Her tuberculin skin test was positive at the time of diagnosis. No source case was identified.

A review of health department records for county Q identified 24 other persons who had epidemiologic links to the county Q cluster. TB was diagnosed for all of these persons before DNA fingerprinting was available (with the exception of patient 35, Figure). Thus, 40 related cases occurred among 35 patients during 54 years in this community. The earliest documented diagnosis (patient 1, Figure) was in 1946.

All but one of the members of this large group are African-Americans; they resided either in neighborhood X or in village Z. A white teenage boy (patient 3, Figure), who resided in neighborhood X, was the exception. The ages of patients at the time of diagnosis ranged from 2 to 80 years (average 47 years). Seven patients were $\geq$65 years old. Numerous children were exposed during this 54-year period; they were treated for TB but were not included in the group. TB caused at least three deaths in this large, epidemiologically related group, although all isolates remained drug-sensitive.

Conclusions

This endemic strain of TB appears to have persisted in this rural community for at least 54 years despite determined TB-control efforts. Repetitive cycles of transmission and infection occurred and were followed by reactivation or progression to disease. The actual transmission path could not be delineated unequivocally. Whether an episode was recrudescent or primary disease was often indeterminable.

The findings of this rural study are similar to those of urban studies. The county Q epidemiologically related group represents a difficult-to-serve population. Many of the members were poor adherents to therapy. They were distrustful of the public health system and resistant to care offered through

the health department. In addition, alcohol and drug abuse, other illicit activities, and mental illness played major roles in the transmission and adherence dynamics of this group.

Contact tracing with recommendation for treatment of latent TB infection among persons with positive tuberculin skin-test results has been unsuccessful in eliminating TB in this population. Reasons for this failure may include concerns about anonymity and reluctance of persons to disclose contacts who are part of social networks involved in illicit activity. Alternative measures to reduce TB transmission are needed. Case-finding and screening for latent infection on the basis of geographic location rather than personal contacts are important options (8). Such a strategy would incorporate the relevant social networks without stigmatization and identify persons who would be missed otherwise. An effective partnership between the health department and this high-risk neighborhood is crucial to eliminating TB from this community (9).

In contrast to urban areas of the United States, where case rates are 2.5 times higher than the national rate, the rural population in Arkansas has a higher case rate than the urban population. (The 1999 statewide case rate was 7.1 per 100,000. The aggregate rate for counties in metropolitan statistical areas was 5.6, and the aggregate rate for rural counties was 8.5.) In this setting, identification of TB-endemic strains followed by case-finding and screening for latent infection in defined geographic areas may play an important role in the successful elimination of this disease in the United States.

This project was supported by the Centers for Disease Control and Prevention (CDC), Tuberculosis Genotyping and Surveillance Network and by the Donald W. Reynolds Center for Aging. The work described in this report was supported by Veterans Administration–CDC Interagency Agreement 98FED10318. Resources and facilities at the Central Arkansas Veterans Healthcare System in Little Rock, Arkansas, were used.

Dr. Dillaha recently joined the Arkansas Department of Health. She completed her training in Geriatric Medicine at the Donald W. Reynolds Department of Geriatrics at the University of Arkansas for Medical Sciences (UAMS) in 2001. Before then, she trained in Internal Medicine and Infectious Diseases at UAMS. Her research interests include tuberculosis in the geriatric population.

References

1. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
2. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City—an analysis by DNA fingerprinting and conventional epidemiological methods. N Engl J Med 1994;330:1710–6.
3. Bishai WR, Graham NMH, Harrington S, Pope DS, Hooper N, Astemborski J, et al. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 1998;280:1679–84.
4. Barnes PF, Yang ZH, Preston-Martin S, Pogoda JM, Jones BE, Otaya M, et al. Patterns of tuberculosis transmission in central Los Angeles. JAMA 1997;278:1159–63.
5. Chin DP, Crane CM, Diul MY, Sun SJ, Agraz R, Taylor S, et al. Spread of *Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. JAMA 2000;283:2968–74.
6. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, 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.
7. Templeton GL, Illing LE, Young L, Cave MD, Stead WW, Bates JH. The risk of transmission of *Mycobacterium tuberculosis* at the bedside and during autopsy. Ann Intern Med 1995;122:922–5.
8. Barnes PF. Reducing ongoing transmission of tuberculosis [editorial]. JAMA 1998;280:1702–3.
9. Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis 2001;33:1801–6.

Address for correspondence: Jennifer A. Dillaha, Arkansas Department of Health, 4815 West Markham St., Mail Slot 41, Little Rock, AR 72205, USA; fax: 501-661-2055; e-mail: jdillaha@healthyarkansas.com

Poster from the collection of Joel and Jahan Montague. Joel Montague is Board Chairman of Partners for Development, a nonprofit organization. Jahan Montague is associate professor of nephrology, University of Massachusetts Medical School in Worcester.

Cross-Jurisdictional Transmission of *Mycobacterium tuberculosis* in Maryland and Washington, D.C., 1996–2000, Linked to the Homeless

Monica Lathan,*† Leonard Ntaate Mukasa,* Nancy Hooper,* Jonathan Golub,* Nancy Baruch,* Donna Mulcahy,‡ William Benjamin,§ and Wendy A. Cronin*

From 1996 to 2000, 23 Maryland and Washington, D.C., tuberculosis cases were identified in one six-band DNA cluster. Cases were clustered on the basis of their *Mycobacterium tuberculosis* isolates. Medical record reviews and interviews were conducted to identify epidemiologic linkages. Eighteen (78%) of the 23 case-patients with identical restriction fragment length polymorphism patterns were linked to another member; half the patients were associated with a Washington, D.C., homeless shelter. Molecular epidemiology defined the extent of this large, cross-jurisdictional outbreak.

A rise in homelessness in particular and poverty in general partially accounted for the resurgence of tuberculosis (TB) in the United States from 1984 to 1992 (1,2). In Maryland, as part of the National Tuberculosis Genotyping and Surveillance Network activities, population-based DNA fingerprinting of *Mycobacterium tuberculosis* isolates from culture-positive patients was conducted from January 1996 through December 2000. Selected Washington, D.C., isolates from TB patients with suspected or known homelessness were DNA fingerprinted as early as 1996. The Washington, D.C. TB Control staff determined suspected or known homelessness from information contained in case histories and medical records. An interjurisdictional investigation was conducted among homeless persons in Washington, D.C., and Maryland to establish epidemiologic linkages.

The Study

Standard methods were used for IS*6110* restriction fragment length polymorphism (RFLP) analysis of *M. tuberculosis* isolates (3). A cluster was defined as a group of two or more cases with a matching DNA fingerprint pattern (+/-1 band). Spoligotyping (secondary typing) was performed on all clustered strains having six or fewer copies of IS*6110* (4). Medical record reviews and interviews were conducted for all clustered cases to identify connections. We used chi-square and Fisher exact tests to compare demographic and clinical characteristics between the homeless and nonhomeless groups.

A homeless person was defined, at the time a case was reported to the health department, as a person who lacked a fixed, regular, and adequate nighttime residence within the past year or a person who gave a history of homelessness in the recent past (1–5 years). Alcoholism and drug use were routinely documented on the TB case report form. Alcohol abuse was defined as patient-reported alcoholism or disclosure of excessive alcohol use. Recent drug use was defined as injecting or noninjecting drug use within the past year.

From January 1996 through December 2000, nearly all (99 % or 1,181/1,191) of the culture-positive isolates from Maryland were DNA fingerprinted. Since the District of Columbia was not a sentinel site in the genotyping network, only 29 (9%) of the 318 *M. tuberculosis* isolates from culture-positive Washington, D.C., cases were fingerprinted (only those for outbreak investigations). Maryland Cluster A6 (Centers for Disease Control and Prevention [CDC] designations 00104 and 00645) consisted of 23 case isolates; 15 (65%) A6 case-patients were residents of Maryland, and the remaining 8 (35%) were District of Columbia residents. Spoligotypes (CDC designation 3) were identical for all 23 isolates. The first known case in Cluster A6 was reported in March 1996; 22 subsequent cases occurred through November 2000 (Figure). All patients were born in the United States and were African Americans. Homelessness was documented for 14 (61%) of the 23 patients. Eighty-seven percent had acid-fast bacilli smear-positive sputum, and 52% had pulmonary, cavitary disease. Other TB risk factors included alcohol abuse (52%), HIV infection (39%), and drug use (22%). Nonhomeless persons differed from homeless persons because they were more likely to be women ($p<0.02$) and were less likely to have identified risk behaviors. Nonhomeless patients were similar to the homeless in the proportion of cases with pulmonary cavities (Table).

Most cases (78% or 18/23) in Cluster A6 had epidemiologic connections. Nine (50%) of the connected case-patients

*Maryland Department of Health and Mental Hygiene, Baltimore, Maryland, USA; †American Public Health Association, Washington, D.C., USA; ‡Alabama Department of Health, Montgomery, Alabama, USA; and §University of Alabama at Birmingham, Birmingham, Alabama, USA

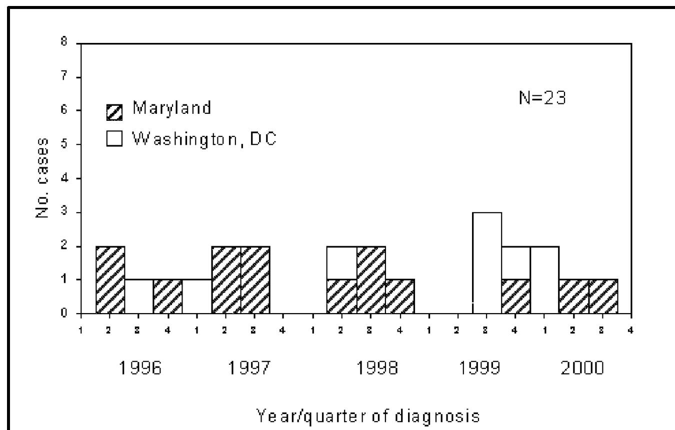

Figure. Cluster A6 tuberculosis cases, Maryland and Washington, D.C.

were directly linked to a large homeless shelter in Washington, D.C. Six case-patients were connected by time and place on the basis of histories of homelessness, socializing with homeless persons, caring for a homeless patient, and sharing boarding or transitional houses. The remaining three were rural nonhomeless case-patients connected by workplace and social links (e.g., drinking); however, they had no other known links with other cluster members. More than one third of the relationships were identified only after the DNA cluster investigation.

In late 1998, active disease caused by the same strain was diagnosed in a nurse who had cared for two hospitalized persons (cluster case-patients), a nonhomeless Maryland resident in 1997 and a homeless Washington, D.C., resident in 1998. Both hospitalized persons were highly infectious with sputum smear-positive and cavitary disease. The nurse cared for the first patient before the patient's diagnosis and subsequent isolation and cared for the second patient only during isolation. Because the second patient reportedly often removed his mask and left his isolation room, we could not determine definitively which case-patient was the source of the nurse's infection.

Our molecular epidemiology study identified TB transmission between homeless and nonhomeless settings in Maryland and provided an opportunity to assess transmission between the state and adjacent Washington, D.C. Population-based molecular epidemiologic studies consistently demonstrate that TB transmission is geographically localized in one or two adjoining jurisdictions (5–8). We are aware of only two DNA-confirmed instances of TB transmission between states (9,10). Interestingly, the Cluster A6 strain exactly matched that seen in a large outbreak among the homeless in Syracuse, New York (J. Driscoll, pers. comm.) (11).The Syracuse outbreak began with a single, highly infectious case that was reported in 1992. We do not know when this strain appeared in Maryland because statewide genotyping was not available in our area before 1996. Further investigation and consultation with Syracuse health department staff showed no additional epidemiologic links between the two states.

The Cluster A6 strain was not identified in isolates from two other sentinel sites in the TB genotyping network (Massachusetts and New Jersey) or more recently in New York City

Table. Selected characteristics among homeless and nonhomeless clustered tuberculosis case-patients, Maryland and Washington, D.C.[a,b]

Characteristics	Homeless No. (%)	Nonhomeless No. (%)	p value[c]
Gender			
Male	13 (92.8)	4 (44.4)	0.02
Female	1(7.2)	5 (55.6)	
Median age in years (range)	42 (27–57)	40 (23–61)	0.97
Substance abuse (IV or non-IV)			
Yes	5 (35.7)	0 (0.0)	
No	9 (64.3)	9 (100.0)	0.11
Alcohol abuse			
Yes	9 (64.3)	3 (33.3)	0.21
No	5 (35.7)	6(66.7)	
HIV status			
Positive	8 (57.1)	1 (11.1)	0.16
Negative	6 (42.9)	6 (66.7)	
Unknown	0 (0.0)	2 (22.2)	
Smear for AFB			
Positive	14 (100.0)	6 (66.7)	0.05
Negative	0 (0.0)	3 (33.3)	
Pulmonary cavities			
Present	7 (50.0)	5 (55.6)	1.00
Absent	7 (50.0)	4 (44.4)	
Mental illness			
Yes	4 (28.6)	1 (11.1)	0.61
No	10 (71.4)	8 (88.9)	
Died			
Yes	4 (28.6)	1 (11.1)	0.61
No	10 (71.4)	8 (88.9)	

[a]NA, not applicable; AFB, acid-fast bacilli.
[b]Washington, D.C., cases were preselected on the basis of a homeless history except for case in the nurse.
[c]Fisher exact tests were used to compare demographic and clinical characteristics between the homeless and nonhomeless groups.

homeless or nonhomeless outbreaks (J. Driscoll, pers. commun.). This observation suggests that the existing links may not be among immediate neighbors but in two non-adjacent juridictions. As disease incidence continues to decline and programs consolidate into regional offices (12), the expanded use of molecular epidemiology will prove increasingly valuable in TB investigations.

Even with expanded (or regional) genotyping of isolates, an active relationship between jurisdictions remains essential to prevent transmission or progression to active disease in patients and their contacts within a region. To investigate Cluster A6, Maryland and Washington, D.C., TB control staff held frequent meetings and teleconferences to review data on RFLP patterns, case characteristics and locations, and contact infor-

mation. Local health department personnel in Washington, D.C., and adjacent Maryland counties routinely share contact data on as many as half of their cases (TB Programs, Washington, D.C., and Maryland, unpub. data, 2002).

Whether the Cluster A6 strain was introduced recently or was a result of prior TB in the Washington, D.C., homeless population is not known. The underlying tuberculin skin-test (TST) positivity was unknown for most case-patients, and only one person had a documented TST conversion. However, persons at high risk for TB, particularly those who were homeless or had HIV infection, are at increased risk for exogenous reinfection (13,14). Homelessness, along with other TB risk factors, can make treatment a daunting task.

Cases reported in 2000 indicate that this large outbreak was not controlled effectively (Figure). Three cases with this strain were also reported in 2001—the most recent in October. The ongoing appearance of patients indicates that TB in the homeless continues to be a challenge in the region. Although 78% of the case connections were found in this difficult-to-reach population, one third of these linkages were identified only after DNA fingerprinting of the *M. tuberculosis* isolates. More thorough contact investigations would not likely have established connections. Cofactors such as substance abuse, mental illness, and HIV infection further complicate the likelihood of obtaining reliable histories and contact reports. Although Cluster A6 may have eventually been traced to the homeless shelter, the magnitude of the outbreak might have never been realized without the inclusion of District of Columbia cases in our DNA fingerprinting sample (regionalization). Universal *M. tuberculosis* genotyping in Washington, D.C., would likely have shown additional cases clustered among the homeless and nonhomeless and further defined the extent of the outbreak.

Conclusions

Molecular epidemiology showed unsuspected TB transmission across jurisdictional borders and transmission involving the homeless and the nonhomeless populations. Investigation of this large interjurisdictional cluster required close collaboration between the Maryland and Washington, D.C., TB Control Programs. As disease incidence continues to decline, regionalizing program efforts with associated *M. tuberculosis* genotyping will be essential to detect and prevent ongoing disease transmission, particularly in difficult-to-reach populations.

Acknowledgments

We thank Margaret Tipple and the staff of the Washington, D.C., Department of Health Tuberculosis (TB) Control Program, and Thomas Walsh and the staff of Montgomery and Prince George's County TB Programs.

Funding for this project was provided by the Centers for Disease Control and Prevention, Cooperative Agreement for the National Tuberculosis Genotyping and Surveillance Network study.

Ms. Lathan is a health science analyst for the American Public Health Association. At the time of this study, she was an epidemiologist with the Maryland Department of Health and Mental Hygiene, Division of Tuberculosis Control, Refugee and Migrant Health. As such, she coordinated interviews and medical record reviews of case-patients with DNA matching strains and investigated a number of tuberculosis outbreaks.

References

1. Brudney K, Dobkin J. Resurgent tuberculosis in New York City. Human immunodeficiency virus, homelessness, and the decline of tuberculosis control programs. Am Rev Respir Dis 1991;144:745–9.
2. Centers for Disease Control and Prevention. Prevention and control of tuberculosis among homeless persons. Recommendations of the advisory council for the elimination of tuberculosis. MMWR Morb Mortal Wkly Rep 1992;41:RR-5.
3. Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
4. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Sooligen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
5. Cronin WA, Golub JE, Magder LS, Baruch NG, Lathan MJ, Mukasa LN, et al. Usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J Clin Microbiol 2001;39:3709–11.
6. Kline SE, Hedemark LL, Davies SF. Outbreak of tuberculosis among regular patrons of a neighborhood bar. N Engl J Med 1995;333:222–7.
7. Bishai WR, Graham NMH, Harrington S, Pope DS, Hooper N, Astemborski J, et al. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 1998;280:1679–84.
8. Bradford WZ, Koehler J, El-Hajj H, Hopewell PC, Reingold AL, Agasino CB, et al. Dissemination of *Mycobacterium tuberculosis* across the San Francisco Bay area. J Infect Dis 1998;177:1101–7.
9. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, et al. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA 1996;275:452–7.
10. Casper C, Singh SP, Rane S, Daley CL, Schechter GS, Riley LW, et al. The transcontinental transmission of tuberculosis: a molecular epidemiological assessment. Am J Public Health 1996;86:551–3.
11. Curtis AB, Ridzon R, Novick LF, Driscoll J, Blair D, Oxtoby M, et al. Analysis of *Mycobacterium tuberculosis* transmission patterns in a homeless shelter outbreak. Int J Tuberc Lung Dis 2000;4:308–13.
12. National Academy of Sciences (NAS). Ending neglect: the elimination of tuberculosis in the United States. Washington: National Academy Press; 2000.
13. Nardell E, McInnis B, Thomas B, Weidhaus S. Exogenous reinfection with tuberculosis in a shelter for the homeless. N Engl J Med 1986;315:1570–5.
14. Sonnenberg P, Murray J, Glynn JR, Shearere S, Kambashi B, Godfrey-Faussett P. HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. Lancet 2001;358:1687–93.

Address for correspondence: Wendy A. Cronin, Maryland Department of Health and Mental Hygiene, 201 West Preston St., Room 307A, Baltimore, MD 21201, USA; fax: 410-669-4215; e-mail: croninw@dhmh.state.md.us

Use of DNA Fingerprinting To Investigate a Multiyear, Multistate Tuberculosis Outbreak

Peter D. McElroy,* Timothy R. Sterling,† Cynthia R. Driver,‡ Barry Kreiswirth,§ Charles L. Woodley,* Wendy A. Cronin,¶ Darryl X. Hardge,*† Kenneth L. Shilkret,# and Renee Ridzon*[1]

In 1998–1999, the Baltimore TB control program detected a cluster of 21 tuberculosis (TB) cases. Patients reported frequent travel to various East Coast cities. An investigation was conducted to determine whether transmission of the same *Mycobacterium tuberculosis* strain was occurring in these other localities. A collaborative investigation among federal, state, and local TB controllers included TB record reviews, interviews of patients, and restriction fragment length polymorphism (RFLP) analysis of selected *M. tuberculosis* isolates from diagnosed TB patients in several cities in 1996–2001. A national TB genotyping database was searched for RFLP patterns that matched the outbreak pattern. Eighteen additional outbreak-related cases were detected outside of Baltimore—the earliest diagnosed in New Jersey in 1996, and the most recent in New York City in late 2001. The outbreak demonstrates the need for strategies to detect links among patients diagnosed with TB across multiple TB control jurisdictions.

Tuberculosis (TB) rates have been declining in the United States since 1993; they reached a low of 5.6 cases per 100,000 population in 2001 (1). To continue this downward trend and eventually achieve the national goal of TB elimination (<1 case per 1 million population per year), ongoing transmission of *Mycobacterium tuberculosis* in social networks needs to be detected earlier (2). When transmission patterns include interstate travel, the epidemiologic connections among TB patients from different localities are often unrecognized, thus inhibiting the ability of local TB controllers to identify a widespread outbreak. New methods that overcome this logistical difficulty will facilitate TB control in this era of increased mobility among difficult-to-reach populations.

Genetic typing methods to differentiate strains of *M. tuberculosis* are useful in identifying disease clusters resulting from recent transmission (3–5). Beginning in August 1999, the Centers for Disease Control and Prevention (CDC) facilitated a cross-jurisdictional investigation of a TB outbreak first recognized in Baltimore, Maryland, and described by Sterling et al. (6). We present results of this extended investigation that uncovered 18 additional outbreak cases after the original 21 were described. Most of these additional patients lived outside the state of Maryland. The outbreak demonstrates the value of a system that allows *M. tuberculosis* strains to be compared across TB control jurisdictions, particularly in situations where unique social and cultural circumstances hinder conventional approaches to contact investigations and control efforts.

Methods

Baltimore Outbreak Profile

Between June 1998 and June 1999, TB was diagnosed in 13 young-adult, U.S.-born African-Americans in Baltimore, Maryland; 8 (62%) were HIV positive. Their *M. tuberculosis* isolates shared a common 11-band DNA fingerprint (6). Upon a review of the National Tuberculosis Genotyping and Surveillance Network database (7), two additional Maryland patients were found to have isolates matching this cluster. Baltimore investigators also suspected that two TB patients in New York City were part of this cluster, a fact that was later verified through genotyping of the patients' isolates. Among these patients with culture-confirmed TB (age range 18–35 years), all but one were associated with the transgender community (6). The transgender persons were members of a "house," a social organization of young-adult transgender persons. House members do not necessarily reside together but regularly engage in dance and dressing competitions known as "balls." A network of several dozen houses exists on the U.S. East

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA; †Baltimore City Health Department, Baltimore, Maryland, USA; ‡New York City Department of Health, New York City, New York, USA: §Public Health Research Institute, New York City, New York, USA; ¶Maryland State Department of Health and Mental Hygiene, Baltimore, Maryland, USA; and #New Jersey Department of Health and Senior Services, Trenton, New Jersey, USA

[1]PDM, TRS, CRD, BK, CLW, WAC, DXH, KLS, and RR were involved in the conception and design of various phases (epidemiologic and laboratory methods) of this investigation. All coauthors were responsible for collection of either epidemiologic or molecular genotyping data. Analysis and interpretation of the data were primarily performed by PDM, TRS, and RR. All coauthors participated in preparation and critical review of the manuscript.

Coast (McElroy, unpub. data). Owing to house members' reported travel patterns, Baltimore investigators suspected that transmission of this same 11-band strain was occurring in other areas, particularly New York City, New Jersey, Atlanta, Boston, Philadelphia, San Francisco, and Washington, D.C. In August 1999, CDC epidemiologists initiated an interstate investigation of this outbreak.

Genetic Typing

Restriction fragment length polymorphism (RFLP) analysis (DNA fingerprinting) of *M. tuberculosis* isolates was performed according to standard methods (8). RFLP patterns were considered to match if the patterns were identical or differed by the addition or subtraction of a single band.

Search of National Tuberculosis Genotyping Network Database

An image of the 11-band Baltimore outbreak strain was compared to RFLP images in the TB genotyping network database at CDC. This database contains >6,000 unique fingerprint images collected in the period January 1996–December 2000 from Arkansas, Maryland, Massachusetts, Michigan, New Jersey, five counties in Texas, and the San Francisco Bay area. Images were analyzed with BioImage Whole Band Analyzer software, version 3.4 (Genomic Solutions, Inc, Ann Arbor, MI). Pairs of patterns were compared for matching bands by using a deviation of ±2.5% for molecular weight of each band. Patterns were clustered by using the unweighted pair group method with arithmetic averages (UPGMA/Average) linkage. All matches were verified by visual comparison.

New York City Investigation

Before 2001, New York City did not routinely perform RFLP analysis. RFLP was retrospectively performed on isolates obtained from 1998 to 2000 from TB patients with demographic characteristics similar to those of the Baltimore outbreak patients (9). After New York City patients associated with the outbreak were identified, their addresses were cross-matched with the entire New York City TB registry. RFLP analysis was performed on isolates from any patient with a street address matching an outbreak-related case. Beginning in January 2001, RFLP analysis was routinely performed on isolates from all TB patients diagnosed in New York City.

Other City Investigations

In Atlanta and Philadelphia, TB charts of patients fitting the outbreak profile were reviewed, and RFLP analyses were performed on isolates from patients strongly suspected of being part of the outbreak. All isolates from Boston and San Francisco are included in the National Tuberculosis Genotyping and Surveillance Network database, which was searched for matches. San Francisco, although part of the genotyping network, also had a database of additional images from cases preceding 1996. Health authorities in Washington, D.C., did not participate in this investigation.

Epidemiologic Investigations

A member of the outbreak investigation team contacted patients with isolates found to have an RFLP pattern matching the outbreak strain. Patients or next-of-kin (for deceased persons) were interviewed, either by telephone or in person. A questionnaire was used to assess regular travel destinations, ball attendance, cross-dressing behaviors, and membership in a house.

Results

As of December 31, 2001, this TB outbreak included 39 patients (36 with culture-confirmed disease) from New Jersey, Maryland, New York City, and Baltimore (Figure). RFLP analysis of the 36 *M. tuberculosis* isolates indicated a matching 11-band pattern in 34 isolates; 2 isolates shared the same 11-band pattern plus 1 additional band. All isolates were susceptible to first-line anti-TB medications. Clinical and demographic characteristics are presented in Tables 1 and 2, respectively.

New Jersey

The initial search of the 6,000 RFLP images in the network database in August 1999 yielded two New Jersey isolates that matched the outbreak strain. The earliest matching RFLP pattern came from an isolate cultured in November 1996 from a 17-year-old woman (patient 3), the sister of a culture-negative 7-year-old girl (patient 2) with clinical TB that had been diagnosed earlier that year (Figure). The second match was an isolate from a 33-year-old woman whose TB was diagnosed in 1997 (patient 5). These three patients were contacts from either the household (patients 2 and 3) or workplace (patient 5) of a 24-year-old man (patient 1) in whom pulmonary TB was diagnosed in 1995, one year preceding initiation of the genotyping network in New Jersey. Thus, although no isolate was available from case-patient 1 at the time of the 1999–2000 outbreak investigation, he was epidemiologically linked to this outbreak. This patient reported a history of frequent travel to

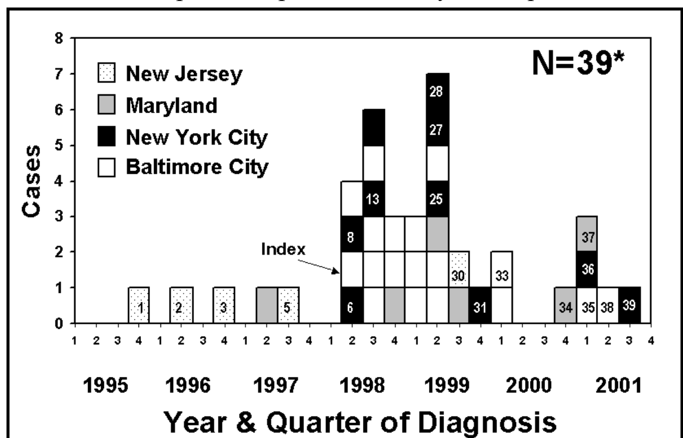

Figure. Epidemic curve representing 38 tuberculosis patients associated with an outbreak involving the cities of Baltimore and New York and the states of Maryland and New Jersey, 1995-2001. *Numbered boxes represent additional patients detected after the investigation was extended beyond Baltimore (August 1999). Unnumbered patients (and patient 28) were previously described by Sterling et al. (6).

Table 1. Clinical characteristics of tuberculosis outbreak patients, New Jersey, New York City, Baltimore, and Maryland, 1995–2001

Characteristic	New Jersey (n=5)	New York City (n=10)	Baltimore (n=18)	Maryland (n=6)	Total (n=39) (%)
Culture-positive	4	10	17	5	36 (92)
Sputum smear–positive	1	4	9	3	17 (44)
Disease site					
Pulmonary only	2	5	11	4	22 (56)
Extrapulmonary only	0	2	4	2	8 (21)
Pulmonary-extrapulmonary	3	3	3	0	9 (23)
Cavitary disease	0	0	2	1	3 (8)
HIV status					
Positive	2	7	11	1	21 (54)
Unknown	1	0	1	1	3 (8)
Deceased[a]	2	2	1	1	6 (15)

[a]Four of six patients died within 3 months of their TB diagnosis.

New York, Baltimore, Philadelphia, and Atlanta to attend and compete in balls but denied ever being a member of a house and had no identified link to any outbreak cases. The most recent New Jersey patient associated with this outbreak was a transsexual who had undergone male-to-female transition surgery (patient 30); her case of TB had been diagnosed in September 1999, and she died in early 2000. As of this writing (mid-2002), no other New Jersey isolates match the outbreak strain.

New York City

RFLP analyses performed on 235 isolates from selected patients diagnosed in 1998–2000 identified seven additional patients (nos. 6, 8, 13, 25, 27, 31, and 36) not previously recognized as being associated with each other or any outbreak case-patients in Baltimore. Interviews indicated that all patients but one (patient 31) were part of the same transgender network described in the Baltimore outbreak. TB was diagnosed in patient 6 in the same month as the case in the Baltimore index patient. The address cross-match identified patient 31, a 1-year old boy who resided in an apartment across the hallway from patient 28. The child's isolate had an RFLP pattern matching the outbreak strain. Neither patient's list of identified contacts included the other patient's name.

Since January 2001, two additional outbreak cases have been identified from New York City after implementation of universal RFLP analysis for isolates obtained from all TB patients in that city. Patient 36 is the only Latino thus far associated with the outbreak. The latest known patient associated with this outbreak (patient 39) was an HIV-coinfected transsexual who had not undergone sex change surgery; plural TB was diagnosed in September 2001.

Maryland

Two additional cases were diagnosed in Maryland after the original report, both in persons from counties bordering Baltimore City. A 42-year-old man with meningeal TB (patient 34) died before he could be interviewed regarding his association with other persons in the outbreak, and his family has not cooperated with this investigation. A 46-year-old woman (patient 37) relocated from New Jersey to Maryland just before her TB diagnosis. The RFLP pattern of her isolate matched the outbreak strain, and subsequent interviews (corroborated by employment records) revealed that in 1995-96 she worked as a custodian at the same county hospital in New Jersey where patient 1 was hospitalized for TB, on the same floor.

Table 2. Demographic and social characteristics of tuberculosis outbreak patients, New Jersey, New York City, Baltimore, and Maryland, 1995–2001

Characteristic	New Jersey (n=5)	New York City (n=10)	Baltimore (n=18)	Maryland (n=6)	Total (n=39) (%)
Median age, yrs (range)	20 (6–33)	30 (1–40)	24 (19–43)	33 (21–46)	26 (1–46)
African-American	4	9	18	5	36 (92)
Born as male	2	10	14	4	30 (77)
House member	1	7	11	0	19 (49)
Pediatric patient	2	1	0	0	3 (8)
Foreign born	1	0	1	0	2 (5)

[a]Probable nosocomial exposure for one of these two patients occurred at a New Jersey hospital before the patient's relocation to Maryland.

Baltimore

Since the first report (6), three additional TB patients with isolates matching the outbreak strain have been diagnosed in Baltimore. Both male patients (patients 33 and 38) were associated with the transgender network. One female patient (patient 35) had contact with a son of the nightclub owner whose club was frequented by most outbreak patients.

Other Sites

RFLP analysis performed on available isolates from patients fitting the outbreak profile in Atlanta (n=26) and Philadelphia (n=3) found that none matched the outbreak strain. RFLP patterns of four San Francisco isolates from the pre-1996 database closely resembled the outbreak strain. These isolates were retyped in 2001 but did not match the outbreak strain. No isolates from Massachusetts or other TB genotyping network sites matched the outbreak strain.

Discussion

Ongoing transmission of *M. tuberculosis* in this outbreak occurred in multiple jurisdictions for at least 4 years (1996–1999). Early recognition of this outbreak by the City of Baltimore TB Control Program staff and their suspicion of a larger outbreak network beyond that city were critical to the initiation of the multistate investigation. The transient nature of the outbreak's social network within Baltimore and the members' propensity for frequent travel and shared accommodations in other cities created an opportunity for extensive transmission of *M. tuberculosis*. Subsequently, the high prevalence of HIV coinfection within this population led to an increased likelihood of progression from latent TB infection to TB disease (10). Through the detection of isolates with matching RFLP patterns, followed by interviews with patients regarding their travel to and participation in transgender social events, TB controllers were able to plan and implement specific interventions for this population. Outbreak-associated cases continue to be detected. Continued transmission of this strain is possible, given that pleural disease was diagnosed in a 29-year-old HIV-coinfected patient in September 2001.

DNA fingerprinting of *M. tuberculosis* isolates has proven to be an extremely useful tool for tracking transmission of various genotypes throughout communities. This technology has a demonstrated role in investigations of transmission within facilities such as prisons (11), hospitals (12), other localized settings such as homeless shelters (13), or in individual states (14). However, few studies have used RFLP to uncover outbreak networks across geographic areas this large. The predictive value of RFLP patterns to document recent *M. tuberculosis* transmission and the relatedness of isolates is often considered stronger in outbreak settings in which links among patients are known (14). This investigation demonstrated that, after a general demographic profile was established among a group of TB patients (e.g., 18- to 35-year-old, HIV-positive, African-American men) from one locality, typing of isolates from other localities allowed investigators to successfully uncover additional matching isolates from patients with characteristics similar to those of the Baltimore cluster. This outbreak demonstrated the value of periodically comparing RFLP fingerprint patterns beyond local jurisdictions to establish whether transmission of a particular strain extends to a broader geographic area. During the second quarter of 1998, four cases with isolates retrospectively matching the outbreak strain were diagnosed in Baltimore and New York City. By early 1998, the national TB genotyping network database already included at least three isolates from New Jersey and Maryland that matched the outbreak strain. Earlier recognition of the common sociodemographic links among the initial case-patients in multiple localities might have averted the subsequent cases diagnosed between 1999 and the present.

Casual transmission of *M. tuberculosis* is defined as transmission from an infectious TB patient to persons who are not household, work, school, or close contacts (15). This outbreak included three patients with isolates matching the RFLP pattern of the outbreak strain but with no identified epidemiologic link to other cases, despite extensive investigation. One should be cautious about viewing cases with no clear epidemiologic link as evidence of casual transmission. These patients' denial of any association with the outbreak network may be a consequence of the secretive nature of some transgender persons, who are often ostracized by society. Some of the male outbreak patients appeared as women only on occasion, and otherwise assimilated into their jobs or schools as men. Many of the transgender patients also engaged in commercial sex work, another potential source of exposure for patients with no identifiable ties to other outbreak patients.

The detection of matching RFLP patterns alone was insufficient to allow investigators to fully characterize the social network in which transmission was occurring. As RFLP analysis of *M. tuberculosis* isolates is performed on a routine basis in large urban areas (e.g., New York City), case-clusters will continue to emerge. As this happens, TB controllers will be expected to direct interventions toward these apparent clusters. Our investigation confirmed that simply detecting a disease cluster, without describing and understanding the social milieu supporting transmission, can lead to incomplete and inefficient TB control. Cutting-edge molecular tools can be enhanced by equally novel epidemiologic approaches. The role of alternative epidemiologic methods such as network analysis, used in tandem with DNA fingerprint analysis, warrants further investigation (16,17).

Acknowledgments

We thank the following persons, who provided epidemiologic or laboratory support during various stages of this investigation during 1999–2001: Athalia Christie, Shama Ahuja, Monica Lathan, Sarah Bur, Nancy Baruch, Leonard Mukasa, Deidre Thompson, Richard Stanley, Azalea Madison, George Brubach, Dawn Tuckey, Jacqueline Elliot, Patrick Brennan, Antonio Paz, Beverly DeVoe, Henry Blumberg, and Diane Stemler. Alejandro Marcel provided expert advice and educational materials regarding transgender health issues.

Portions of this work were supported by Centers for Disease Control and Prevention cooperative agreement funds for the National Tuberculosis Genotyping and Surveillance Network.

Dr. McElroy is an epidemiologist with the Outbreak Investigation Section (OIS) in the Division of Tuberculosis Elimination, Centers for Disease Control and Prevention. OIS responds to requests from state and local health departments for on-site epidemiologic assistance when ongoing transmission of *Mycobacterium tuberculosis* is suspected. Dr. McElroy is also project officer for a study examining the usefulness of network analysis as a tool for tuberculosis control.

References

1. Centers for Disease Control and Prevention. Reported tuberculosis in the
2. Institute of Medicine. Ending neglect: the elimination of tuberculosis in the United States. Washington: National Academy Press; 2000.
3. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
4. Valway SE, Sanchez MPC, Shinnick TF, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med 1998;338:633–9.
5. Agerton TB, Valway SE, Blinkhorn RJ, Orme I, Agerton T, Hoy D, et al. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis,* across the United States. Clin Infect Dis 1999;29:85–92.
6. Sterling TR, Thompson D, Stanley RL, McElroy PD, Madison A, Moore K, et al. A multi-state outbreak of tuberculosis among members of a highly mobile social network: implications for tuberculosis elimination. Int J Tuberc Lung Dis 2000;4:1066–73.
7. Hale YM, Pfyffer GE, Salfinger M. Laboratory diagnosis of mycobacterial infections: new tools and lessons learned. Clin Infect Dis 2001;33:834–46.
8. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
9. Centers for Disease Control and Prevention. HIV-related tuberculosis in a transgender network—Baltimore, Maryland and New York City area, 1998–2000. MMWR Morb Mortal Wkly Rep 2001;49:317–20.
10. Selwyn PA, Sckell BM, Alcabes P, Friedland GH, Klein RS, Schoenbaum EE. High risk of active tuberculosis in HIV-infected drug users with cutaneous anergy. JAMA 1992;268:504–9.
11. March F, Coll P, Guerrero RA, Busquets E, Cayla JA, Prats G, et al. Predictors of tuberculosis transmission in prisons: an analysis using conventional and molecular methods. AIDS 2000;14:525–35.
12. Hennessey KA, Schulte JM, Valway SE, Joglar OT, Rios N, Sheppard JD, et al. Using DNA fingerprinting to detect transmission of *Mycobacterium tuberculosis* among AIDS patients in two health-care facilities in Puerto Rico. South Med J 2000;93:777–82.
13. Curtis AB, Ridzon R, Novick LF, Driscoll J, Blair D, Oxtoby M, et al. Analysis of *Mycobacterium tuberculosis* transmission in a homeless shelter outbreak. Int J Tuberc Lung Dis 2000;4:308–13.
14. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, 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
15. Onorato IM. Tuberculosis outbreaks in the United States. Int J Tuberc Lung Dis 2000;4(12 Suppl 2):S134–8.
16. Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, et al. Networks and tuberculosis: an undetected community outbreak involving public places. Soc Sci Med 2001;52:681–94.
17. Fitzpatrick LK, Hardacker JA, Heirendt W, Agerton T, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis 2001;33:1801–6.

Address for correspondence: Peter D. McElroy, Division of Tuberculosis Elimination, Mailstop E10, Centers for Disease Control and Prevention, 1600 Clifton Rd, NE, Atlanta, GA 30333; fax: 404-639-8959; e-mail: pgm9@cdc.gov

OPPORTUNITIES FOR PEER REVIEWERS

The editors of Emerging Infectious Diseases seek to increase the roster of reviewers for manuscripts submitted by authors all over the world for publication in the journal. If you are interested in reviewing articles on emerging infectious disease topics, please e-mail your name, address, qualifications or curriculum vitae, and areas of expertise to eideditor@cdc.gov

At Emerging Infectious Diseases, we always request reviewers' consent before sending manuscripts, limit review requests to three or four per year, and allow 2-4 weeks for completion of reviews. We consider reviewers invaluable in the process of selecting and publishing high-quality scientific articles and acknowledge their contributions in the journal once a year.

Even though it brings no financial compensation, participation in the peer-review process is not without rewards. Manuscript review provides scientists at all stages of their career opportunities for professional growth by familiarizing them with research trends and the latest work in the field of infectious diseases and by improving their own skills for presenting scientific information through constructive criticism of those of their peers. To view the spectrum of articles we publish, information for authors, and our extensive style guide, visit the journal web site at www.cdc.gov/eid .

For more information on participating in the peer-review process of Emerging Infectious Diseases, e-mail eideditor@cdc.gov or call the journal office at 404-371-5329.

EID
Online
www.cdc.gov/eid

Mycobacterium tuberculosis Transmission between Cluster Members with Similar Fingerprint Patterns

Kashef Ijaz,* Zhenhua Yang,† H. Stewart Matthews,‡ Joseph H. Bates,‡§ and M. Donald Cave§¶

Molecular epidemiologic studies provide evidence of transmission of *Mycobacterium tuberculosis* within clusters of patients whose isolates share identical IS*6110*-DNA fingerprint patterns. However, *M. tuberculosis* transmission among patients whose isolates have similar but not identical DNA fingerprint patterns (i.e., differing by a single band) has not been well documented. We used DNA fingerprinting, combined with conventional epidemiology, to show unsuspected patterns of tuberculosis transmission associated with three public bars in the same city. Among clustered TB cases, DNA fingerprinting analysis of isolates with similar and identical fingerprints helped us discover epidemiologic links missed during routine tuberculosis contact investigations.

The use of DNA fingerprinting has led to important advances in understanding the epidemiology of tuberculosis (TB) (1–5). However, *Mycobacterium tuberculosis* isolates that possess fewer than six copies of IS*6110* do not generate sufficient polymorphism to be readily distinguishable by this technique and therefore require secondary genotyping with another probe (6). A few strains of *M. tuberculosis* lack IS*6110* and cannot be fingerprinted with this technique (7,8). Although the epidemiologic importance of *M. tuberculosis* strains that have more than five copies of IS*6110* with similar but not identical fingerprint patterns (differing by one or two IS*6110* hybridizing bands) is unknown, in some cases epidemiologic links among patients infected with such strains have been established (9). Failure to identify these similar, but not identical, strains results in a misinterpretation of the extent of TB transmission in a community.

We conducted this study to determine the epidemiologic evidence of transmission among patients whose isolates have similar DNA fingerprint patterns (i.e., differing by a single band). Another purpose was to evaluate how useful such evidence, when combined with conventional epidemiology, would be in determining epidemiologic links that may have been missed during routine TB contact investigations in a cluster of patients.

The Study

We obtained *M. tuberculosis* isolates from persons whose cultures were positive for TB from the Arkansas Department of Health and from hospital and reference laboratories where clinical samples from Arkansas patients were submitted. At least one isolate was obtained for each person who had TB. During the period of study (January 1992–December 2000), we identified 1,977 patients with TB; 1,495 were culture positive. IS*6110*-based restriction fragment length polymorphism (RFLP) analysis on the DNA extracted from the isolates of 1,141 patients (77% of culture-positive patients) was performed as described previously (10). Identical DNA fingerprint patterns are present when two or more patients' isolates have indistinguishable IS*6110*-DNA fingerprints of six or more bands or when isolates with fewer than six copies of IS*6110* have identical RFLP patterns and secondary typing with polymorphic GC-rich sequence (PGRS) shows them to be indistinguishable (10–13). Similar DNA fingerprints exist when two or more patients' isolates share an IS*6110*-DNA fingerprint that differs by a single band (i.e., has an additional band [+ 1], lacks a band [- 1], or differs in the size of a single hybridizing band) and has an identical pattern by PGRS (10–13). We performed secondary typing with PGRS for identical IS*6110*-based DNA fingerprint patterns with fewer than six bands or for patterns with six or more bands that were similar but differed by a single band (10–13).

When DNA fingerprints for isolates analyzed from 1992 to 2000 were reviewed for clustering, we identified a cluster of 15 TB patients. For all patients in the cluster, we examined routine TB contact investigation records, medical histories, and laboratory records. By using standard interviews, we established epidemiologic links among case-patients prospectively and retrospectively.

Eleven patients in the cluster resided in a small city in a predominantly rural county (county 1) with a population of 80,268 (14). The other four patients with isolates showing this DNA fingerprint resided in three other Arkansas counties, including one geographically contiguous to county 1. The other two counties were in different parts of the state and were not geographically contiguous to each other. Patients with iso-

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA; †University of Michigan, Ann Arbor, Michigan, USA; ‡Arkansas Department of Health, Little Rock, Arkansas, USA; §University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA; and ¶Central Arkansas Veterans Healthcare Services, Little Rock, Arkansas, USA

lates displaying this DNA fingerprint were not encountered at any other site during the study period.

The 15-patient cluster included 7 patients whose isolates had an identical 13-band fingerprint pattern (pattern A) and 5 patients whose isolates had a DNA-fingerprint pattern that differed by one less hybridizing band (pattern B). The DNA fingerprint pattern of isolates from three other patients differed from one of the two primary patterns by a single band. Secondary typing showed that all 15 case-patients had an identical PGRS pattern (Figure 1). Before DNA fingerprinting, routine contact investigations had established epidemiologic links for only four patients. After this cluster was discovered, epidemiologic investigations demonstrated that 12 of the patients resided in the two geographically contiguous counties, and 10 additional epidemiologic links were established among patients in the 15-member cluster (Figure 2).

Of seven patients (patients 1–7) with pattern A isolates, we discovered epidemiologic links for six. These six patients were connected to three local bars (i.e., commercial drinking establishments). TB was diagnosed for four of the six patients in 1992 and 1993. We found links between four patients connected to bar 1, including the bar's owner (patient 1), a bartender (patient 2), the bartender's husband (patient 3), and a patron (patient 4). Patient 4 also frequented bars 2 and 3. TB was diagnosed in 1996 for patient 5 and patient 6, who frequented bars 2 and 3, respectively. Patients 5 and 6 were in contact in these bars with patient 4 when he was infectious (Figure 2). Patient 7 was not linked to any of these bars; we found no epidemiologic links for him.

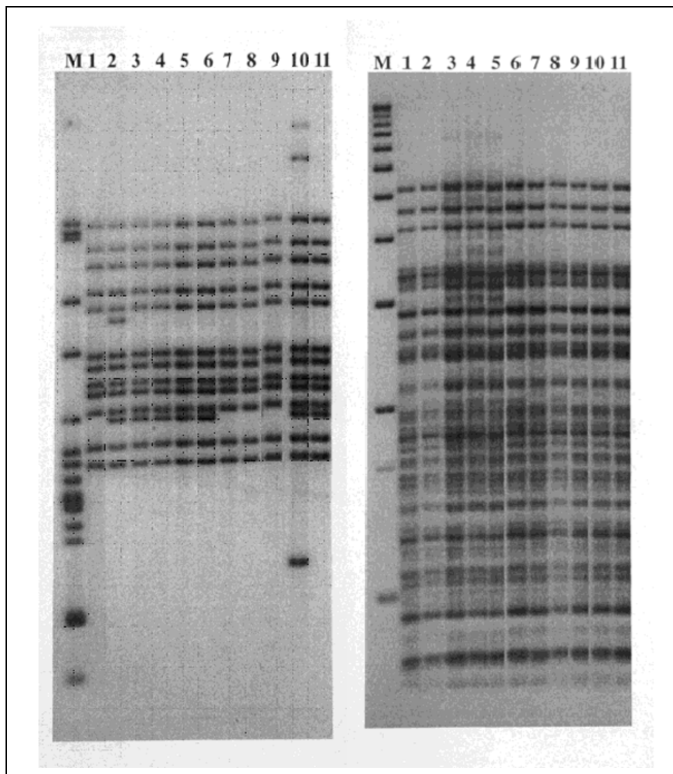

Figure 1. Restriction fragment length polymorphism patterns of *Mycobacterium tuberculosis* isolates from 11 patients residing in two geographically contiguous counties, Arkansas, 1992–1998. IS*6110* patterns are shown on the left and polymorphic GC-rich sequence on the right. Lane M shows *M. tuberculosis* strain H37Rv DNA marker (left) and 1-kb DNA ladder (right). Lane 1, isolate from patient 11; Lane 2, patient 13; Lanes 3–6, patients 4, 1, 3, and 2; Lanes 7–9, patients 10, 9, and 8; Lane 10, patient whose isolate differed by three bands and was not included in the study; and Lane 11, patient 5.

We found epidemiologic links for three of five patients who had pattern B isolates. TB was diagnosed in 1995 for the three patients; two (patients 8 and 9) frequented bar 2, and the third (patient 10) was a granddaughter of patient 9. We could not find epidemiologic links for two patients who had pattern B isolates. Patient 11 was a nurse at the local hospital; no TB patients had been admitted to the hospital where she worked, and she did not remember caring for any TB patients. Patient 12 did not reside in the same city or county. Both patients 11 and 12 denied frequenting any of the bars (Figure 2).

An indirect epidemiologic link was found for one (patient 13) of three patients (patients 13–15). Patient 13's isolate showed a DNA fingerprint pattern that differed from patterns A and B by a single band. Patient 13 was linked indirectly to patient 2 (bar 1) through acquaintances with patient 2's family members (patients 16–18); TB was diagnosed for these patients in 1991. DNA fingerprinting was not conducted for isolates from patients 16, 17, and 18 because they had TB before the beginning of the study (Figure 1). No epidemiologic links were discovered for patients 14 and 15.

Conclusions

Several factors are known to account for changes in RFLP patterns, including single nucleotide mutations that create a new restriction site or lead to loss of a preexisting site. Insertions, duplications, inversions, and deletions can cause changes in restriction fragments. In IS*6110* RFLP patterns, changes can also be accounted for by transposition of the insertion sequence itself. Although the IS*6110* RFLP pattern is sufficiently stable to enable us to make inferences about the linking of patients in a transmission chain, we have observed minor changes in the pattern (15).

DNA fingerprint clusters with similar but not identical IS*6110*-DNA fingerprint patterns should be investigated for epidemiologic links. In this investigation, we used secondary typing with PGRS and spoligotyping (data not reported) to show the isolates as indistinguishable. Previous studies demonstrate that the biological processes measured by PGRS typing or spoligotyping progress more slowly than those of transposable elements like IS*6110* (11). The molecular epidemiologic data indicate that the cases reported were caused by a strain of *M. tuberculosis* with the IS*6110* RFLP pattern that diverged only recently.

During the period 1996–2000, 719 TB cases were confirmed by culture in Arkansas; 707 were genotyped by using IS*6110*. Of cases with isolates having more than five copies of IS*6110*, 234 were in 59 clusters ranging from 2–16 cases. Nineteen of these clusters, including 101 cases, included at least one isolate that was similar but not identical to other isolates in the cluster. In a preliminary study, we discovered epidemiologic

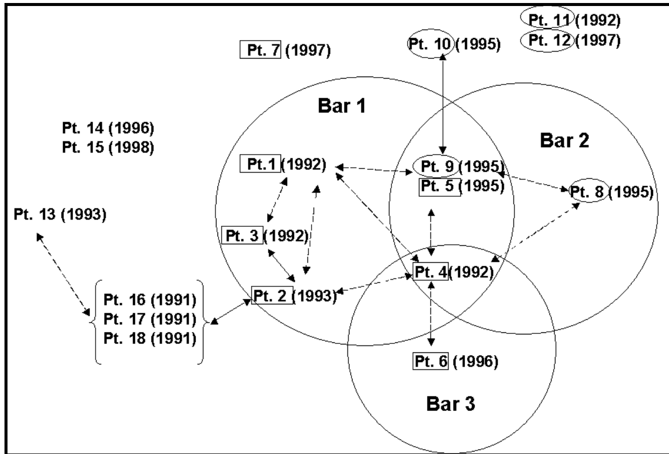

Figure 2. Epidemologic links among tuberculosis patients, Arkansas, 1992–1998. The circles represent the three public bars associated with the cluster of patients. Patient (Pt.) numbers enclosed with boxes and oval circles show patient isolates with patterns A and B, respectively. Parentheses show year of diagnosis. Solid black lines show epidemiologic links found during contact investigations. Dashed black lines show additional epidemiologic links discovered after DNA fingerprinting was done on the isolates and after standardized interviews were conducted with the clustered patients. Absence of lines means that no epidemiologic links were discovered for patients.

links among 15% of patients with IS*6110* RFLP patterns similar to those in clusters. During the same study period, we found epidemiologic links among 32% of Arkansas patients whose isolates shared an identical IS*6110* RFLP pattern.

In this cluster investigation, we discovered additional epidemiologic links among 10 (66%) of 15 patients that were missed during routine contact investigations. In addition, the investigation helped us link eight (53%) of the patients in the cluster to three local bars as common sites of TB transmission (16,17). Our investigation highlighted an extensive social network with multiple epidemiologic links (18).

Acknowledgments

We gratefully acknowledge the contributions of Don Cunningham, Bill Starrett, and Susan Rowland.

The work described in this report was supported by the Veterans Administration and Centers for Disease Control and Prevention Interagency Agreement 98FED10318. The authors used resources and facilities at the Central Arkansas Veterans Health Services Center in Little Rock, Arkansas.

Dr. Ijaz worked with the Arkansas Department of Health TB Program as a medical epidemiologist conducting molecular epidemiology and DNA fingerprinting of *Mycobacterium tuberculosis*. In January 2002, Dr. Ijaz joined the Surveillance and Epidemiology Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention.

References

1. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprint and conventional epidemiologic methods. N Engl J Med 1994;330:1710–6.
2. Edlin BR, Tokars JY, Grieco MH, Crawford JT, Williams J, Sordelo M, et al. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
3. Rastougi NB, Ross BC, Dwyer B, Goh KS, Clavel-Seres S, Jeantils V, et al. Emergence during unsuccessful chemotherapy of multiple drug resistance in a strain of *Mycobacterium tuberculosis*. Eur J Clin Microbiol Infect Dis 1992;11:901–7.
4. Small PM, Hopewell PC, Singh SP, Paz J, Parsonnet PC, Ruston DF, 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.
5. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, et al. Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. N Engl J Med 1993;328:1137–44.
6. van Soolingen D, Hermans PWM, Haas PEW, Soll DR, van Embden JDA. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J Clin Microbiol 1991;29:2578–86.
7. van Soolingen D, Hermans PWM, Haas PEW, Groenen PMA, van Embden JDA. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 1993;31:1987–95.
8. Yuen LKW, Ross BC, Jackson KM, Dwyer B. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. J Clin Microbiol 1993;31:1615–8.
9. Tabet SR, Goldbaum GM, Hooton TM, Eisenach KD, Cave MD, Nolan CM. Restriction fragment length polymorphism analysis detecting a community-based tuberculosis outbreak among persons infected with human immunodeficiency virus. J Infect Dis 1994;169:189–92.
10. Chaves F, Yang Z, El Hajj H, Alonso M, Burman W, Eisenach KD, et al. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J Clin Microbiol 1996;34:1118–23.
11. Yang Z, Chaves F, Barnes PF, Burman WJ, Koehler J, Eisenach KD, et al. Evaluation of method for secondary DNA typing of *Mycobacterium tuberculosis* with pTBN12 in epidemiologic study of tuberculosis. J Clin Microbiol 1996;34:3044–8.
12. Burman WJ, Reves RR, Hawkes AP, Rietmeijer CA, Yang Z, El-Hajj H, et al. DNA fingerprinting with two probes decreases clusters of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:1140–6.
13. Barnes PF, Yang Z, Preston-Martin S, Eisenach KD, Jones BE, Otaya M, et al. Patterns of tuberculosis transmission in central Los Angeles. JAMA 1997;278:1159–63.
14. Fort Smith Chamber of Commerce. 2002. Available from: URL: http://www.fortsmith.org .
15. Cave MD, Eisenach KD, Templeton G, Salfinger M, Mazurek G, Bates J, et al. Stability of DNA fingerprint pattern produced with IS*6110* strains of *Mycobacterium tuberculosis*. J Clin Microbiol 1994;32:262–6.
16. Yaganehdoost A, Graviss EA, Ross MW, Adams GJ, Ramaswamy S, Wanger GJ, et al. Complex transmission dynamics of clonally related virulent *Mycobacterium tuberculosis* associated with barhopping by predominantly human immunodeficient virus-positive gay men. J Infect Dis 1999;180:1245–51.
17. Kline SE, Hedemark LL, Davies SF. Outbreak of tuberculosis among the regular patrons of a neighborhood bar. N Engl J Med 1995;333:222–7.
18. Fitzpatrick LK, Hardacker JA, Heirendt W, Tracy A, Streicher A, Melnyk H, et al. A preventable outbreak of tuberculosis investigated through an intricate social network. Clin Infect Dis 2001;33:1801–6.

Address for correspondence: M. Donald Cave, University of Arkansas for Medical Sciences, 4301 W. Markham, Slot: 510, Little Rock, AR 72205, USA; fax: 501-664-6748; e-mail: cavedonald@uams.edu

A Prospective, Multicenter Study of Laboratory Cross-Contamination of *Mycobacterium tuberculosis* Cultures

Robert M. Jasmer,* Marguerite Roemer,† John Hamilton,‡ John Bunter,§ Christopher R. Braden,¶ Thomas M. Shinnick,¶ and Edward P. Desmond#

A prospective study of false-positive cultures of *Mycobacterium tuberculosis* that resulted from laboratory cross-contamination was conducted at three laboratories in California. Laboratory cross-contamination accounted for 2% of the positive cultures. Cross-contamination should be a concern when an isolate matches the genotype of another sample processed during the same period.

Culture remains the reference standard for diagnosis of disease caused by *Mycobacterium tuberculosis*. However, false-positive results can be caused by cross-contamination of cultures in the laboratory, e.g., when *M. tuberculosis* bacilli are transferred from one specimen to another specimen that does not contain viable bacilli (1–10). Historically, determining whether false-positive results are caused by laboratory cross-contamination has been difficult because of the lack of specific strain identification and nonsystematic criteria. False-positive cultures for *M. tuberculosis* have important implications for clinical management of patients. Many patients are treated on the basis of the results; therefore, patients can be exposed to unnecessary, potentially toxic, and costly treatment. Genotyping of *M. tuberculosis* strains has become the standard method for determining whether isolates are clonal (11–14). This technique, in combination with a review of clinical and radiographic data, allows a determination of the incidence of laboratory cross-contamination of *M. tuberculosis* cultures. In this study, we used predefined criteria to investigate possible laboratory cross-contamination *of M. tuberculosis* cultures and prospectively determine its incidence in an effort to find methods to decrease the occurrence of cross-contamination.

Methods

Study Laboratories and Patients

This study was conducted by staff of the Microbial Diseases Laboratory, California Department of Health Services, Berkeley, California, which is one of seven participants in the National Tuberculosis Genotyping and Surveillance Network of the Centers for Disease Control and Prevention. Participating California laboratories included those at the San Francisco General Hospital, Santa Clara Valley Medical Center (San Jose), and Solano County Public Health Laboratory (Vallejo). The study was conducted from January 1, 1998, to June 30, 1999.

Laboratory Methods

M. tuberculosis isolates from all sources underwent IS*6110*-based DNA restriction fragment length polymorphism (RFLP) analysis (11) if they were 1) the patient's first *M. tuberculosis*–positive culture derived from a sample cultured in the participating laboratory; 2) cultured from a specimen collected >30 days after an *M. tuberculosis* culture–negative specimen was obtained; or 3) cultured from a specimen collected >90 days after the start of appropriate anti-tuberculosis (TB) therapy. When five or fewer bands were present, the isolates underwent secondary genotyping with a RFLP analysis based on a polymorphic GC-rich sequence (12–14).

RFLP pattern images were entered into a database and compared to identify isolates with matching genotypes. Any of the following cultures were considered potentially cross-contaminated and underwent further investigation: 1) the first *M. tuberculosis*–positive culture for a patient whose isolate had a genotype that matched that of another isolate cultured or used in the participating laboratory 2 days before or after the potentially cross-contaminated culture; 2) an *M. tuberculosis* culture from a specimen obtained >30 days after the collection of an *M. tuberculosis* culture-negative specimen that had an isolate with a genotype different from that of any previous isolate from the same patient; or 3) an *M. tuberculosis* culture, from a specimen collected >90 days after the start of appropriate anti-TB therapy, in which the isolate had a genotype different from

*San Francisco General Hospital Medical Center and University of California, San Francisco, California, USA; †San Francisco General Hospital, San Francisco, California, USA; ‡Santa Clara Valley Medical Center, San Jose, California, USA; §Solano County Public Health Laboratory, Vallejo, California, USA; ¶Centers for Disease Control and Prevention, Atlanta, Georgia, USA; and #California Department of Health Services, Berkeley, California, USA

that of any previous isolate from the same patient.

Patients with specimens meeting the above criteria and for whom potential source isolates were identified (i.e., their isolates had a genotype that matched that of a potential source isolate) underwent further investigation. The investigation included a review of all clinical data and radiologic studies, if applicable, and possible epidemiologic connections between patients with potential source and contaminated specimens. Clinical data included prior history of TB, results of tuberculin skin tests, treatment of latent TB infection, symptoms of present illness, results of diagnostic evaluations for TB, and alternative diagnoses. Personnel in laboratories where the potentially cross-contaminated specimens were processed also investigated potential sources of cross-contamination within their facilities.

Determination of Cross-Contamination

A final determination of laboratory cross-contamination was made after a panel of experts reviewed each case. The panel comprised three genotyping network investigators from sites other than California. Panel members met once at the conclusion of the study and reviewed all information from laboratory and medical records, epidemiologic investigations, and genotyping images (upon request). Cases were presented to the panel members in person by one of the authors (RJ). Panel members independently recorded their conclusions on whether cross-contamination was likely, as well as their final diagnosis, and submitted these results to the senior investigator of the study (ED).

Results

Review showed that similar methods for specimen processing were used in all three participating laboratories. However, one laboratory used a common flask for dispensing decontaminant reagent and phosphate buffer rather than an individual tube or pipette for each specimen.

During the study period, 21,835 specimens were submitted for mycobacterial culturing at the three laboratories. Of these, 988 (4.5%) from 296 different patients were positive for *M. tuberculosis*. Twenty-seven had only a single positive culture. Of these, specimens from 10 patients met criteria for an investigation of possible laboratory cross-contamination (Table).

After the panel's review, laboratory cross-contamination was identified as the cause for positive culture results for six patients (2% of all patients with cultures positive for *M. tuberculosis*) (Table). Rates were similar at two of the three laboratories, ranging from 2.8% in both Laboratories 1 (1 of 36 patients) and 2 (5 of 179 patients) to 0% (0 of 81 patients) with a culture positive for *M. tuberculosis* in Laboratory 3. At Laboratory 2, a common flask was used for dispensing reagents during the study period. One of the cross-contamination incidents (involving Patient 4) probably resulted from a malfunctioning broth-culturing system (BACTEC 460) (Becton Dickinson Microbiology Systems, Sparks MD). Patient 4's specimen was in the BACTEC instrument immediately after another patient's specimen (not processed on the same day), and genotypes of the two isolates matched. Cultures from two patients (Patients 7 and 9) were cross-contaminated from the same source patient during sample processing. All six of the laboratory cross-contamination incidents occurred with the initial rather than follow-up specimens for mycobacterial culture. Five of these six patients were treated for TB. Of the remaining four patients whose isolates were suspected of being cross-contaminated in the laboratory, one had a false-positive culture attributed to specimen mislabeling by a health-care provider; one had either a mislabeled specimen or mixed infection; one had active TB; and one had either a mislabeled specimen or a cross-contaminated specimen (Table). Having only a single positive culture was highly associated with laboratory cross-contamination ($p<0.001$, Fisher exact test).

Discussion

Laboratory cross-contamination was the cause for a positive culture result in 2% of all patients with *M. tuberculosis*–positive cultures. Cross-contamination accounted for one fifth (22%) of patients having only one culture positive for *M. tuberculosis*. Of the six patients who had cross-contaminated cultures, five were treated unnecessarily with multiple anti-TB medications.

The rate of cross-contaminated cultures in our study is similar to the rates in previous studies of *M. tuberculosis* cultures. Most population-based studies found rates of 0.9% to 3.5% (1–8). However, such studies were retrospective and did not assess the extent of the problem in different types of clinical mycobacteriology laboratories. In this study, we used predefined criteria, which were based largely on DNA genotyping, to identify suspected cases of laboratory cross-contamination prospectively in an effort to correct factors associated with its occurrence. Our study included all clinical specimens submitted during a 1.5-year period to one county public health laboratory and two county hospital laboratories. This study was possible because of the large databank of RFLP results conducted as part of being a member of the genotyping network.

Multiple factors can cause false-positive cultures, including contaminated clinical equipment (e.g., bronchoscope), clerical errors, and cross-contamination that occurs in the laboratory. The last category can be caused by batch processing, transfer of viable bacilli from the sample needle of a broth-culturing system, e.g., BACTEC (15), a faulty exhaust hood (4), and contamination from species identification procedures such as the niacin production test (6).

Five of the six cross-contamination incidents were in a single laboratory. In four of these five cases, contamination probably occurred when reagents were dispensed with a common flask. Previous studies have reported that the step of adding the phosphate buffer was likely to have been the source of the cross-contamination (1,6,9). This procedure was later discontinued on the basis of the results of this study. None of the laboratories used positive control cultures.

Table. Clinical and laboratory characteristics of patients suspected of having false-positive cultures for *Mycobacterium tuberculosis*[a]

Patient	Lab	Specimen type	Initial or follow-up specimen	Sputum smear result	Clinical signs and symptoms	Comments	Panel decision	Final diagnosis
1	1	Sputum	Initial	Negative	36-yr-old man with AIDS hospitalized with cough, dyspnea, right lower lobe infiltrate; he improved with trimethoprim and sulfamethoxazole alone	Genotype of isolate matched that of another patient whose specimen was processed same day	Cross-contamination	Bacterial pneumonia
2	2	Sputum	Initial	2+ Positive	28-yr-old man with HIV infection with fever, cough, right lower lobe infiltrate; improved with ceftazidime	Genotype of isolate matched that of another patient hospitalized on the same ward	Mislabeled specimen	Bacterial pneumonia
3	2	Sputum	Initial	Negative	31-yr-old woman hospitalized with cough, left lower lobe infiltrate, and leukocytosis; chest radiograph showed improvement with clindamycin and ofloxacin	Genotype of isolate matched that of another patient whose specimen was processed same day	Cross-contamination	Bacterial pneumonia
4	2	Sputum	Initial	Negative	44-yr-old man hospitalized with cough and right lower lobe infiltrate; improved with levofloxacin before anti-TB therapy initiated	Genotype of isolate matched that of another patient whose specimen in the BACTEC instrument immediately preceded the case patient	Cross-contamination	Lung abscess
5	2	Sputum	Follow-up	Negative	68-yr-old woman newly immigrated from China with cough; chest radiograph showed bi-apical fibronodular changes	Genotype of second isolate >30 days later did not match that of initial isolate or any other isolate in database	Mislabeled specimen or mixed infection	Bacterial pneumonia versus TB with mixed infection
6	2	Sputum	Initial	4+ Positive	82-yr-old man with cough, fever; chest radiograph showed chronically increased right mid-lung interstitial markings	Genotype of isolate matched that of another patient hospitalized on the same ward	Mislabeled specimen or cross-contamination	Bacterial pneumonia
7	2	Sputum	Initial	Negative	27-yr-old woman with upper respiratory symptoms; chest radiograph was normal	Genotype of isolate matched that of another patient whose specimen was processed same day	Cross-contamination	Upper respiratory tract infection
8	2	Sputum	Initial	Negative	44-yr-old man with hemoptysis and known pulmonary metastases of squamous cell carcinoma of trachea; chest radiograph showed three large cavities with air-fluid levels; bronchoscopy culture grew *H. influenzae*, and patient improved on antibiotics alone	Genotype of isolate matched that of another patient whose specimen was processed same day	Cross-contamination	Lung abscess
9	2	Broncho-alveloar lavage	Initial	Negative	55-yr-old woman hospitalized with dyspnea and left lower lobe infiltrate; she improved with broad-spectrum antibiotics alone	Genotype of isolate matched that of another patient whose specimen was processed same day	Cross-contamination	Bacterial pneumonia
10	3	Sputum	Follow-up	Negative	34-yr-old homeless man with HIV infection, fever, and cervical lymphadenopathy; chest radiograph showed left lung nodule	Genotype of initial isolate matched that of another homeless TB patient; follow-up specimen (35 days later) had a unique genotype	TB	TB

[a]Lab, laboratory; TB, tuberculosis; *H. influenzae, Haemophilus influenzae*.

As in other studies, we found that a single positive culture for *M. tuberculosis* was a sensitive but nonspecific marker for detection of a false-positive culture, since most patients (78%) with a single positive culture had TB. In a New York study (7), 12 (44.4%) of 27 patients with a single positive culture had a false-positive culture. These findings suggest that clinicians and laboratorians should be increasingly suspicious of a single false-positive culture. Additional specimens should be collected in cases of a single false-positive culture and the patient evaluated carefully for TB and other illnesses; the laboratory

should also retain the isolate and others processed that day for genotyping. Because all specimens that met the inclusion criteria in our study were from the respiratory tract, we cannot draw any conclusions about the rate of cross-contamination of nonrespiratory specimens (e.g., cerebrospinal or pleural fluid). Nor can we draw any conclusions about a single positive culture when only a single specimen is submitted to the laboratory, as is often the case with nonrespiratory specimens.

Clinical judgment is also important in raising suspicion about cross-contamination. TB classically is accompanied by symptoms of prolonged cough, fever, weight loss, and night sweats, but other diseases such as bacterial pneumonia can cause these symptoms. Therefore, no specific clinical criterion alone can be used to definitively state that TB is present. However, an inconsistent clinical course or absence of symptoms should certainly raise suspicion that cross-contamination may have occurred (3,8,16). A determination regarding the presence of cross-contamination requires a thorough evaluation of a patient's symptoms and clinical course as well as laboratory evaluation, with additional specimens obtained if only a single culture is positive as described above. Genotyping should be performed if cross-contamination is suspected on the basis of an inconsistent clinical course or the presence of only one positive culture (8).

Our assessment of the rate of cross-contamination did not include private laboratories; thus, our results may not be generalizable to all types of clinical laboratories. In addition, our methods depended on identifying, obtaining, and genotyping an isolate from a positive source culture; thus, we may have underestimated the true rate of laboratory cross-contamination.

Acknowledgments

The authors gratefully acknowledge the efforts of Donald Cave and Lisa Fitzpatrick, who served as expert reviewers for cases.

This study was supported by funds from the Centers for Disease Control and Prevention (CDC Cooperative Agreement CCU900515-15-2).

Dr. Jasmer is an assistant professor of medicine in the Division of Pulmonary and Critical Care Medicine at the University of California, San Francisco. His primary research interest is in clinical studies of tuberculosis.

References

1. Maurer JR, Desmond EP, Lesser MD, Jones WD Jr. False-positive cultures of *Mycobacterium tuberculosis*. Chest 1984;86:439–43.
2. Nivin B, Kaye K, Munsiff SS. Detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures. Clin Infect Dis 1997;25:943.
3. Bhattacharya M, Dietrich S, Mosher L, Siddiqui F, Reisberg BE, Paul WS, et al. Cross-contamination of specimens with *Mycobacterium tuberculosis*: clinical significance, causes, and prevention. Am J Clin Pathol 1998;109:324–30.
4. Segal-Maurer S, Kreiswirth BN, Burns JM, Lavie S, Lim M, Urban C, et al. *Mycobacterium tuberculosis* specimen contamination revisited: the role of laboratory environmental control in a pseudo-outbreak. Infect Control Hosp Epidemiol 1998;19:101–5.
5. Burman WJ, Stone BL, Reves RR, Wilson ML, Yang Z, El-Hajj H, et al. The incidence of false-positive cultures for *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:321–6.
6. Van Duin JM, Pijnenburg JEM, van Rijswoud CM, de Haas PEW, Hendriks WDH, van Soolingen D. Investigation of cross contamination in a *Mycobacterium tuberculosis* laboratory using IS*6110* DNA fingerprinting. Int J Tuberc Lung Dis 1998;2:425–9.
7. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996;77:407–13.
8. Braden CR, Templeton GL, Stead WW, Bates JH, Cave D, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
9. Burman WJ, Reves RR. Review of false-positive cultures for *Mycobacterium tuberculosis* and recommendations for avoiding unnecessary treatment. Clin Infect Dis 2000;31:1390–5.
10. Centers for Disease Control and Prevention. Misdiagnosis of tuberculosis resulting from laboratory cross-contamination of *Mycobacterium tuberculosis* cultures—New Jersey, 1998. MMWR Morb Mortal Wkly Rep 2000;49:413–6.
11. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
12. Chaves F, Yang Z, El Haji H, Alonso M, Burman WJ, Eisenach KD, et al. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J Clin Microbiol 1996;34:1118–23.
13. Ross BC, Raios K, Jackson K, Dwyer B. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J Clin Microbiol 1992;30:942–6.
14. Hopewell PC, Small PM. Applications of molecular epidemiology to the prevention, control, and study of tuberculosis. In: Rom WN, Garay SM, editors. Tuberculosis. Boston: Little, Brown; 1996. p. 113–27.
15. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, Mickelsen PA. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 1993;31:1677–32.
16. Centers for Disease Control and Prevention. Misdiagnoses of tuberculosis resulting from laboratory cross-contamination of *Mycobacterium tuberculosis* cultures—New Jersey, 1998. MMWR Morb Mortal Wkly Rep 2000;49:413–6.

Address for correspondence: Robert M. Jasmer, Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, Room 5K-1, 1001 Potrero Ave., San Francisco, CA 94110, USA; fax: 415-695-1551; e-mail: rjasmer@itsa.ucsf.edu

Estimated Costs of False Laboratory Diagnoses of Tuberculosis in Three Patients

Jill M. Northrup,* Ann C. Miller,* Edward Nardell,*† Sharon Sharnprapai,* Sue Etkind,* Jeffrey Driscoll,‡ Michael McGarry,‡ Harry W. Taber,‡ Paul Elvin,* Noreen L. Qualls,§ and Christopher R. Braden§

We estimated direct medical and nonmedical costs associated with a false diagnosis of tuberculosis (TB) caused by laboratory cross-contamination of *Mycobacterium tuberculosis* cultures in Massachusetts in 1998 and 1999. For three patients who received misdiagnoses of active TB disease on the basis of laboratory cross-contamination, the costs totaled U.S.$32,618. Of the total, 97% was attributed to the public sector (local and state health departments, public health hospital and laboratory, and county and state correctional facilities); 3% to the private sector (physicians, hospitals, and laboratories); and <1% to the patient. Hospitalizations and inpatient tests, procedures, and TB medications accounted for 69% of costs, and outpatient TB medications accounted for 18%. The average cost per patient was $10,873 (range, $1,033–$21,306). Reducing laboratory cross-contamination and quickly identifying patients with cross-contaminated cultures can prevent unnecessary and potentially dangerous treatment regimens and anguish for the patient and financial burden to the health-care system.

To date, studies investigating cases of laboratory cross-contamination have described only the resources to care for patients who received false diagnoses of tuberculosis (TB) (1–6); to our knowledge, the costs attributable to cross-contamination have not been reported. We estimated direct medical and nonmedical costs for three patients whose misdiagnoses of active TB disease resulted from laboratory cross-contamination of *Mycobacterium tuberculosis* cultures. The costs totaled U.S.$32,618. By examining the costs from the perspective of the patient and the public and private sectors, we documented the financial costs to the health-care system caused by laboratory cross-contamination.

The rate of patients having false-positive *M. tuberculosis* cultures resulting from laboratory cross-contamination may be up to 33% of culture-confirmed TB patients (1–3,7–11). Reportedly two thirds of patients with false-positive cultures are treated for active TB disease (4) and may undergo unnecessary, potentially toxic anti-TB therapy. Public health departments may initiate costly activities such as contact investigations and directly observed therapy. Dunlap et al. report that if persons who receive misdiagnoses resulting from laboratory cross-contamination were treated as TB case-patients with contact investigations and 6 months of directly observed therapy, the costs to the health-care system would be $2,500 per patient (12) in 1993 U.S. dollars, or $3,111 in 1999 dollars, when the Medical Care component of the Consumer Price Index is used to convert 1993 dollars to 1999 dollars.

*Massachusetts Department of Public Health, Boston, Massachusetts, USA; †Harvard Medical School, Boston, Massachusetts, USA; ‡New York State Department of Health, Wadsworth Center, Albany, New York, USA; and §Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Methods

Identifying Patients

As part of the Centers for Disease Control and Prevention–funded National Tuberculosis Genotyping and Surveillance Network, the Massachusetts Department of Public Health, Division of Tuberculosis Prevention and Control (TB Division), conducted a population-based study to determine the rate of TB misdiagnosis in Massachusetts caused by laboratory cross-contamination of *M. tuberculosis* specimens.

The study also evaluated the following criteria that may assist TB control programs to identify patients with potentially cross-contaminated cultures: 1) the patient had a single respiratory specimen positive for *M. tuberculosis,* regardless of acid-fast bacilli (AFB) smear status; a single extrapulmonary body fluid specimen positive for *M. tuberculosis,* regardless of AFB status; or a single tissue specimen positive for *M. tuberculosis* without evidence of AFB or granuloma on histologic examination; 2) the patient had an *M. tuberculosis* culture-positive specimen collected >30 days after the collection of an *M. tuberculosis* culture-negative specimen, and the isolate had a unique genotype compared with any previous isolate from the same patient; 3) the patient had an *M. tuberculosis* culture-positive specimen collected >90 days after the start of appropriate, continuous anti-TB therapy, and the isolate had a unique genotype compared with any previous isolate from the same patient; 4) a caretaker indicated that an *M. tuberculosis* culture-positive result was clinically inconsistent; or 5) a laboratorian indicated that the *M. tuberculosis* culture-positive result might be false.

The Massachusetts Department of Public Health Human Research Review Committee reviewed the protocol and

waived oversight. Personnel in 24 mycobacteriology laboratories (all the laboratories that were processing specimens for AFB for persons in Massachusetts at the time) and public health professionals worked together to identify patients with potentially cross-contaminated specimens. Persons who were reported in Massachusetts as possible TB patients and were reported as having *M. tuberculosis*–positive cultures between January 1, 1998, and June 30, 1999, were prospectively screened. Persons meeting one or more of the criteria were included in the study.

We reviewed laboratory records to identify potential sources of cross-contamination, i.e., any *M. tuberculosis* culture-positive specimen or laboratory control strain processed, reprocessed, or subcultured within 2 working days of the potentially cross-contaminated specimen. For laboratories that did not record usage dates for control strains, the controls were designated as potential sources of cross-contamination and were obtained for genotyping.

The genotype was determined for isolates by IS*6110*-based restriction fragment length polymorphism (RFLP) (13) at the Northeast Regional Genotyping Laboratory, Wadsworth Center, Albany, New York. Spoligotyping (14) was used as a secondary typing method for isolates with five or fewer IS*6110* copies.

Patients with potentially cross-contaminated isolates that matched organisms from potential sources of contamination by genotype and patients for whom a DNA fingerprint could not be produced were investigated. Investigations included reviews of medical and public health department records, abstracts from laboratory data, and patient interviews. Because the criteria would potentially identify not only patients with false-positive *M. tuberculosis* cultures that resulted from laboratory cross-contamination but also patients with false-positive cultures that were caused by other errors as well as true TB cases, a panel of three TB investigators representing other sentinel sites in the genotyping network reviewed the findings. The panel judged whether laboratory cross-contamination was possible, likely, or unlikely and whether the patient had active TB disease or another clinical diagnosis.

Estimating Costs

The cost of TB misdiagnosis was estimated retrospectively for patients who had *M. tuberculosis* culture-positive specimens judged to be possibly or likely caused by laboratory cross-contamination and who received inappropriate diagnoses and were treated for TB because of the false-positive results. (Patients judged to have false-positive *M. tuberculosis* cultures caused by other error were not included in the cost analysis.) Costs for the patient, public sector (local and state health departments, public health hospital and laboratory, and county and state correctional facilities), and private sector (private physicians, hospitals, and laboratories) incurred specifically as a result of the cross-contaminated cultures are included (Table 1). Cost information was collected from the time of initial misdiagnosis until the patient was no longer followed for active TB disease. If patients had other, unrelated medical costs at the same time, the TB medical officer (EN) determined which costs could be attributed to cross-contaminated cultures.

Table 1. Cost inventory for three patients who received misdiagnoses of active tuberculosis disease on the basis of laboratory cross-contamination of *Mycobacterium tuberculosis* specimens[a]

Patient	Public sector[b]	Private sector[c]
Direct medical costs		
TB medications	Outpatient visits TB medications and PPD DOT provision Tests and procedures Contact investigations Hospitalizations	Outpatient visits TB medications and PPD Tests and procedures Contact investigations Hospitalizations
Direct nonmedical costs		
	Case management[d] Overhead[e]	

[a]TB, tuberculosis; PPD, purified protein derivative of tuberculin; DOT, directly observed therapy.
[b]Local and state public health departments, public health hospital and laboratory, and county and state correctional facilities.
[c]Private physicians, hospitals, and laboratories.
[d]Health department case management and administrative support.
[e]Overhead costs, including rent, utilities, and supplies.

Data were collected on direct medical and nonmedical costs for the following: public health department case management and administrative support; outpatient visits; TB medications (started, continued, or changed); directly observed therapy; tests and procedures (bacteriologic, radiologic, chemical, hematologic, pathologic, immunologic, bronchoscopic, and biopsy); health department and hospital contact investigations and diagnostic and treatment services for contacts; and hospitalizations or transfers to hospital isolation rooms. Indirect and intangible costs were excluded.

The resources expended for TB care and treatment were identified from records from these sources: local and state public health departments; inpatient and outpatient medical departments; hospital, clinic, and laboratory billing departments; pharmacies; and mycobacteriology laboratories. We obtained information about contact investigations from health department and hospital infection control personnel, and we asked nurses about public health case management.

Cost estimates were obtained from several sources described below; detailed cost information for these estimates are available upon request. Public health department personnel costs for case management, administrative support, directly observed therapy provision, and contact investigations were estimated by multiplying the sum of annual salaries, fringe benefits, and overhead (rent, utilities, and supplies) by the fraction of the year spent on the activity (as estimated by the health department staff). Costs for providing directly observed therapy at a correctional facility were estimated by multiplying hourly salary by the number of hours spent on the activity, as estimated by the health services administrator. Costs for outpatient visits to health department TB clinics and for tests and

procedures at these clinics were based on the TB Division's reimbursement to the clinics. Costs of private outpatient visits, tests, and procedures were estimated on the basis of information from provider and laboratory billing departments. Costs of TB medications and purified protein derivative (PPD) of tuberculin were based on the TB Division's expenditures for TB drugs and PPD for state fiscal years 1999–2000. The mycobacteriology supervisors estimated costs for mycobacteriology procedures at the public health laboratory and one private laboratory. Charges for hospitalizations and inpatient tests and procedures were obtained from patient billing records and were adjusted to market prices by using Medicare provider-specific, cost-to-charge ratios. The medical services' senior financial analyst estimated the costs for hospitalization at a correctional facility's infirmary on the basis of a flat daily bed rate.

Calculating Cost

Total costs by health-care sector, cost category, and patient were calculated and reported in 1999 U.S. dollars and rounded to the nearest whole dollar. 1998 dollars were adjusted by using the Consumer Price Index Medical Care component. Costs were not discounted because all costs occurred within 1 year of diagnosis.

Results

Rate of TB Misdiagnosis

Between January 1, 1998, and June 30, 1999, 342 of the persons reported as possible TB case-patients in Massachusetts had *M. tuberculosis* positive cultures; of these, 5 (1.5%) had cultures judged to be cross-contaminated in the laboratory. Three (0.9% of 342) of the five persons received misdiagnoses for active TB disease on the basis of the results (Table 2). Each case had been reported as a verified case of TB for national surveillance, but the status was revoked when information from this investigation became available.

Despite their positive cultures, two patients with cross-contaminated cultures were not treated for active TB disease, largely because their physicians did not believe a TB diagnosis was clinically consistent. The mycobacteriology laboratory that processed one patient's specimen questioned the result and performed in-house RFLP typing that confirmed laboratory cross-contamination. Both patients were informed about the false-positive results and reassured about the findings.

Costs by Health-Care Sector

The costs of caring for the three patients whose misdiagnoses and treatment for active TB resulted from laboratory cross-contamination are summarized in Table 3. The total was estimated to be $32,618 in 1999 U.S. dollars. Ninety-seven percent of costs ($31,552) occurred within the public sector: $14,319 at the public hospital, $9,024 within the correctional system, $7,075 to local and state public health departments, and $1,134 to the public health laboratory. Three percent ($949) occurred within the private sector: $381 at hospitals, $316 from laboratories, and $252 for physicians. The patient incurred <1% of the total costs—$118 that went for TB medications.

Costs by Category

Across all sectors, hospitalizations (daily inpatient bed rate and differential for transfer to isolation room) accounted for 59% ($19,348) of total costs. This category was followed by TB medications and PPD ($68 inpatient/$5,774 outpatient), tests and procedures ($3,046 inpatient/$1,695 outpatient), personnel time for directly observed therapy provision ($1,376), outpatient visits ($686), personnel time for health department case management and administrative support ($615), and personnel time for contact investigations ($10). In all, $22, 462 (69%) of the total cost came from hospitalizations and inpatient TB medications, tests, and procedures.

Costs by Patient

The total costs for health care for patients 1, 2, and 3 were $1,033, $10,279, and $21,306, respectively. The average cost per patient was $10,873. Sixty-seven percent of the costs for patient 1 occurred in the private sector: $369 at the hospital, $253 for physicians, and $67 from the laboratory; 22% ($226) by public health departments for case management and administrative support; and 11% ($118) by the patient for TB medications. For patient 2's care, 97% of the costs occurred in the public sector: $7,809 at the public health hospital, $1,491 to health departments, and $720 to the public health laboratory. Three percent of the costs occurred in the private sector: $248 from the laboratory and $11 at the hospital. All costs for patient 3 were within the public sector with $9,024 to county and state correctional facilities, $6,510 at the public health hospital, $5,358 to public health departments, and $414 to the public health laboratory.

Discussion

Rate of TB Misdiagnosis

In Massachusetts, the rate of patients having false-positive cultures resulting from laboratory cross-contamination of *M. tuberculosis* specimens was 1.5% of the culture-confirmed possible TB cases. This rate is within the range demonstrated in other population-based studies (1–3,7–11). In our study, 60% of the patients with cross-contaminated cultures received misdiagnoses and were treated for active TB disease, yielding a rate of TB misdiagnoses caused by laboratory cross-contamination of 0.9% of patients with culture-confirmed TB. These findings corroborate those of Burman and Reves, who estimated that two thirds of patients with false-positive cultures are treated for active TB (4).

Costs of Misdiagnosis

For the three patients, the costs of TB false diagnoses from laboratory cross-contamination fell largely to the public health

Table 2. Characteristics of patients who received misdiagnoses of active tuberculosis disease resulting from laboratory cross-contamination of *Mycobacterium tuberculosis* specimens[a]

Characteristics	Patient 1	Patient 2	Patient 3
Demographic information			
Age at diagnosis (yrs)	59	29	38
Sex	Female	Male	Male
Clinical information			
Site of disease	Lymphatic	Pulmonary	Soft tissue, right index finger
Symptoms when examined	Chronic cough, weight loss, increasing fatigue, night sweats (Sept 1998)	Abdominal discomfort, diarrhea, flank pain, high fever, cough with blood, delirium tremens (Nov 1998)	Infection of right index finger,[b] great pain, lymphangitic streaks up arm (Aug 1998)
Radiology, initial	CAT scan: lymphadenopathy, densities in upper lobes suggestive of infiltration or scarring	Chest x-ray: right lower lobe infiltrate, improved with intravenous ceftriaxone	X-ray right hand: swelling over right index DIP and PIP joints; chest x-ray: normal
Pathology	Lymph node biopsy positive for lymphoma, chemotherapy started	Not applicable	Not done/missing
TST result	Negative	Negative	Negative
Underlying conditions and TB risk factors	History of Hodgkin lymphoma and treatment for active TB disease in 1995,[c] non–U.S.-born	History of chronic alcohol abuse and cocaine use	HIV positive, history of IVDU and incarceration
TB health care			
TB health-care provider	Private physician	Public health department TB clinic	Public health department TB clinic, correctional facility clinic
Type of TB therapy	Self-administered	Daily DOT by public health nurse	Daily DOT by correctional facility staff
Duration of TB therapy	<1 month (started Dec 1998)	<2 months (started Dec 1998)	11 months (treated for 2 weeks in Oct 1998, restarted December 1998)
Hospitalization(s) following TB diagnosis	5 days in private hospital (Jan 1999) with increasing respiratory distress, treated for community acquired pneumonia, died of presumed progression of non–Hodgkin lymphoma	11 days in private hospital with acute gastritis secondary to alcohol abuse (Jan 1999), TB therapy discontinued secondary to increased LFTs; 15 days at public health hospital for TB management; TB ruled out	8 days at public health hospital to start anti-TB therapy and rule out pulmonary and bone involvement (Oct 1998); 5 days in correctional facility infirmary
Contact investigations			
By public health department	Not done	One household contact identified, TST-negative	Not done
By hospital infection control	Not done	Not done	Not done
Information on cross-contaminated specimen			
Specimen type	Right inguinal lymph node tissue	Sputum	Swab of finger cellulitis
AFB smear result	Negative	Negative	Negative
AFB culture result	1 colony at 60 days (reported Dec 1998), sensitive to INH, RIF, EMB, Strep (PZA not tested)	1 colony at 40 days (reported Dec 1998), slightly resistant to INH	"Rare" colonies at 42 days (reported Sept 1998), INH resistant
NTGSN IS*6110* RFLP analysis	10-band pattern (reported April 1999), RFLP match to an isolate from a known TB patient	9-band pattern (reported April 1999), RFLP match to an isolate from a known TB patient	16-band pattern (reported Oct 1999), RFLP match to laboratory control strain H37Ra
Case appraisal results[d]			
Case diagnosis	Lymphoma, nosocomial bacterial pneumonia	Community-acquired pneumonia	Streptococcus cellulitis
Did laboratory cross-contamination occur?	Likely	Likely	Likely

[a]TST, tuberculin skin test; TB, tuberculosis; CAT, computerized axial tomograpy; AFB, acid-fast bacilli; NTGSN, National Tuberculosis Genotyping and Surveillance Network; RFLP, restriction fragment length polymorphism; INH, isoniazid; RIF, rifampin; EMB, ethambutol; Strep, streptomycin; PZA, pyrazinamide; DOT, directly observed therapy; LFTs, liver function tests; DIP, distal interphalangeal; PIP, proximal interphalangeal; and IVDU, intravenous drug use.
[b]Infection of right index finger ultimately resulting in amputation; specimen grew Streptococcus Group A.
[c]Patient treated for active TB disease in 1995, although there was not enough evidence to verify the case for national surveillance.
[d]Case appraisals performed by a panel of three TB investigators representing other NTGSN sentinel sites.

Table 3. Estimated costs for three patients who received misdiagnoses of active tuberculosis disease on the basis of laboratory cross-contamination of *Mycobacterium tuberculosis* specimens[a,b]

	Estimated costs (U.S.$)			
Cost category	Patient 1	Patient 2	Patient 3	Total
Case management[c]	226	288	100	614
Outpatient visits	186	58	443	687
TB medications and PPD	175	606	5,061	5,842
DOT provision[d]	0	508	868	1,376
Tests and procedures	134	1,904	2,703	4,741
Contact investigations[e]	0	10	0	10
Hospitalizations[f]	312	6,905	12,131	19,348
Total	1,033	10,279	21,306	32,618

[a]Costs reported in 1999 U.S. dollars, rounded to the nearest whole dollar. Costs adjusted to 1999 dollars by using the Medical Care group of the Consumer Price Index.
[b]TB, tuberculosis; PPD, purified protein derivative of tuberculin; DOT, directly observed therapy.
[c]Personnel time for health department case management and administrative support.
[d]Personnel time to provide directly observed therapy
[e]Personnel time to perform contact testing.
[f]Daily inpatient bed rate and differential for transfer to isolation room.

and correctional system. Hospitalizations and inpatient tests, procedures, and TB medications accounted for 69% of total health-care sector costs. This finding is consistent with the findings of Brown et al., who demonstrated that inpatient care accounted for 60% of TB health-care expenditures in 1991 even though TB is considered an ambulatory disease (15). Moreover, most of the inpatient costs were for the care of patients 2 and 3, whose underlying circumstances contributed to their hospitalization.

The costs of TB misdiagnosis varied greatly between the three patients (range $1,033–$21,306) and reflected their unique clinical circumstances and treatment courses. Patient 1 had the lowest costs of the three patients. She was treated with anti-TB therapy for only a few weeks before she died of probable lymphoma. Because extrapulmonary TB was diagnosed, no contact investigation was conducted.

Patient 2, with cost of care totaling $10,279, completed <2 months of anti-TB therapy. It was discontinued secondary to elevated liver function tests. The physician found no clinical correlation for a TB diagnosis and did not restart treatment, thus averting additional costs for a full 6-month course of TB therapy. Although pulmonary TB was diagnosed, the patient was AFB smear-negative, so only a limited contact investigation was performed: one household contact was tested.

The greatest cost ($21,306) was for patient 3. The patient's false-positive culture was discovered only as a result of genotyping through the TB genotyping network, and the patient completed 11 months of an intended 12-month course of anti-TB therapy. A diagnosis of single drug-resistant TB and an HIV-positive status further complicated his care. Because extrapulmonary TB was diagnosed, no contact investigation was conducted.

The average cost per patient was $10,873; however, because of the small sample size (n=3), we cannot conclude whether this is a representative estimate of the average cost per TB misdiagnosis. Since two of the patients were hospitalized in a public, long-term care facility rather than an acute-care hospital, the costs were probably much lower than they could have been. However, these preliminary findings demonstrate that substantial costs can result from misdiagnoses caused by laboratory cross-contamination. Additional research with a larger sample size is warranted.

Limitations to the Study

This study included only three patients and did not include estimates of indirect and intangible costs. Since these costs largely affect the patient, we likely underestimated the effects of TB misdiagnosis on patients. Although the consequences were not collected formally, the patients had negative indirect or intangible consequences attributable to the misdiagnoses. The hospitalization of patient 2 and patient 3 represented 13 and 15 days of lost productivity, respectively. In addition, patient 3 underwent a painful bone marrow biopsy to rule out TB involvement of the bone. At the county correctional facility, patient 3 was placed in solitary confinement after testing positive for an illegal substance because of a false-positive reaction from rifampin. The patients likely experienced emotional anxiety, fear, stress, and stigmatization; they were also exposed to unnecessary treatment with potential risks for adverse effects (which did not occur).

Another limitation is that personnel time costs were derived from staff's retrospective estimates of the time involved in various activities, which could have resulted in error. We excluded payers such as Medicare, Medicaid, or private insurers from our study. We showed where costs were incurred within the health-care sectors, but we did not address who actually paid for the resources. This is another area for future research.

Averting Costs of TB Misdiagnoses

This study demonstrates that substantial financial burden can be placed on the health-care system as a result of laboratory cross-contamination. The study also underscores the need for primary prevention of laboratory cross-contamination and the timely recognition of patients who have cross-contaminated *M. tuberculosis* cultures.

Investigators have recommended actions that laboratories, clinicians, and health departments can take to minimize the negative consequences of false-positive *M. tuberculosis* cultures: standardizing laboratory procedures, establishing surveillance for identifying false-positive *M. tuberculosis* cultures, and prospectively screening for patients who may have false-positive cultures (1,4–6,9). We found that clinicians may play an important role in averting the costs associated with TB misdiagnosis resulting from laboratory cross-contamination. Of the five patients with cross-contaminated *M. tuberculosis* cultures, two were not diagnosed with TB because their physicians did not believe the false-positive results. Anti-TB therapy was discontinued after only 2 months for patient 2 because his TB care provider did not believe a TB diagnosis was clinically consistent. Thus interventions may be targeted at physicians who submit samples that test positive for *M. tuberculosis*.

Selective genotyping of isolates from patients who have single positive *M. tuberculosis* cultures may also play a role in limiting the costs of TB misdiagnosis resulting from laboratory cross-contamination. For two of the three study patients who received misdiagnoses for active TB disease, neither the clinicians nor the laboratory personnel reported having suspected that the *M. tuberculosis* culture-positive result might be false; the errors were only detected through routine genotyping by the TB genotyping network.

Actions taken to minimize the negative consequences of false-positive *M. tuberculosis* cultures would require health-care resources. Even so, the costs of implementing these actions would likely be less than the costs of misdiagnosis, especially when intangible costs to the patient are considered. Increased efforts to avoid laboratory cross-contamination and to detect its occurrence as quickly as possible could help prevent unnecessary and potentially dangerous treatment, anguish for the patient, and financial costs to the health-care system.

Acknowledgments

The authors thank Alissa Scharf for assisting in laboratory data abstraction and for subculturing and sending isolates for restriction fragment length polymorphism (RFLP) analysis; Debra Thimas for assisting in clinical data abstraction and organizing the panel review; Jo-Ann Dopp for managing the RFLP database at the Wadsworth Center; Frank Wilson and Zhiyuan Liu for judging potential cases of laboratory cross-contamination; Denise O'Connor for estimating cost data and assisting in obtaining patient clinical information; and John Bernardo, Al DeMaria, Barbara A. Ellis, and Jack T. Crawford for reviewing the manuscript. We are indebted to the nurses at the Massachusetts Tuberculosis Surveillance Area and the Boston Public Health Commission Tuberculosis Program for their assistance in identifying patients with potentially cross-contaminated specimens and obtaining patient interviews; and to the directors and staff of the 24 mycobacteriology laboratories for assisting in prospective screening for laboratory cross-contamination, laboratory data abstraction, and submitting isolates for RFLP analysis.

This research was funded by the Centers for Disease Control and Prevention Cooperative Agreement number U52/CCU100516-16-1. This work was presented in part at the 97th International Conference of the American Thoracic Society in May 2001 in San Francisco, California (Abstract 133441), and at the Northeast Tuberculosis Controllers Conference in October 2001 in Ocean City, Maryland (public health poster session).

Ms. Northrup is an epidemiologist in the Division of Tuberculosis Prevention and Control at the Massachusetts Department of Public Health. Her research interests include molecular epidemiology of tuberculosis.

References

1. Braden CR, Templeton GL, Stead WW, Bates JH, Cave MD, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
2. Gutiérrez MC, Vincent V, Aubert D, Bizet J, Gaillot O, Lebrun L, et al. Molecular fingerprinting of *Mycobacterium tuberculosis* and risk factors for tuberculosis transmission in Paris, France, and surrounding area. J Clin Microbiol 1998;36:486–92.
3. Nivin B, Fujiwara PI, Hannifin J, Kreiswirth BN. Cross-contamination with *Mycobacterium tuberculosis*: an epidemiological and laboratory investigation. Infect Control Hosp Epidemiol 1998;19:500–3.
4. Burman WJ, Reves RR. Review of false-positive cultures for *Mycobacterium tuberculosis* and recommendations for avoiding unnecessary treatment. Clin Infect Dis 2000;31:1390–5.
5. Centers for Disease Control and Prevention. Misdiagnoses of tuberculosis resulting from laboratory cross-contamination of *Mycobacterium tuberculosis* cultures—New Jersey, 1998. MMWR Morb Mortal Wkly Rep 2000;49:413–6.
6. Small PM, McClenny NB, Singh SP, Schoolnik GK, Tompkins LS, Mickelsen PA. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 1993;31:1677–82.
7. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
8. Torrea G, Offredo C, Simonet M, Gicquel B, Berche P, Pierre-Audigier C. Evaluation of tuberculosis transmission in a community by 1 year of systematic typing of *Mycobacterium tuberculosis* clinical isolates. J Clin Microbiol 1996;34:1043–9.
9. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996;77:407–13.
10. van Deutekom H, Gerritsen JJJ, van Soolingen D, van Ameijden EJC, van Embden JDA, Coutinho RA. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. Clin Infect Dis 1997;25:1071–7.

11. Chaves F, Dronda F, Cave MD, Alonso-Sanz M, González-López A, Eisenach KD, et al. A longitudinal study of transmission of tuberculosis in a large prison population. Am J Respir Crit Care Med 1997;155:719–25.
12. Dunlap NE, Harris RH, Benjamin WH, Harden JW, Hafner D. Laboratory contamination of *Mycobacterium tuberculosis* cultures. Am J Respir Crit Care Med 1995;152:1702–4.
13. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
14. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
15. Brown RE, Miller B, Taylor WR, Palmer C, Bosco L, Nicola RM, et al. Health-care expenditures for tuberculosis in the United States. Arch Intern Med 1995;155:1595–600.

Address for correspondence: Jill M. Northrup, Division of Tuberculosis Prevention and Control, State Laboratory Institute, 305 South Street, Boston, MA 02130, USA; fax: 617-983-6990; e-mail: Jill.Northrup@State.MA.US

OPPORTUNITIES FOR PEER REVIEWERS

The editors of Emerging Infectious Diseases seek to increase the roster of reviewers for manuscripts submitted by authors all over the world for publication in the journal. If you are interested in reviewing articles on emerging infectious disease topics, please e-mail your name, address, qualifications or curriculum vitae, and areas of expertise to eideditor@cdc.gov

At Emerging Infectious Diseases, we always request reviewers' consent before sending manuscripts, limit review requests to three or four per year, and allow 2-4 weeks for completion of reviews. We consider reviewers invaluable in the process of selecting and publishing high-quality scientific articles and acknowledge their contributions in the journal once a year.

Even though it brings no financial compensation, participation in the peer-review process is not without rewards. Manuscript review provides scientists at all stages of their career opportunities for professional growth by familiarizing them with research trends and the latest work in the field of infectious diseases and by improving their own skills for presenting scientific information through constructive criticism of those of their peers. To view the spectrum of articles we publish, information for authors, and our extensive style guide, visit the journal web site at www.cdc.gov/eid .

For more information on participating in the peer-review process of Emerging Infectious Diseases, e-mail eideditor@cdc.gov or call the journal office at 404-371-5329.

EID
Online
www.cdc.gov/eid

"Hepatitis C...
what clinicians and other health professionals need to know"
CME, CNE, CEU accreditation available
www.cdc.gov/hepatitis

Molecular Epidemiology of Tuberculosis in a Low- to Moderate-Incidence State: Are Contact Investigations Enough?

Wendy A. Cronin,* Jonathan E. Golub,* Monica J. Lathan,† Leonard N. Mukasa,* Nancy Hooper,* Jafar H. Razeq,* Nancy G. Baruch,* Donna Mulcahy,‡ William H. Benjamin,§ Laurence S. Magder,¶ G. Thomas Strickland,¶ and William R. Bishai#

To assess the circumstances of recent transmission of tuberculosis (TB) (progression to active disease $\leq$2 years after infection), we obtained DNA fingerprints for 1,172 (99%) of 1,179 *Mycobacterium tuberculosis* isolates collected from Maryland TB patients from 1996 to 2000. We also reviewed medical records and interviewed patients with genetically matching *M. tuberculosis* strains to identify epidemiologic links (cluster investigation). Traditional settings for transmission were defined as households or close relatives and friends; all other settings were considered nontraditional. Of 436 clustered patients, 115 had recently acquired TB. Cluster investigations were significantly more likely than contact investigations to identify patients who recently acquired TB in nontraditional settings (33/42 vs. 23/72, respectively; $p<0.001$). Transmission from a foreign-born person to a U.S.-born person was rare and occurred mainly in public settings. The time from symptom onset to diagnosis was twice as long for transmitters as for nontransmitters (16.8 vs. 8.5 weeks, respectively; $p<0.01$). Molecular epidemiologic studies showed that reducing diagnostic delays can prevent TB transmission in nontraditional settings, which elude contact investigations.

Although tuberculosis (TB) remains a major public health threat worldwide (1), its declining incidence in the United States has led health policy makers to develop plans for disease elimination (less than one patient per million) by 2010 (2). Although targeted screening and treatment of latent TB infection has been recommended for groups at high risk (3), learning more about recent TB transmission will help identify specific program interventions that may prevent infection and disease.

Molecular epidemiology has been used to identify groups most at risk for recent TB transmission in high-incidence urban and rural areas of the United States (4–9), but little data have been available from sites with a low-to-moderate disease incidence. Maryland's varied culture and geography provide a microcosm for the study of TB transmission in the United States. The population of 5.1 million resides in distinct areas: urban (city of Baltimore), suburban (5 counties), and rural coastal and mountainous areas (18 counties). Baltimore reports high rates of homelessness, HIV infection, and illegal drug use. Foreign immigration to suburban and some rural areas of the state has increased by 53% in the past decade, causing Maryland to rank third in the nation in rate of foreign population growth (10).

In spite of the presence of these groups at high risk of acquiring TB, excellent treatment regimens utilizing directly observed therapy (87% vs. 47% nationally) and four-drug initial therapy (89% vs. 77% nationally) resulted in a 15-year decline in disease incidence (11). Since 1989, the state's TB incidence has remained lower than the national average (4.9/100,000 vs. 5.6/100,000 population, respectively, in 2001) (11), and Baltimore ranks 26th among 31 major U.S. cities for TB incidence (Centers for Disease Control and Prevention [CDC], unpub. data, 2000).

As part of the CDC-supported National Tuberculosis Genotyping and Surveillance Network, we used DNA fingerprinting of *Mycobacterium tuberculosis* isolates and patient information to conduct a statewide epidemiologic study of culture-positive TB patients. We quantified the problem of recent TB transmission in Maryland, characterized circumstances and settings for transmission, and used our findings to review programmatic interventions.

*Maryland Department of Health and Mental Hygiene, Baltimore, Maryland, USA; †American Public Health Association, Washington, D.C., USA; ‡Alabama Department of Public Health, Montgomery, Alabama, USA; §University of Alabama at Birmingham, Birmingham, Alabama, USA; ¶University of Maryland—Baltimore, Maryland, USA; and #The Johns Hopkins University, Baltimore, Maryland, USA

Methods

Collection of Isolates and DNA Fingerprinting

M. tuberculosis isolates from all patients with a positive culture reported to the Maryland Department of Health and Mental Hygiene (DHMH) between 1996 and 2000 were retrieved from respective reporting laboratories. Restriction fragment length polymorphism (RFLP) analysis of IS*6110* was performed with the standard method (12). Spoligotyping was performed for all matching strains that had six or more IS*6110* copies by using a commercially available kit, according to the manufacturer's instructions (Isogen Bioscience BV, Maarssen, the Netherlands) (13). Patients with genetically related *M. tuberculosis* strains were considered clustered. For high-copy (more than six) IS*6110* strains, patients whose isolate patterns matched exactly, or differed by one band, were assigned a single cluster designation (14,15). For low-copy (six or fewer) strains, cluster designations were assigned to patients whose isolates matched exactly by RFLP analysis and spoligotyping.

Demographic and TB Risk Information

For all culture-positive patients with available DNA fingerprints, we obtained routinely reported demographic and risk factor information (HIV infection, homelessness, incarceration, alcohol abuse and illegal drug use, long-term care residence, and foreign birth) from the state case registry. These data were used to compare patients by estimated time of TB acquisition.

Cluster Investigation

After obtaining genotype results, we abstracted medical records of the clustered patients to determine whether epidemiologic links existed with other patients in the same cluster. We obtained medical histories and information on workplaces, schools, social settings, known or suspected TB exposures, tuberculin skin test results, and contact investigation records. Locatable clustered patients who had no documented links were interviewed to determine whether an existing relationship had eluded the local health department staff who conducted the contact investigations. We assigned epidemiologic links to patients who were named by another TB patient or were in the same place at the same time as another TB patient, even when they did not name each other. When the date and location of specimen collection and laboratory processing suggested that a clustered patient's specimen was falsely positive, a pulmonologist reviewed medical records and chest radiographs to determine whether clinical TB was likely (16,17). Researchers used standardized forms to abstract records and interview patients. The study was approved by DHMH and CDC's institutional review boards, and patients signed informed consent forms before interviews.

Estimated Time of TB Acquisition

Patients with "recent TB" were defined as those who had become infected within 2 years of disease diagnosis by an identified source patient with a matching fingerprint and whose transmission setting was known. Symptom onset had to occur at least 1 month after the onset date of the source's symptoms. The onset date was obtained from the patient's report or conservatively estimated to be 14 days before the date the first positive specimen was collected or the date that treatment was begun, whichever came first. Patients with "probable recent TB" were defined as all clustered patients who had no known transmission from source patients or evidence of past infection, e.g., no history of previous disease or documented positive results of a tuberculin skin test.

The category of "reactivated TB" from latent TB infection was assigned to clustered patients with documented past infection or disease and no identified source case, and to all patients with unique *M. tuberculosis* strain patterns (4,5). Although disease acquisition from a patient residing in another state or from exogenous reinfection could not be completely excluded, we assumed that these events were rare (18,19).

Traditional and Nontraditional Transmission Settings

Traditional settings for transmission were defined as those settings routinely investigated during contact investigations, e.g., households and transmission between close friends and relatives in any location. All other settings where transmission occurred were considered nontraditional.

Time from Symptom Onset to Treatment Initiation

Using only clustered patients as a convenience sample, we compared the times from reported symptom onset to treatment initiation between transmitters (persons who were the source of infection for a patient with recent TB) and nontransmitters (persons who were never identified as a source for another patient). The possibility of transmission was evaluated through September 2002, 21 months after the last patient in the study was reported.

Exclusions

Patients with *M. bovis* infection were excluded. Those with a DNA-confirmed TB relapse (disease occurring $\geq$12 months after treatment was completed, due to an identical *M. tuberculosis* strain) (20) were counted only for the first disease episode. All patients whose time of TB acquisition was undetermined were excluded, including those whose cultures were negative for *M. tuberculosis* and the first patient in a cluster if no source patient was identified (5). Although the infections of patients >5 years of age were recent by definition, children whose cultures were negative were not included in this molecular epidemiologic study, and those results are described elsewhere (21). Finally, because spoligotyping poorly differentiates clustered *M. tuberculosis* strains with low copy IS*6110* in population-based studies (22), we could not confidently determine when TB was acquired by patients who had low-copy IS*6110* strains and no known source acquired TB. These patients were excluded from our comparison between patients by time of TB acquisition.

Analysis

Chi-square tests were conducted for all categorical analyses; Fisher exact test was used when expected cell values were <5. Student *t* test was used for continuous variables.

Results

Culture-Positive TB Patients

Of 1,554 TB patients reported from 1996 through 2000, a total of 1,198 (77%) had positive cultures. The cluster investigations revealed that specimens from 11 patients were false positive, and these patients were deleted from our state TB registry. Five patients with non–BCG *M. bovis* were excluded, and three patients who had DNA-confirmed disease relapse were counted once. No instances of exogenous reinfection from a different *M. tuberculosis* strain were identified. DNA fingerprints were available for 1,172 (> 99%) of 1,179 patient isolates.

Of the 1,172 patient isolates, 436 (37%) were grouped in 111 clusters (median patients per cluster 2; mean 3; range 2–19). Eighty-eight (79%) clusters included persons who resided in one or two adjacent jurisdictions within the state. Overall, 155 (36%) clustered patients were epidemiologically linked to another patient in the cluster; among 336 with high-copy IS*6110* strains, 148 (44%) were linked.

Time of TB Acquisition

The time of TB acquisition could not be determined for 42 patients who were the first symptomatic patient in their respective cluster and had no known source patient, and 145 patients who had low-copy IS*6110* strains and no known source patients (Table 1). These 187 were excluded from our comparison between patients by time of acquisition. However, 29 of the 187 patients were the source for another patient and were included in our analyses of paired source and secondary patients, and of transmitters and nontransmitters.

Of the 985 patients with a known time of infection and subsequent disease, 115 (12%) had recent TB and an additional 82 (8%) had probable recent TB. Fourteen (17%) of these 82 had documented previous negative skin tests. Our extensive case review showed no sources for 56 clustered patients who had documented histories of past infection or disease. We presumed that these 56 patients plus the 732 patients with unique *M. tuberculosis* strains had reactivated disease (n=788).

Patients with recent TB were significantly more likely than patients with probable recent TB to be young and U.S. born, but the proportions of patients with urban residence, HIV infection, illegal drug use, and homelessness were similar for both groups (Table 2). Among the 25 patients with probable recent TB who were >64 years old, 4 were foreign-born, 10 were users of illegal drugs or alcohol, and 2 were homeless. Patients with recent TB were more likely than those with reactivated disease to be urban residents, young, black, U.S.-born, homeless, HIV-infected, and abusers of alcohol or illegal drugs.

Risk Factors among Paired Source and Secondary Patients

Of the 115 patients with recent TB, 114 had 69 sources with available risk information. The mean number of secondary patients per source was 1.6 (median 1; range 1–12). Six (5%) of the 114 secondary patients acquired a resistant *M. tuberculosis* strain (primary resistance) from their source; 2 of these 6 were foreign-born. Five patient-strains were resistant to streptomycin and one was resistant to isoniazid. Risks, particularly illegal drug use, were frequently the same for respective source and secondary patients. Risks were identical (e.g., both source and secondary patients were injection drug users, homeless, HIV-infected) for 47 (72%) of 65 patient pairs aged 15–44 years. We found no transmission from U.S.-born persons to foreign-born persons. Other than birth in a country with a high disease incidence, only 2 (11%) of 18 foreign-born sources had risks compared with 46 (90%) of 51 U.S.-born sources (p<0.001). Foreign-born persons were the sources for all 10 foreign-born secondary patients and eight U.S.-born secondary patients. Among the latter, two were young children who acquired infection from a relative. Nonhousehold transmission from foreign-born persons to the remaining six U.S.-born persons occurred in a school, a hospital (22), two churches, and two workplaces. Five of these U.S.-born patients were immunocompetent, and their only risk for TB was exposure to the infectious source patient.

Identification of Recent Transmission before and after Genotyping

Source cases and settings of transmission were identified for all instances of recent transmission except one, a 3-year-old child (n=114). Fifty-six (49%) patients with recent TB

Table 1. Estimated time of infection and disease acquisition among *Mycobacterium tuberculosis* culture-positive patients by DNA cluster status[a]

DNA cluster status of patients' isolates	No. patients with recent TB	No. patients with probable recent TB	No. patients with reactivated TB	No. patients with unknown time of TB acquisition	Totals
Clustered strains with >6 IS*6110* copies	89	82	56[b]	42[c]	269
Clustered strains with ≤6 IS*6110* copies	22	0	0	145	167
Nonclustered strains	4[d]	0	732	0	736
Total	115	82	788	187	1,172

[a]TB, tuberculosis.
[b]History of previous positive tuberculin skin test or extensive past exposure to a patient.
[c]First patient in a cluster by estimated date of symptom onset and no identified source patient.
[d]Known link to another patient outside the study area or timeframe whose isolate had the same DNA fingerprint.

Table 2. Selected characteristics of culture-positive patients with comparison between categories

Characteristic	No. patients with recent TB (%) (n=115) Group A	No. patients with probable recent TB (%) (n=82) Group B	No. patients with reactivated TB (%) (n=788) Group C	p value A vs. B	p value A vs. C
Residence					
Baltimore City	57 (50.0)	38 (46.3)	157 (19.9)	0.66	<0.001
Other state jurisdictions	58 (50.0)	44 (53.7)	631 (80.1)		
Age group (yrs)					
0–14[a]	6 (5.2)	3 (3.7)	5 (0.6)	<0.001	<0.001
15–24	21 (18.3)	7 (8.5)	86 (10.9)		
25–44	46 (40.0)	27 (32.9)	275 (34.9)		
45–64	33 (28.7)	20 (24.3)	178 (22.6)		
≥65	9 (7.8)	25 (30.4)	244 (31.0)		
Race/ethnicity					
White, non-Hispanic	20 (17.4)	17 (20.7)	162 (20.6)	0.03	<0.001
Black, non-Hispanic	89 (77.4)	51 (62.2)	341 (43.3)		
Hispanic	1 (0.9)	5 (6.1)	84 (10.7)		
Asian	5 (4.3)	9 (11.0)	200 (25.4)		
Native American	0	0	1 (0.1)		
Country of birth					
United States	105 (91.3)	66 (80.5)	360 (45.7)	0.03	<0.001
Other	10 (8.7)	16 (19.5)	428 (54.3)		
Long-term care resident					
Yes	7 (6.1)	5 (3.3)	25 (3.2)	1.00	0.11
No	108 (93.9)	77 (96.7)	763 (96.8)		
Homeless					
Yes	18 (15.7)	8 (9.8)	23 (2.9)	0.22	<0.001
No	97 (84.3)	74 (90.2)	765 (97.1)		
Prison resident					
Yes	13 (11.3)	10 (13.1)	21 (2.7)	0.85	<0.001
No	102 (88.7)	72 (86.9)	767 (97.3)		
Uses illegal drugs or abuses alcohol					
Yes	53 (46.0)	30 (36.6)	76 (9.6)	0.18	<0.001
No	62 (54.0)	58 (63.4)	712 (90.4)		
HIV-infected					
Yes	28 (24.3)	19 (23.2)	75 (9.5)	0.85	<0.001
No	87 (75.7)	63 (76.8)	713 (90.5)		

[a]Includes one child <6 years old without a known source patient; the case was classified as recent based on age.

acquired their infection and disease in nontraditional settings (Table 3). Less than two-thirds of the recent patients' epidemiologic links to their source patients were identified by routine contact investigations before genotyping. Patients identified by contact investigations were significantly less likely to have acquired TB in nontraditional settings than those identified by cluster investigations (23/72 vs. 33/42, respectively; $p<0.001$).

The importance of nontraditional settings among persons at high risk was influenced in part by large outbreaks (three or more secondary patients) (23–29). Nine of these began in nontraditional settings and ultimately expanded to traditional settings, and cluster investigation identified additional outbreak-related infections in patients who had not been identified through routine contact investigations (25–27).

TB acquisition in nontraditional settings was associated with age >14 years ($p=0.033$, compared to younger patients aged 0 to 14 years), U.S. birth ($p=0.012$, compared to foreign birth), and illegal drug use ($p<0.001$, compared to nonusers).

Table 3. Identified transmission settings for 114 patients with recently acquired tuberculosis (TB)

Settings	Total patients with known settings (%)	Setting identified by routine contact investigation (%)	Setting identified by DNA cluster investigation (%)
Traditional			
Household	28 (24.6)	25 (34.7)	3 (7.1)
Close relative	13 (11.4)	13 (18.1)	0
Close friend	17 (14.9)	11 (22.2)	6 (14.3)
Nontraditional			
Hospital (24,28)	10 (8.8)	5 (6.9)	5 (11.9)
Other workplace (25)	6 (5.3)	6 (8.3)	0
Social club (26)	11 (9.6)	7 (9.7)	4 (9.5)
Homeless shelter	5 (4.4)	0	5 (11.9)
Bar	10 (8.8)	1 (1.4)	9 (21.4)
Prison/jail (26)	5 (4.4)	3 (4.2)	2 (4.8)
Store (27)	2 (1.8)	0	2 (4.8)
Church	2 (1.8)	0	2 (4.8)
Nursing home	2 (1.8)	0	2 4.8)
School	1 (0.9)	0	1 (2.4)
Ship	1 (0.9)	1 (1.8)	0 (2.4)
Mortuary (29)	1 (0.9)	0 (1.4)	1 (2.4)
Total	114 (100.0)	72 (100.0)	42 (100.0)

At least 5 of 15 patients who acquired TB in public settings, i.e., churches, hospitals, a school, and a store, had only brief or distant (casual) exposure to a highly infectious person (24,28). Nine (60%) of the 15 had no apparent TB risk factor except exposure to their source patient.

TB transmission occurred in households for all 10 foreign-born persons with recent TB, and the sources for all but one foreign-born patient were found by contact investigations (Table 4). Cluster investigations were significantly more likely than contact investigations to identify source cases for patients who were homeless, abusers of alcohol, or both. Recent patients with other common TB risk factors, i.e., HIV infection, illegal drug use, incarceration, and long-term care residence, were equally likely to have epidemiologic links identified by cluster or contact investigations.

Time from Symptom Onset to Treatment Initiation

The estimated time of symptom onset was available for 69 transmitters and 99 nontransmitters. TB transmitters were significantly more likely than nontransmitters to have pulmonary disease (68/69 vs. 73/99; $p<0.001$). Among patients with pulmonary disease, transmitters were more likely than nontransmitters to have lung cavitation (40/68 vs. 14/73; $p<0.001$) and sputum smears positive for acid-fast bacilli (64/68 vs. 59/73; $p=0.034$). Among transmitters, the mean time from symptom onset to treatment initiation was 16.8 weeks compared with 8.5 weeks among nontransmitters (median 11 vs. 6 weeks, respectively; $p=0.008$). Transmitters also were more likely than nontransmitters to have at least one risk factor for TB, e.g., homelessness, HIV infection, alcohol abuse, or illegal drug use, residence in a long-term care facility, incarceration, foreign birth (60/69 vs. 38/99, respectively; $p<0.001$).

Discussion

Our 5-year molecular epidemiologic study featured a complete sampling of patients' isolates from the entire state (30,31) and a multifaceted study site. We also compared our patient groups by time of disease acquisition to more clearly define the relationship between clustering and recent transmission in the state. Even though Maryland has low-to-moderate TB incidence, results from our comparison between groups with recent, probable recent, and reactivated TB were similar to those from studies conducted among clustered and nonclustered patients in high incidence urban and rural areas. Recent and probable recent TB acquisition were associated with patients who were young, homeless, users of alcohol and illegal drugs, HIV-infected, and incarcerated. These findings further support the assumption that clustering is a reasonable, though not exact, surrogate for recent transmission (4–9).

The importance of clustered patients who do not have identifiable links has remained unclear (32). By assuming that clustered patients without links and with histories of old infections or previous TB had reactivated disease, we attempted to be more specific in identifying those for whom recent TB was plausible. Our patients with probable recent TB were older and more likely to be foreign-born than were patients with recent TB. Half of the elderly patients had other high-risk factors that made exposure and recent infection likely. Among the foreign-born, acquisition of endemic strains in their countries of origin could account for some clustering (33). However, patients with probable recent TB had risk factors similar to those of patients with known recent TB. The most likely explanation for most clustered patients in this group is that existing epidemiologic links remained unidentified by contact or cluster investigations, and that some had casual exposures to their source patients in unidentified settings.

Patients with reactivated disease were rarely misclassified. Among clustered patients with histories of old infection, disease, or both, our extensive review revealed no source patients. In addition, as of July 2002, we found no instances of exogenous reinfection by a different *M. tuberculosis* strain even among HIV-infected patients. Because genotyping was

Table 4. Comparison of selected risk-group characteristics of 114 recent tuberculosis (TB) patients by method of source patient identification

Characteristic	Total recent TB patients (n=114) (%)	Source patient identified by routine contact investigation (n=72) (%)	Source patient identified by cluster investigation (n=42) (%)	p value
Residence				
Baltimore city	56 (49.0)	33 (45.8)	23 (54.8)	0.38
Other state jurisdictions	58 (51.0)	39 (54.2)	19 (45.2)	
Country of birth				
United States	104 (91.3)	63 (87.5)	41 (97.6)	0.07
Other	10 (8.7)	9 (12.5)	1 (2.4)	
Long-term care resident				
Yes	7 (6.1)	3 (4.2)	4 (9.5)	0.25
No	107 (93.9)	69 (95.8)	38 (90.5)	
Homeless				
Yes	18 (15.8)	6 (8.3)	12 (28.6)	0.004
No	96 (84.2)	66 (91.7)	30 (71.4)	
Prison resident				
Yes	13 (11.4)	8 (11.1)	5 (11.9)	0.86
No	101 (88.6)	64 (88.9)	37 (88.1)	
Abuses alcohol				
Yes	40 (35.0)	19 (26.4)	21 (50.0)	0.01
No	74 (65.0)	53 (73.6)	21 (50.0)	
Uses injection drugs				
Yes	17 (14.9)	11 (15.3)	6 (14.3)	0.89
No	97 (85.1)	61 (84.7)	36 (85.7)	
Uses noninjection drugs				
Yes	35 (30.7)	23 (31.9)	12 (28.6)	0.71
No	79 (69.3)	49 (68.1)	30 (71.4)	
HIV-infected				
Yes	28 (24.6)	16 (22.2)	12 (28.6)	0.45
No	86 (75.4)	56 (77.8)	30 (71.4)	

not conducted in adjoining states, we could not eliminate the possibility of cross-jurisdictional transmission to patients who had unique *M. tuberculosis* strains. Recent TB was transmitted from three patients in Washington, D.C., to four Maryland residents (DHMH, unpub. data, 2001); disease incidence is greater in Washington, D.C., than in Maryland (14.9 vs. 5.3 per 100,000 population, respectively, in 2000) (34). Only 13% of TB patients resided in rural counties that form most of Maryland's border. With low incidence in adjacent Delaware, Pennsylvania, Virginia, and West Virginia (3.6, 3.1, 4.1, and 1.8 patients per 100,000 population, respectively, in 2000) (34), transmission between states was probably minimal.

Transmission to and from Foreign-Born Persons

Most recent transmission occurred between U.S.-born persons who had at least one common urban risk factor such as HIV infection or illegal drug use. In contrast, transmission between foreign-born persons occurred exclusively in households among persons who had no other risk except their arrival from a high-incidence country of origin or close exposure to their source patient. We found no instances of transmission between U.S.-born and foreign-born persons. These results differed from other studies, which reported that foreign-born patients who acquired TB from U.S.-born sources shared risks such as homelessness, HIV infection, and illegal drug use with those source patients (35,36). In the past decade, few immigrants and refugees settled in the city of Baltimore where urban risks are common (10). From 1996 through 2000, only 36 (9%) of 423 Baltimore patients were foreign-born compared with 642 (57%) of 1,120 patients in other Maryland areas (DHMH, unpub. data, 2001). In general, foreign migration to Maryland is relatively new (10), and we may observe more shared risks among U.S.- and foreign-born patients as time of residence increases. This study is unique in reporting that infectious foreign-born sources to U.S.-born persons primarily transmitted the disease to persons whose only risk was exposure in their workplace or a public setting, such as a church or school.

Identification of Recent Transmission before and after Genotyping

In spite of a recommended concentric circle approach for routine contact investigations that includes leisure and social locations (37), we found that investigations usually had been conducted in the homes of patients and rarely extended beyond friends and relatives. Nonetheless, the high proportion of recent patients who acquired TB in these traditional settings clearly represented numerous missed opportunities for disease prevention. Although recent TB was diagnosed among some patients during the initial contact investigation, not all identified contacts had received postexposure tuberculin skin tests or treatment for latent infection (25). More timely and complete contact investigations could reduce the risk for transmission in traditional settings.

Perhaps more importantly, almost half of recent TB cases were acquired in nontraditional settings. Many of these patients were from marginalized groups at high risk, who may have been reluctant or unable to provide names of their associates to contact investigators. However, cluster and contact investigations were equally effective in identifying sources for patients with recent TB who were illegal drug users, incarcerated, and HIV-infected, and more aggressive contact investigations would probably not substantially improve patient reporting. Instead, our data suggest that the setting, and not the risk group, eludes routine contact investigators.

In addition, TB genotyping and cluster investigations indicated unsuspected transmission to immunocompetent persons in public locations such as churches, hospitals, and stores. In these instances, the possibility of casual transmission must be considered. Casual transmission was likely in the store outbreak (28) and conceivably could account for some patients with probable recent TB for whom epidemiologic links were not found. Rarely reported, casual transmission occurs when the bacterial load of the source patient is high, the infecting organism has inherent increased virulence, or the environment is enclosed (28,38). Without creative intervention, the proportional contribution of casual transmission will increase substantially as the disease incidence decreases.

Delayed Diagnosis Among Transmitters

The mean time between reported symptom onset and initiation of treatment among transmitters was twice that identified for nontransmitters. Whether treatment delays are due to patients who delay in seeking care or to providers who do not include TB in the differential diagnosis, treatment delays provide ample time for pulmonary TB patients to develop smear-positive disease and cavitation (39–41). Our findings led to a study to determine what time period defines a diagnostic delay and to identify related client and provider factors that will guide future program interventions (42).

Conclusion

Even with excellent treatment indices, one sixth of Maryland's patients with positive cultures had recent or probable

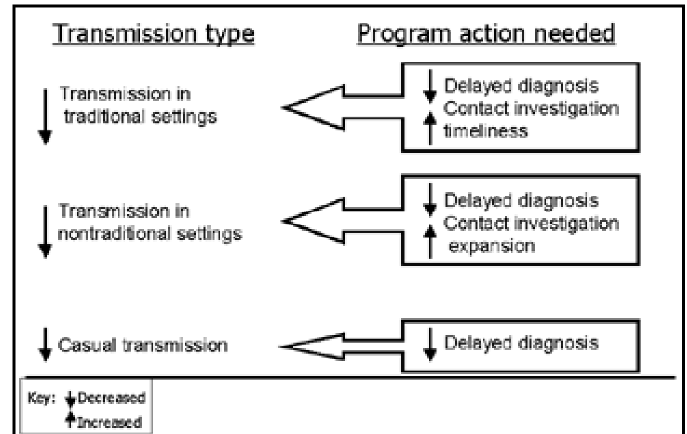

Figure. Actions needed to decrease recent tuberculosis transmission in various settings. Decreasing diagnostic delays can potentially eliminate large point source clusters and substantially reduce transmission in both traditional and nontraditional settings. CI, contact investigation.

recent disease. The new guidelines for targeted testing and treatment for latent TB infection (3) will require time and substantial resources for successful implementation, and more practical and timely interventions are needed to minimize TB transmission. In the figure, we summarize the program activities that are needed to reduce transmission from infectious TB patients in the various scenarios described in this article. Program implications include the need for improved contact investigations tailored more carefully to each patient's particular situation, with increased emphasis on activities and patient contacts outside the immediate household. However, contact investigations cannot fully address the problem of transmission in nontraditional settings. Decreasing diagnostic delays can potentially eliminate large point source clusters and substantially reduce transmission in both traditional and nontraditional settings. This method may be the only way to prevent casual transmission. Additional molecular epidemiologic investigations are needed to determine the importance of casual transmission, clarify the importance of clustered patients without links to other patients, and evaluate the long-term effectiveness of new program interventions.

Acknowledgments

We thank Sarah Bur, Richard E. Chaisson, and Timothy R. Sterling for their thoughtful reviews of this manuscript, Bianca Oden for her work on the figures, and the Division of Tuberculosis Elimination, Centers for Disease Control and Prevention, for supporting the Maryland statewide molecular epidemiology project (U52-CCU300500).

Dr. Cronin is an epidemiologist with the Division of Tuberculosis Control, Refugee and Migrant Health, Maryland Department of Health and Mental Hygiene, and currently the coprincipal investigator for the Maryland site of the Tuberculosis Epidemiologic Studies Consortium, supported by the Centers for Disease Control and Prevention. Her primary research interest includes epidemiology, particularly molecular epidemiology related to the prevention of tuberculosis and nosocomial infections.

References

1. Dye C, Scheele S, Dolin P, Pathania MC. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. JAMA 1999;282:677–86.
2. Institute of Medicine. Ending neglect: the elimination of tuberculosis in the United States. Washington: National Academy Press; 2000.
3. Centers for Disease Control and Prevention. Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Morb Mortal Wkly Rep 2000;49 (No. RR-6).
4. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiological methods. N Engl J Med 1994;300:1710–6.
5. Small PM, Hopewell PC, Singhe SP, Paz A, Parsonnet J, Ruston DC, 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.
6. Barnes PF, Yang Z, Preston-Martin S, Pogoda JM, Jones BE, Otaya M, et al. Patterns of tuberculosis transmission in Central Los Angeles. JAMA 1997;278:1159–63.
7. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, 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.
8. Burman WJ, Reves RR, Hawkes AP. DNA fingerprinting with two probes decreases clustering of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:1140–6.
9. Bishai WR, Graham NMH, Harrington S, Pope DS, Hooper N, Astemborski J, et al. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 1998;280:1679–84.
10. Maryland Office for New Americans. Immigration to Maryland, then and now. New American 2000;6:4–5.
11. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2001. Atlanta: Department of Health and Human Services, 2002.
12. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenbach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
13. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Sooligen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
14. Das S, Chan SL, Allen BW, Mitchison DA, Lowrie DB. Application of DNA fingerprinting with IS*986* [IS*6110*] to sequential mycobacterial isolates obtained from pulmonary tuberculosis patients in Hong Kong before, during, and after short-course therapy. Tuber Lung Dis 1993;74:47–51.
15. Cave MD, Eisenbach KD, Templeton G, Salfinger M, Mazurek G, Bates JH, et al. Stability of DNA genotype pattern produced with IS*6110* in strains of *Mycobacterium tuberculosis*. J Clin Microbiol 1994;32:262–6.
16. Braden CR, Templeton GL, Stead WW, Bates JH, Cave MD, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
17. Burman WJ, Reves RR. Review of false-positive cultures for *Mycobacterium tuberculosis* and recommendations for avoiding unnecessary treatment. Clin Infect Dis 2000;31:1390–5.
18. Nardell E, McInnis B, Thomas B, Weidhaus S. Exogenous reinfection with tuberculosis in a shelter for the homeless. N Engl J Med 1986;315:1570–5.
19. van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, et al. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. N Engl J Med 1999;341:1174–9.
20. Centers for Disease Control and Prevention. Tuberculosis information management system (TIMS) user's guide; surveillance appendix I. Atlanta: Centers for Disease Control and Prevention; 1998.
21. Cronin WA, Magder LS, Baruch NG, Strickland GT, Bishai WR. Source identification for pediatric tuberculosis cases in Maryland. In: Abstracts of the 7th Annual North American Regional Conference of the International Union Against Tuberculosis and Lung Disease; Vancouver, Canada: March 2002. Vancouver, BC: British Columbia Lung Association;2002.
22. Cronin WA, Golub JE, Magder LS, Baruch NG, Lathan MJ, Mukasa LN, et al. Usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J Clin Microbiol 2001;39:3709–11.
23. Michele TM, Cronin WA, Graham NHM, Dwyer DM, Pope DS, Harrington S, et al. Transmission of *Mycobacterium tuberculosis* by a fiberoptic bronchoscope: identification by DNA fingerprinting. JAMA 1997;278:1093–5.
24. Lathan M, Cronin W, Mukasa L, Hooper D, Benjamin W. Outbreak of tuberculosis at a hospital identified by DNA fingerprinting. In: Abstracts of the National Tuberculosis Genotyping and Surveillance Network Annual Meeting; San Francisco, California, July 1999. Atlanta: Centers for Disease Control and Prevention; 1999.
25. Cronin WA, Lathan MJ, Benjamin WH. Unusually high rates of tuberculosis infection and disease among oyster fishery workers in a rural community (abstract). Int J Tuberc Lung Dis 1998;11:S281.
26. Sterling TR, Stanley RL, Brubach GA, Madison A, Harrington S, Bishai WR, et al. HIV-related tuberculosis in a transgender network—Baltimore, Maryland, and New York City Area. JAMA 2000;283:2515–6.
27. Bur S, Golub J, Armstrong J, Johnson B, Maltas G, Barker L, et al. Anatomy of a large community and jail tuberculosis outbreak investigation. In: Abstracts of the 98th American Thoracic Society International Conference; San Francisco, California, April 2001. New York: American Thoracic Society; 2001.
28. Golub JE, Cronin WA, Obasanjo OO, Coggin W, Moore K, Pope DS, et al. Transmission of *Mycobacterium tuberculosis* through casual contact with an infectious patient. Arch Intern Med 2001;161:2254–8.
29. Sterling TR, Pope DS, Bishai WR, Harrington S, Gershon RR, Chaisson RE. Transmission of *Mycobacterium tuberculosis* from a cadaver to an embalmer. N Engl J Med 2000;342:246–8.
30. Glynn JR, Vynnycky E, Fine PEM. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am J Epidemiol 1999;149:366–71.
31. Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol 2002;155:565–71.
32. Murray M, Nardell E. Molecular epidemiology of tuberculosis: achievements and challenges to current knowledge. Bull WHO 2002;80:477–82.
33. Hermans PWM, Messadi F, Guebrexabher H, van Sooligan D, Haas PEW, Heersman H, et al. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J Infect Dis 1995;171:1504–13.
34. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2000. Atlanta: Department of Health and Human Services; 2001.
35. Chin DP, DeReimer K, Small PM, Ponce de Leon A, Steinhart R, Schecter GF, et al. Differences in contributing factors to tuberculosis incidence in US-born and foreign-born persons. Am J Respir Crit Care Med 1998;158:1797–803.
36. Jasmer RM, Ponce de Leon A, Hopewell PC, Alarcon RG, Moss AR, Paz EA, et al. Tuberculosis in Mexican-born persons in San Francisco: reactivation, acquired infection and transmission. Int J Tuberc Lung Dis 1997;1:536–41.
37. Etkind SC. Contact tracing. In: Reichman L, Hershfield E, editors. TB: a comprehensive international approach. New York: Marcel Dekker, Inc.; 1993. p. 275–89.

38. Valway SE, Sanchez MPC, Shinnick TF, Orme I, Agerton T, Hoy D, et al. An outbreak of extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med 1998;338:633–9.
39. Asch S, Leake B, Anderson R, Gelberg L. Why do symptomatic patients delay obtaining care for tuberculosis? Am J Respir Crit Care Med 1998;157:1244–8.
40. Mather P, Sacks L, Auten G, Sall R, Levy C, Gordin F. Delayed diagnosis of pulmonary TB in city hospitals. Arch Intern Med 1994;154:306–10.
41. Chin DP, Crane CM, Diul MY, Sun SJ, Agraz R, Taylor RS, et al. Spread of *Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. JAMA 2000; 283:2968–74.
42. Golub J, Bur S, Cronin W, Baruch N, Comstock G, Chaisson R. Patient and health care delays in the diagnosis of Maryland tuberculosis patients. In: Abstracts of the 7th Annual North American Regional Conference of the International Union Against Tuberculosis and Lung Disease; Vancouver, Canada: March 2002. Vancouver, BC: British Columbia Lung Association; 2002.

Address for correspondence: Wendy A. Cronin, Division of Tuberculosis Control, Refugee and Migrant Health, Maryland Department of Health and Mental Hygiene, 201 W. Preston St., Room 307A, Baltimore, MD 21201, USA; fax: 410-669-4215; e-mail: croninw@dhmh.state.md.us

ScholarOne™
Web-Based Solutions for Scholarly Publishers

EMERGING INFECTIOUS DISEASES
A Peer-Reviewed Journal Tracking and Analyzing Disease Trends

Emerging Infectious Diseases to launch Web-based manuscript submission and peer review in 2003

Emerging Infectious Diseases will be launching **Manuscript Central**, a Web-based system for manuscript submission and peer review. Manuscript Central is operated by **ScholarOne**, a software company that specializes in scholarly publishing applications. The system allows authors, reviewers, editors, and editorial office staff direct access to journal operations through the Web.

Digital workflow expedites peer-review and administrative processes

- Allows manuscripts and reviews to be transmitted electronically between authors, the journal office, editors, and reviewers around the world
- Enables authors to submit their manuscripts as Microsoft Word, RTF, PostScript, HTML, and PDF files
- Accepts a wide variety of graphic file types and multimedia formats
- Makes manuscript files accessible to reviewers and editors through the journal's Manuscript Central site

Digital workflow eliminates mailing costs and has many other benefits

- Encourages global participation
- Enables authors to check the status of their manuscripts or update their contact information
- Allows editors direct access to an extensive database from which to select reviewers—by name, areas of expertise, key words

Built-in security ensures that only editors and reviewers assigned to a particular manuscript can view that manuscript. Reviewers, whose identities are kept confidential, have access to the manuscript only until they have completed and submitted their reviews.

Emerging Infectious Diseases, now in its 8th year of publication, welcomes submissions from infectious disease specialists in academia, industry, clinical practice, and public health, as well as from specialists in economics, demography, sociology, and other disciplines. For information on the types of articles published in Emerging Infectious Diseases, see http://www.cdc.gov/ncidod/eid/instruct.htm.

More detailed information on electronic submission and peer review will be published in upcoming issues of Emerging Infectious Diseases and on the Web at www.cdc.gov/eid.

Skin-Test Screening and Tuberculosis Transmission among the Homeless[1]

Po-Marn Kong,* Jan Tapy,* Patricia Calixto,* William J. Burman,*† Randall R. Reves,*† Zhenhua Yang,‡§ and M. Donald Cave‡§

We describe the implementation of a mandatory tuberculosis (TB) screening program that uses symptom screening and tuberculin skin testing in homeless shelters. We used the results of DNA fingerprinting of *Mycobacterium tuberculosis* isolates to evaluate the effect of the program on TB incidence and transmission. After the program was implemented, the proportion of cases among homeless persons detected by screening activities increased, and the estimated TB incidence decreased from 510 to 121 cases per 100,000 population per year. Recent transmission, defined by DNA fingerprinting analysis as clustered patterns occurring within 2 years, decreased from 49% to 14% (p=0.03). Our results suggest that the shelter-based screening program decreased the incidence of TB by decreasing its transmission among the homeless.

Homelessness is one of the greatest risk factors for contracting tuberculosis (TB) in the United States (1,2). The incidence in the general population is extremely low; however, the incidence among the homeless, particularly among minority ethnic groups, can be as high as 450 cases per 100,000 population (3). In a previous study, homelessness was the key factor associated with DNA fingerprint clustering, suggesting that ongoing transmission of TB in Denver, Colorado, occurs primarily among the homeless (4).

Several studies suggest that active case finding may decrease transmission and the overall case rate (2,5,6). We describe the implementation of a mandatory screening program consisting of symptom screening and tuberculin skin testing in homeless shelters and evaluate its effect on TB incidence and transmission.

Methods

Description of Screening Program

The Denver Metro TB Clinic of Denver Public Health Department provides TB control activities for metropolitan Denver, Colorado, which includes the counties of Adams, Arapahoe, Denver, Douglas, and Jefferson. Active TB is treated by use of directly observed therapy in the clinic or in the community; outside the clinic, outreach workers deliver the therapy using incentives and enablers (i.e., transport, food coupons, temporary housing, and treatment for substance abuse or mental illness). Temporary housing for the homeless in Denver is available in large communal shelters, through residential drug and alcohol treatment programs, and in individual or family transitional housing. Our program focused on the large communal shelters and the drug and alcohol treatment programs because these settings were the likely loci of TB transmission.

The screening program included a simple symptom assessment and a tuberculin skin test for persons who had no history of a positive skin test. The symptom assessment was a brief interview to detect TB symptoms such as chronic cough, weight loss, night sweats, and hemoptysis.

We used the Mantoux method of tuberculin skin testing (6) and defined a positive result as induration of >10 mm. Patients with symptoms that suggested active TB (regardless of tuberculin skin-test results) and those with a previous or current positive skin test were sent to the Denver Metro TB Clinic for further evaluation and chest radiograph. To stay in a homeless shelter or residential substance abuse treatment program for more than 3 to 5 days, patients were required to complete the screening. All persons with active TB were treated with directly observed therapy, and persons with latent TB infection who met criteria for treatment were offered directly observed preventive treatment. After completion of screening, homeless persons were given a card from the clinic that allowed them to remain in one of the shelters for up to 6 months.

The screening program, started on a voluntary basis in 1993 at a day shelter and a clinic for the homeless, was expanded to four large shelters and six residential drug and alcohol treatment programs by 1995. One experienced public health nurse and four assistant outreach workers staff the program 15 hours each week, Monday to Friday. They provide tuberculin skin testing at some overnight shelters and at a day

*Denver Public Health, Denver, Colorado, USA; †University of Colorado Health Sciences Center, Denver, Colorado, USA; ‡Regional Tuberculosis Genotyping Laboratory, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA; and §Central Arkansas Veterans Healthcare System, Little Rock, Arkansas, USA

[1]This study was presented in part at the 2001 International Meeting of the American Thoracic Society, San Francisco, California.

shelter and a clinic for the homeless. In our analysis of the program, we included all persons identified as homeless or in the residential drug/alcohol treatment programs between January 1995 and December 1998.

Data Analysis

Persons with active TB were considered homeless if they lived in one of the shelters or residential substance-treatment centers or were reported as homeless in clinic records between 1988 and 1998. Data were available for 1995 to 1998 on whether TB diagnoses were detected by the screening program or by staff at a medical facility. We obtained the numbers of homeless persons from studies done in 1995 and 1998 by the Colorado Department of Human Services (7). These studies enumerated the persons living in shelters and residential substance treatment centers and also estimated the number of persons living in other temporary housing and on the streets. Data for homeless populations in 1996 and 1997 were not available, so we estimated this number, assuming a linear increase in the total. We analyzed TB incidence for the 11-year period, from 1988 (the year that isolates were first available for DNA fingerprinting) through 1998, compiling data for 7 years before and 4 years after the screening program was implemented for the homeless in 1995.

We analyzed identical fingerprinting patterns in homeless patients from 1988 through 1998. Initial DNA fingerprinting used the IS*6110* technique, and secondary fingerprinting used pTBN12 for isolates having five or fewer hybridizing bands (4,8). To estimate recent TB transmission in the homeless population, we calculated the proportion of cases clustered within a 2-year period of a preceding case with the identical DNA fingerprint (9). We compared the proportion of cases caused by recent transmission during the 7-year period before the program (1988–1994) with the proportion caused during the first 4 years of the program (1995–1998) to assess the effect of the screening program on TB transmission.

Analyses were done by using Epi Info, version 6.0 (Centers for Disease Control and Prevention, Atlanta, GA). The proportion clustered in the two time periods was compared by using the Fisher exact test.

Results

Screening Program Evaluation

The estimated number of homeless persons in Denver, Colorado, increased from 3,330 in 1995 to 5,792 in 1998. This increase was largely attributed to the combination of population growth in the metropolitan area and rising housing costs (7). The four large communal shelters could accommodate ≤1,176 persons and the transitional housing and treatment programs ≤980 persons. The number of homeless persons who had tuberculin skin tests increased from 893 in 1995 to 3,897 by 1998 (Table 1). The screening ratio (number of completed skin tests divided by the estimated population) increased from 26.7 to 67.3 per 100 persons during this period. The proportion of persons with positive tuberculin skin tests decreased during the study period (17% in 1995, 12% in 1998, $p<0.01$) (Table 1). Only those persons not already known to be positive were tested.

Table 1. Effect of screening program on tuberculin skin testing and treatment of latent tuberculosis among the homeless, Denver Health Tuberculosis Clinic, 1995–1998

Year	Tuberculin testing ratio[a]	Tuberculin skin-test positivity (%)	Completion of treatment of latent tuberculosis infection (%)
1995	893/3,350 (26.7)	150/893 (17)	7/37 (18.9)
1996	1,799/4,164 (43.2)	267/1,799 (15)	24/83 (28.9)
1997	3,438/4,978 (69.1)	397/3,438 (12)	24/93 (25.8)
1998	3,897/5,792 (67.3)	470/3,897 (12)	22/59 (37.3)

[a]Completed tuberculin tests divided by estimated homeless population (tests per 100 persons).

Active TB was diagnosed in 94 homeless persons during the 11-year period from 1988 to 1998. Of those, 87 (93%) had positive cultures. The number of confirmed cases increased to a peak of 17 cases in 1995 and then decreased to 7 during each of the next 3 years (Table 2). When the increase in the number of homeless persons during this period is considered, the estimated incidence of active TB decreased from 510 to 121 per 100,000 persons from 1995 to 1998.

The screening program was more successful in early identification of TB cases than in treatment of latent TB infection, which had low acceptance and completion rates (Table 1) that did not change substantially from 1995 to 1998. Over the 4-year period, 1,284 positive tuberculin skin tests were recorded, but only 272 homeless persons initiated isoniazid treatment; of those, 77 (28%) completed therapy. Five (29%) of the 17 confirmed cases of TB diagnosed in 1995 were detected through the screening program. Of seven TB cases diagnosed during each of the 3 subsequent years (1996–1998), the screening detected three, five and four cases, respectively, for a mean of 57%.

DNA Fingerprinting Analysis

DNA fingerprinting results were available for 76 (87%) of the 87 culture-positive cases (Table 2). Isolates from the 76 patients demonstrated 51 different DNA fingerprinting patterns. Nine clusters, in which identical patterns occurred at any time during the study period, consisting of 2–12 patients included 34 (45%) of the 76 patients. Clusters in which cases occurred within 2 years of each other were present in six DNA fingerprinting patterns and accounted for 27 (36%). In three other clusters, occurrences of TB in patients were separated by >2 years. DNA fingerprinting patterns unique in the homeless population were found in the remaining 42 patients' isolates, 9 of which had patterns also found in nonhomeless TB patients in the community. Of these, we detected one matching set of isolates for each of eight patterns in nonhomeless patients; one case in a homeless patient was followed 3 years later with three cases in nonhomeless persons who had isolates of the

Table 2. DNA fingerprinting results for culture-confirmed tuberculosis cases among homeless persons, Denver Health Tuberculosis Clinic, 1988–1998[a]

	Years											
RFLP clusters	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	Total
A			1					1				2
B		1		[1][b]	[2]	1						2 [3]
C	1						1				1	3
D				1		1						2
E				1	1							2
F								2				2
G				1		1				[2]		2 [2]
H			1	3 [1]	2		1			[1]		7 [2]
I			2	3	1 [2]	3 [1]	1	2 [2]	[2]			12 [7]
Total clusters	1	1	4	9 [2]	4 [4]	6 [1]	3	5 [2]	0 [2]	0 [3]	1	34 [14]
Unique RFLP	1	1	3	2	3	6	3	7	5	6	5	42
Test ratio[c]	2/4 (50)	2/2 (100)	7/7 (100)	11/11 (100)	7/7 (100)	12/12 (100)	6/6 (100)	12/17 (71)	5/7 (71)	6/7 (86)	6/7 (86)	76/87 (87)

[a]RFLP, restriction fragment length polymorphism.
[b]Numbers in brackets refer to the 14 case isolates from nonhomeless patients that share DNA fingerprinting patterns with the homeless patients.
[c]Number of case isolates with DNA fingerprinting results divided by total culture-positive cases (% tested).

identical DNA fingerprint type (data not shown). Among the homeless patients, persons born outside the United States constituted 7 (17%) of the 42 cases with unique isolates and 3 (9%) of the 34 cases with clustered isolates.

The estimated proportion of TB cases resulting from recent transmission within the homeless population, defined as cases that were clustered within 2 years, decreased from 49% (23/47) in the 7-year period (1988–1994) before the screening program was implemented to 14% (4/29) in the 4-year period (1995–1998) after the program was implemented (p=0.03).

DNA fingerprinting results were available for 272 (55%) of 491 of culture-positive cases among persons who were not homeless. The proportion of nonhomeless case isolates with DNA fingerprinting results increased from 47% (167/353) to 53% (112/210) for the years 1988–1994 and 1995–1998, respectively. Case isolates in the nonhomeless population showed 233 DNA fingerprinting patterns with 24 clusters of two to seven patients. In four of the clusters among homeless persons, cases in homeless persons were followed by two to seven cases with the same DNA fingerprinting pattern in nonhomeless persons (Table 2). The largest cluster (I) originated with a case diagnosed in a homeless person in Denver who had stopped treatment in another state in 1990; the same strain was detected in 11 additional cases in homeless persons over a 6-year period and was also detected in 7 cases in nonhomeless persons, the first 2 of which were nosocomially acquired in 1992.

Discussion

The institution of a TB-screening program for the homeless population in congregate settings in Denver in 1995 was associated with a substantial decrease in the incidence of active TB in this high-risk population. The mechanisms by which the program could have caused this reduction in TB incidence include treating active cases earlier and treating persons with latent TB infection, thus preventing future active TB cases. Treating persons with latent TB infection was unlikely to have played a major role, however, since we found low rates of starting and completing treatment for this type of infection, a finding noted in previous reports (1,10).

The screening program more likely reduced TB transmission by earlier diagnosis and treatment of active TB cases (the top priority activity in TB control). The program was effective in finding TB patients, detecting over half the active TB cases reported in homeless persons from 1996 to 1998. In contrast to the low acceptance and completion of latent TB infection treatment, high rates of treatment completion for active TB cases can be achieved with directly observed therapy supported by public health regulations (11,12). The results of the DNA fingerprint analysis are consistent with the conclusion that early detection of TB through the screening program leads to a reduction in transmission. New, unique strains of *M. tuberculosis* have continued to appear among the Denver homeless population, but cases from recent transmission (clustered in the homeless population within 2 years) decreased.

The estimated number of active TB cases decreased despite an increasing population of homeless persons, leading to a drop in the annual incidence rate from 510 to 121 per 100,000 persons.

Other potential explanations for the reduction in TB cases in the homeless population include failure to detect subsequent cases and errors in population estimates. The mandatory TB screening program for residents in shelters and residential substance treatment programs may have led some homeless persons to live in other settings or to leave Denver; the latter reason led to the diagnosis of subsequent cases in other jurisdictions. However, the Denver Metro TB Clinic receives reports of cases diagnosed in Metropolitan Denver and conveys them to the TB Control Program at the Colorado Department of Public Health and Environment; cases were not likely missed unless diagnosed outside Colorado. Obtaining an accurate count of homeless persons is difficult so we based our denominator in this study upon estimates obtained by the Colorado Department of Housing (7). Even if the population at risk had remained the same, despite indications otherwise, the incidence rate would still have decreased from 510 to 210 (rather than 121) per 100,000 persons per year from 1995 to 1998. We detected substantial TB transmission in the homeless population for only two of nine homeless clusters, raising the possibility that the increase in cases was the result of one or two outbreaks running their natural course. Further studies with larger numbers of cases are needed to address this issue; however, our data support the concept that early case detection results in lower rates of transmission in homeless populations.

The possibility exists that clusters that occurred over several years were strains that lingered in the community but were not detected in our study; we obtained DNA fingerprints for only 55% of the cases in the community among nonhomeless persons and cannot rule out that possibility. For clustered cases occurring many years apart, we may have missed links in transmission because of patients who were not tested in the community or who moved out of the area without diagnosis. As we increase our understanding of how to interpret the results of DNA fingerprinting, we may find that clusters that are far apart represent a remote transmission that resulted in latent TB infection with later reactivation. The secondary cases would then share the same pattern as the original (source) case, but the connection between them would be more difficult to identify since the contact in these cases occurred many years earlier.

Of the 42 homeless cases with unique patterns, seven (17%) were diagnosed in foreign-born persons. Of those, one patient arrived in United States <2 years before TB diagnosis, three 2–5 years before diagnosis, two 5–10 years before diagnosis, and one for an unknown period of time. Few of the unique patterns found among homeless patients represent cases imported from foreign countries.

Various screening programs have been proposed to decrease TB transmission among homeless persons (13). Chest radiographic screening provides a rapid assessment for potentially transmissible disease, but this method would have been too expensive for us to implement. Sputum examination by smear has been used (14) but does not detect smear-negative cases and is difficult to implement for large numbers of persons. Tuberculin skin testing does not provide an immediate assessment of infectious TB and gives negative test results for up to 25% of persons with active pulmonary TB. Furthermore, skin testing requires a second visit 2–3 days later for a reading.

Despite reaching only two thirds of the homeless population in 1998, the screening program was notably successful. The true incidence of skin test positivity is likely higher than that shown in Table 1 because those already known to be positive were not tested. The symptom survey was an important component of our program. In spite of the challenges of skin testing, our screening program proved to be feasible for a large population staying at a number of shelters and residential programs. Our results suggest that congregate settings, the focus of our program, are the most important loci of TB transmission among homeless persons.

The low completion rate of treatment for latent TB infection provided further evidence that the primary effect of the program was to increase early detection of active cases of TB, not to identify and treat latent TB infection. Despite the decrease in TB incidence following the screening program, the TB case rate remained high (121 per 100,000 population) compared to the overall rate for the metropolitan area (8.2 per 100,000 population). Further progress in decreasing TB among homeless persons will require improvements in the acceptance and completion of latent infection treatment (6).

Acknowledgments

The authors acknowledge the assistance of Lourdes Yun in the analysis and review of the data.

The work described in this report was supported by the Veterans Administration and the Centers for Disease Control and Prevention Interagency Agreement 98FED10318. Resources and facilities were used at the Central Arkansas Veterans Health Services Center in Little Rock, Arkansas.

Dr. Kong is currently a consultant respiratory physician at the Department of Respiratory Medicine and Tuberculosis Control Unit, Tan Tock Seng Hospital, Singapore. The work for this study was done at Denver Public Health in Denver, Colorado. His research interests include tuberculosis epidemiology and pulmonary disease caused by atypical mycobacteria.

References

1. Barnes PF, Yang Z, Preston-Martin S, Pogoda JM, Jones BE, Otaya M, et al. Patterns of tuberculosis transmission in central Los Angeles. JAMA 1997;278:1159–63.
2. Nolan CM, Elarth AM, Barr H, Saeed AM, Risser DR. An outbreak of tuberculosis in a shelter for homeless men: a description of its evolution and control. Am Rev Respir Dis 1991;143:257–61.
3. Moss AR, Hahn JA, Tulsky JP, Daley CL, Small PM, Hopewell PC. Tuberculosis in the homeless: a prospective study. Am J Respir Crit Care Med 2000;162:460–4.

4. Burman WJ, Reves RR, Hawkes AP, Reitmeijer CA, Yang Z, El-Hajj H, et al. DNA fingerprinting with two probes decreases clustering of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:1140–6.
5. Rendleman NJ. Mandated tuberculosis screening in a community of homeless people. Am J Prev Med 1999;17:108–13.
6. American Thoracic Society and Centers for Disease Control and Prevention. Targeted testing and treatment of latent tuberculosis infection. Am J Respir Crit Care Med 2000;161:S221–47.
7. D'Alanno TA. Homelessness in the Denver Metropolitan Area: a base line point in time study, June 15, 1998. The Colorado Department of Human Services. Available from: URL: http://www.cdhs.state.co.us/shhp/
8. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
9. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991-1997. Ann Intern Med 1999;130:971–8.
10. Tulsky JP, Pilote L, Hahn JA, Zolopa AJ, Burke M, Chesney M, et al. Adherence to isoniazid prophylaxis in the homeless: a randomized controlled trial. Arch Intern Med 2000;160:697–702.
11. Burman WJ, Cohn DL, Rietmeijer CA, Judson FN, Sbarbaro JA, Reves RR. Noncompliance with directly-observed therapy for tuberculosis: epidemiology and effect on the outcome of treatment. Chest 1997;111:1168–73.
12. Burman WJ, Cohn DL, Rietmeijer CA, Judson FN, Sbarbaro JA, Reves RR. Short-term incarceration for the management of noncompliance with tuberculosis treatment. Chest 1997;112:57–62.
13. Centers for Disease Control and Prevention. Prevention and control of tuberculosis among homeless persons. Recommendations of the Advisory Council for the Elimination of Tuberculosis. MMWR Recomm Rep 1992;41(RR-5):13–23.
14. Kimerling ME, Shakes CF, Carlisle R, Lok KH, Benjamin WH, Dunlap NE. Spot sputum screening: evaluation of an intervention in two homeless shelters. Int J Tuberc Lung Dis 1999;3:613–9.

Address for correspondence: Kong Po Marn, Dept. of Respiratory Medicine, Tan Tock Seng Hospital,11, Jalan Tan Tock Seng, Singapore 308433, Republic of Singapore; fax: 65-6357-7871; e-mail: pmkong@singnet.com.sg

"Everything should be made as simple as possible, but not simpler."
–Albert Einstein

Impact of Genotyping of *Mycobacterium tuberculosis* on Public Health Practice in Massachusetts

Ann C. Miller,* Sharon Sharnprapai,* Robert Suruki,* Edward Corkren,* Edward A. Nardell,*† Jeffrey R. Driscoll,‡ Michael McGarry,‡ Harry Taber,‡ and Sue Etkind*

Massachusetts was one of seven sentinel surveillance sites in the National Tuberculosis Genotyping and Surveillance Network. From 1996 through 2000, isolates from new patients with tuberculosis (TB) underwent genotyping. We describe the impact that genotyping had on public health practice in Massachusetts and some limitations of the technique. Through genotyping, we explored the dynamics of TB outbreaks, investigated laboratory cross-contamination, and identified *Mycobacterium tuberculosis* strains, transmission sites, and accurate epidemiologic links. Genotyping should be used with epidemiologic follow-up to identify how resources can best be allocated to investigate genotypic findings.

Genotyping of *Mycobacterium tuberculosis* isolates is accepted as a useful tool in many public health settings (1–13). Conventional epidemiology may be supported by analysis of isolates that cluster genetically by using IS*6110* fingerprinting methods. Genotyping can confirm or disprove previously known epidemiologic connections or suggest unsuspected associations. We explore the impact that genotyping has had on public health practice in Massachusetts by investigating the following premises. Genotyping plays an indirect role in interrupting the transmission of tuberculosis (TB) by identifying unexpected epidemiologic links and unsuspected sites of transmission. By identifying and confirming outbreaks and their impact and aiding the investigation of laboratory cross-contamination, genotyping determines whether epidemiologic links found through conventional contact investigations represent true cases of transmission. This determination helps coordinators of TB-control programs to direct and evaluate their program activities.

Before Massachusetts become one of seven sentinel surveillance sites in the National Tuberculosis Genotyping and Surveillance Network (genotyping network), *M. tuberculosis* isolates were sent by the Massachusetts Department of Public Health, Division of Tuberculosis Prevention and Control (TB Division) to the regional laboratory (New York State Department of Health's Wadsworth Center in Albany, New York) for genotyping on an as-needed basis. During the genotyping network study (July 1, 1996–December 31, 2000), one isolate from each new patient whose culture was positive for *M. tuberculosis* was sent for genotyping (14).

*Massachusetts Department of Public Health, Jamaica Plain, Massachusetts, USA; †Harvard University, Boston, Massachusetts, USA; and ‡New York State Department of Health, Albany, New York, USA

Methods

Laboratory Methods

Cowan and Crawford (15) and Crawford et al. (16) describe the laboratory protocol and design and methods of the genotypying network. The Wadsworth Center performed IS*6110* restriction fragment length polymorphism (RFLP) and spoligotyping according to standardized procedures (16–18).

Genotype Cluster Identification

The Centers for Disease Control and Prevention (CDC) developed a standard protocol for cluster investigations. For IS*6110* strains with more than six copies in their fingerprint patterns (high-copy strains), a single cluster designation was allocated if two or more patients had isolates with identical RFLP patterns. Strains with six or fewer copies in their patterns (low-copy strains) were assigned a single cluster designation if two or more patients had both identical RFLP and spoligotype patterns. The cluster investigation by the genotyping network took place from January 1, 1998, to December 31, 2000.

Epidemiologic Investigation of Clusters

To identify sources of transmission, we defined an epidemiologic link as two persons who shared space or time. For example, if a cluster included three persons (the source [A] and two subsequent patients [B and C]), the epidemiologic relationships identified for inclusion in the database would be A to B and A to C. However, when a common source was determined from outside the study, each relationship was counted only once without reference to direction (B and C, representing one link with no reference to A). An "expected link" was one found through conventional contact investiga-

tion, and an "unexpected link" was identified as a result of an RFLP cluster investigation.

To determine epidemiologic links, we reviewed medical records from health departments, hospitals, and clinics for each genotypically clustered case. Patients or their proxies were interviewed. Collected data included medical information (i.e., tuberculin skin-test results and BCG vaccine status, potential exposures, and previous diagnoses), demographics, and a 2-year social and work history.

CDC collaborated with the genotyping network sentinel sites in developing forms for data collection, patient consent, and interviews. The Massachusetts Department of Public Health institutional review board approved the protocol and forms. The patient consent form was translated into Spanish, Haitian Creole, Portuguese, Khmer, Vietnamese, and Chinese. Interpreters were present during interviews, when needed.

Laboratory Cross-Contamination

For the period 1994–1997, laboratory cross-contamination or other error was suspected if more than one specimen processed at the same time was positive for *M. tuberculosis.* At the time, we considered that contamination was less likely to occur between specimens processed in different batches on the same day; therefore, this type of contamination was not addressed in the study. Contamination in different batches was, however, addressed in a broader 1998 genotyping network study on laboratory cross-contamination. Data and methods from that study are described elsewhere (19).

Results

Transmission Dynamics and Public Health Implications

From July 1, 1996, to December 31, 2000, a total of 1,281 TB cases were counted in Massachusetts. Of these, 1,043 (81%) were culture confirmed as TB. Positive cultures were obtained from 1,032 patients in Massachusetts, and 95% (984) of the isolates obtained from these patients were genotyped. The remaining isolates were not genotyped for various reasons, including inability to obtain growth when subcultured, contamination with other bacteria or fungi, and inability to acquire isolates from outside laboratories. Of the 984 isolates, 776 (79%) had high-copy strains (seven or more IS*6110* copies in their patterns); 712 (72%) had unique DNA fingerprints; 272 (28%) aggregated into 82 clusters. Of 208 cases with low-copy strains, 100 (48%) met the study definition of genotype cluster.

Links Established by Conventional Contact Investigation

Overall, 129 expected epidemiologic relationships in Massachusetts were identified. Of these, 37 relationships were between a person in the study and a person without a genotyped isolate (e.g., children, clinical patients). Of 92 relationships analyzed (88 persons), 67 (72%) demonstrated exact or similar (±1 hybridizing band) RFLP matches. Twenty-five relationships between 38 persons were not supported by genotyping, i.e., the isolates had different genotypes.

The 25 expected relationships not supported by genotyping were as follows: 11 in persons staying at the same homeless shelter at the same time; 5 in household contacts; 2 in nonhousehold family members, 3 in friends and social contacts, and 4 in co-workers. The five relationships involving household contacts were in persons from countries with a high prevalence of TB. These relationships included roommates from Mali, siblings from Kenya and Ethiopia, and parents and adult children from China and India.

Additional Epidemiologic Links in RFLP Cluster Investigation

In addition to the 129 epidemiologic relationships identified before RFLP results, 11 unexpected epidemiologic relationships involving 21 persons were identified. We also identified additional and some previously unrecognized places of transmission (Table). Several of these are believed to have been caused by either casual contact or unsuspected settings.

Table. Settings of transmission for unexpected epidemiologic links within genotyped clusters

Cluster designation (no. IS*6110* copies)	No. unexpected epidemiologic relationships (persons)	No. expected relationships (persons)	Settings of transmission for unexpected epidemiologic links
1 (10)	1 (2)	2 (3)	Prison
2 (7)	1 (2)	2 (3)	Neighborhood, same public housing
3 (15)	1 (2)	0 (0)	Long-term care facility
4 (8)	1 (2)	0 (0)	Fast food restaurant
5 (12)	2 (3)	0 (0)	Hair salon, college building
6 (1)	1 (2)	0 (0)	Buddhist temple
7 (5)	1 (2)	0 (0)	Community barbecue[a]
8 (8)	1 (2)	1 (2)	Bars
9 (12)	1 (2)	0 (0)	Neighborhood, same markets
10 (9)	1 (2)	0 (0)	Neighborhood

[a]Community barbecue was held at different sites.

For example, cluster 5 in the table consisted of three patients who ostensibly had nothing in common. However, RFLP cluster investigation established that one patient had been the hairdresser of the second patient, who was a college student. The third patient was a security guard who had worked in a college dormitory frequented by the second case. At the onset of the guard's illness, he was working at another facility.

Extent of an Outbreak in a Homeless Shelter

Of 18 men with TB in the homeless or associated populations in 2000, isolates from 15 persons had RFLP pattern 437 or 5309. These patterns differ by a one-copy addition. Thirteen of the men were long-term users of one shelter, although eight used the shelter sporadically over many years. Two patients had positive tuberculin skin tests in the past, and two other patients were clinical patients in past years (1997 and 1995). Eleven of the 15 patients were interviewed, and only one interview provided a definite epidemiologic link between two patients. Other interviews could not establish specific dates that the cases had overlapped. We reviewed 199 bed logs from the shelter for October 1999 through June 2000, a time period between documented negative tuberculin skin tests and diagnosis of active disease for three of the patients.

Of 14 homeless men in the cluster, 2 were not on the bed logs for the entire 9-month period. The remaining 12 patients were never all present on the same day. However, 11 of the 12 patients were present on 1 day in November, and 10 of the 12 were present on 6 nights in October and November 1999. Of 15 persons in the cluster (14 guests and 1 employee), we established that 13 were epidemiologically linked. Of the other two, we believe that one man's diagnosis resulted from a false-positive culture, and the other man arrived at the shelter after June 2000 and was likely exposed later.-

Laboratory Cross-Contamination

In April and May 1994, two related instances of cross-contamination in one laboratory were identified 3 weeks apart. Five of ten cultures processed at the same time in the first instance and three of the ten in the second were submitted for RFLP analysis; they matched a strain used as a control in the decontamination and digestion procedure. The TB Division and the laboratory sent all isolates tested 2 weeks before the first instance through 2 weeks after the second instance for RFLP typing. Clinical records were reviewed, and epidemiologic follow-up was completed on all cases. No other specimens were affected.

Discussion

We identified unexpected epidemiologic links between 21 known cases of TB (Table). However, over the course of the study, we did not unearth any new unknown infectious cases of TB by using genotyping. Two possible explanations for this lack of new cases include a historically strong programmatic interest in case finding, or conversely, the difficulty of reopening contact investigations after substantial time has elapsed.

Although the Massachusetts TB Division has always practiced the concentric-circle method of contact investigation, RFLP typing has identified enough unexpected links and sites of transmission that we now focus on the time involved in the case-contact interaction even if that contact is not within the home or workplace. We have always emphasized the importance of considering the nonhousehold contact as a means of disease transmission, and we address the potential of the more casual contact on the basis some of the unexpected links we found (such as the student and security guard mentioned previously). Bishai et al. report that TB transmission continues despite active case finding and 15 years of directly observed therapy (20) and suggest that this transmission may have been in difficult-to-treat populations (such as the homeless) or the result of casual transmission. Other authors have also commented on the likely importance of casual transmission and the important role of RFLP typing in documentation (21–23). If the trend in Maryland noted by Bishai et al. (20) and Sterling et al. (24) is true in otherwise well-served, financially viable TB control programs, genotyping of *M. tuberculosis* isolates will be essential for public health interventions in these settings.

Nearly three fourths of our cases did not cluster. Although the patients with unique patterns in our dataset may have clustered with someone outside of the study (thus underestimating clustering in our population), the high numbers of unique cases still suggest that most of our cases are due to reactivation of latent TB infection. Thus, we should concentrate resources on early identification of latent TB infection and follow-up to interrupt TB transmission. In addition, the finding that our conventional contact investigations were overestimating transmission in persons born in disease-endemic countries supports the 1998 findings of Behr et al. (25). Isolates from contacts born in foreign countries were significantly more likely to have different strains than isolates from contacts born in the United States (data not shown). This finding suggests that the public health program makes certain assumptions about the definition of close contacts that are not correct in every case.

The utility of genotyping in outbreak response cannot be overestimated. A strain designated as the genotyping network pattern 437 has been responsible for 20 cases since 1993. The strain's reemergence in the homeless in 2000 was unsuspected, partly because men who were believed to spend few nights in homeless shelters were found to have this strain. In addition, the shelter, identified as a common site for many persons with TB, had excellent control measures. In this situation, genotyping provided evidence that warranted a thorough investigation and the development of new educational materials. Educational activities were conducted in shelters, emergency rooms, and other sites; and contacts from a large-scale screening at the site are being followed. The discovery of other unexpected clusters has resulted in involving less traditional sites in education and prevention efforts. One such educational video about TB was pilot tested in hairdressing salons in a Massachusetts community. In addition, genotyping can be applied to situa-

tions of increased cases to determine that an outbreak is not occurring, as in Hanau-Bercot et al. (8). Determining that cases have different genotypes can mean that costly mobilization of public health resources to combat an outbreak is unnecessary.

Policies to reduce cross-contamination risks have been instituted at the laboratory where cross-contamination occurred in 1994 and 1995. Live control strains are no longer used to process specimens. When a new patient's culture grows five or fewer colonies, laboratory records are reviewed to establish the likelihood of error. However, because cross-contamination or other misclassification of a specimen or result does not rule out the possibility of a TB diagnosis, care must be taken before attributing a positive culture to error. In addition, public health programs should be able to prove or disprove laboratory contamination on a real-time basis for clinical purposes. This confirmation is important for laboratory internal quality control, can result in cost savings of several thousand dollars per misdiagnosed patient (19,26), and can save emotional and other costs to the patient.

Limitations of Genotyping Techniques

In the cluster investigation study, 105 (66%) of 159 persons in clusters were interviewed or had epidemiologic links to everyone else in their cluster, so they were not interviewed according to the protocol. Of the remaining 54 (33%), patients refused to be interviewed (10 patients). An additional 20 patients were defined as "administratively closed" at the end of the study. These patients did not refuse to participate in the study but could not be reached after six phone attempts or three visit attempts. Several either agreed to participate or asked to be contacted later but were unreachable. Of 45 clusters investigated as part of the official cluster investigation study with patients eligible for interview, we interviewed all persons in 13 clusters.

The RFLP technique is expensive and technically demanding; it also requires microbiologic expertise and resources to obtain cultures as well as sophisticated software for RFLP pattern comparison. Thus, turnaround time between specimen collection and availability of RFLP result may be lengthy (27). In our study, this turnaround time averaged 7 months, which may have contributed to our difficulties in obtaining patient interviews. However, identifying clusters would be a slow process even with negligible turnaround times, as latent TB infection reactivates slowly. Reopening contact investigations after receipt of RFLP results was impractical in most instances. In addition, interpreting results is more difficult because of the lack of specificity in patterns with fewer than seven IS*6110* copies.

A recently reported limitation of RFLP relates to sampling (28). A certain number of cases in every population cannot be assigned RFLP types. Either the specimens do not grow *M. tuberculosis*, or TB is diagnosed on the basis of clinical information without bacteriological confirmation. During the study period in Massachusetts, RFLP typing provided information on 95% of our culture-confirmed cases, 77 % of our cases overall.

RFLP and any other genotyping tool can be useful for TB control if certain assumptions hold true. *M. tuberculosis* strains are relatively stable but not immutable; thus, if persons fall into a cluster, they probably have the same TB strain. Clustering implies recent transmission because strains change over time. A few studies (29–32) have analyzed the stability of RFLP patterns; one determined that half of the strains will demonstrate a shift in 3–4 years (29), while another study found no change (32). Neimann et al. saw almost no change in patterns among 75 isolates in chains of transmission, i.e., multiple patients (30). These studies cannot make assumptions about changes of patterns during latency, and rates of change of strains may vary. In addition, we assume that clustering in the TB control program's population implies recent transmission and not simply endemic strains with a lack of diversity. We assume that sampling is complete and that no difference exists between the patients with or without available isolates for genotyping. For public health programs to benefit fully from a genotyping technique, it must be rapid, inexpensive, and reproducible from laboratory to laboratory. In general, the currently available DNA fingerprinting tools should be used only in conjunction with epidemiology. Universal fingerprinting, in particular, shows its greatest utility when the health department can respond to this new information by allocating additional resources to conduct investigations of unexpected clusters (33). Genotyping can then have as great an impact on public health practice nationwide as it had in Massachusetts.

Acknowledgments

The authors gratefully acknowledge the contributions of the following: Paul Elvin, Alissa Scharf, Denise O'Connor, Michael Malone, Barbara McInnis, John Bernardo, nurses at the Boston Public Health Commission TB Division and the Massachusetts Department of Public Health Tuberculosis Division, Muriel Day, JoAnn Dopp, The Shelters; Jill Northrup, Al DeMaria, and Christopher R. Braden.

The research in this study was supported in part by the Centers for Disease Control and Prevention, National Tuberculosis Genotyping and Surveillance Network Cooperative Agreement U52/ CCU100156.

Ms. Miller has worked as an epidemiologist for the Massachusetts Department of Public Health, Division of Tuberculosis Prevention and Control, since 1996. She coordinated the activities of the Massachusetts Sentinel Surveillance Site of the National Tuberculosis Genotyping and Surveillance Network. Her primary research interests are genotyping of *Mycobacterium tuberculosis* isolates, antimicrobial resistance, and tuberculosis occurrence in homeless shelters.

References

1. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of TB in New York City; an analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–6.

2. Edlin BR, Tokars JI, Greico MH, Crawford JT, Williams J, Sordillo EM, et al. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
3. Beck-Sague C, Dooley SW, Hutton MD, Otten J, Breeden A, Crawford J, et al. Hospital outbreak of multidrug-resistant *Mycobacterium tuberculosis* infections: factors in transmission to staff and HIV-infected patients. JAMA 1992;268:1280–6.
4. Bauer J, Kok-Jensen A, Faurschou P, Theusen J, Taudorf E, Anderson AB. A prospective evaluation of the clinical value of nationwide DNA fingerprinting of tuberculosis isolates in Denmark. Int J Tuberc Lung Dis 2000;4:295–9.
5. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
6. Dobbs KG, Lok KH, Bruce F, Mulcahy D, Benjamin WH, Dunlap NE. Value of *Mycobacterium tuberculosis* fingerprinting as a tool in a rural state surveillance program. Chest 2001;120:1877–82.
7. Chavez F, Dronda F, Cave MD, Alonso-Sanz M, González-López A, Eisenach KD, et al. A longitudinal study of transmission of tuberculosis in a large prison population. Am J Respir Crit Care Med 1997;155:719–25.
8. Hanau-Bercot B, Grémy I, Raskine L, Bizet J, Gutierrez MC, Boyer-Mariotte S, et al. A one-year prospective study (1994–1995) for a first evaluation of tuberculosis transmission in French prisons. Int J Tuberc Lung Dis 2000;4:853–9.
9. Southwick KL, Hoffmann K, Ferree K, Matthews J, Salfinger M. Cluster of tuberculosis cases in North Carolina: possible association with atomizer reuse. Am J Infect Control 2001;29:1–6.
10. Michele TM, Cronin WA, Graham NMH, Dwyer DM, Spies Pope D, Harrington N, et al. Transmission of *Mycobacterium tuberculosis* by a fiberoptic bronchoscope: identification by DNA fingerprinting. JAMA 1997;278:1093–5.
11. Cronin W, Rodriguez E, Valway S, Bur S, Hooper N, Smithwick R, et al. Pseudo-outbreak of tuberculosis in an acute-care general hospital: epidemiology and clinical implications. Infect Control Hosp Epidemiol 1998;19:739–40.
12. Wurtz R. Specimen contamination in mycobacteriology laboratory detected by pseudo-outbreak of multidrug-resistant tuberculosis: analysis by routine epidemiology and confirmation by molecular technique. J Clin Microbiol 1996;34:1017–9.
13. Braden CR, Templeton GL, Stead WW, Bates JH, Cave MD, Valway S. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* culture with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
14. Sharnprapai S, Miller A. Genotyping analyses among foreign-born and U.S.-born persons in Massachusetts. Emerg Infect Dis 2002; 8;1239-1245.
15. Cowan LS, Crawford JT. Genotype analysis of *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. Emerg Infect Dis 2002;8;1294-1302.
16. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192-6.
17. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
18. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen DS, Bunschoten A, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
19. Northrup JM, Miller AC, Nardell E, Sharnprapai S, Etkind S, Driscoll J, et al. Estimated costs of false laboratory diagnoses of tuberculosis in three patients. Emerg Infect Dis 2002;8;1264-70.
20. Bishai WR, Graham NM, Harrington S, Pope DS, Hooper N, Astemborski J, et al. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA 1998;280:1679–84.
21. Solsona J, Caylà JA, Verdú E, Estrada MP, Garcia S, Roca D, et al. Molecular and conventional epidemiology of tuberculosis in an inner city district. Int J Tuberc Lung Dis 2000;5:724–31.
22. Onorato IM. Tuberculosis outbreaks in the United States. Int J Tuberc Lung Dis 2000;4:121–6.
23. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, et al. An outbreak involving extensive transmission of a virulent strain of tuberculosis. N Engl J Med 1998;338:633–9.
24. Sterling TR, Thompson D, Stanley RL, McElroy PD, Madison A, Moore K, et al. A multistate outbreak of TB among members of a highly mobile social network: implications for TB elimination. Int J Tuberc Lung Dis 2000;4:1066–73.
25. Behr MA, Hopewell PH, Paz EA, Kawamura LM, Schecter GF, Small PM. Predictive value of contact investigation for identifying recent transmission of *Mycobacterium tuberculosis.* Am J Respir Crit Care Med 1998;158:465–9.
26. Dunlap NE, Harris RH, Benjamin WH, Harden JW, Hafner D. Laboratory contamination of *Mycobacterium tuberculosis* cultures. Am J Respir Crit Care Med 1995;152:1722–4.
27. van Soolingen D. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. J Intern Med 2001;249:1–26.
28. Murray M. Sampling bias in the molecular epidemiology of tuberculosis. Emerg Infect Dis 2002;8:363–9.
29. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJD, van Embden JDA, van Soolingen D. Analysis of rate of change of IS*6110* RFLP patterns *of Mycobacterium tuberculosis* based on serial patient isolates. J Infect Dis 1999;180:1238–44.
30. Niemann S, Rusch-Gerdes S, Richter E, Thielen H, Heykes-Uden H, Diel R. Stability of IS*6110* restriction fragment length polymorphism patterns of *Mycobacterium tuberculosis* strains in actual chains of transmission. J Clin Microbiol 2000;38:2563–7.
31. Yeh RW, Ponce de Leon A, Agasino CB, Hahn JA, Daley CL, Hopewell PC, et al. Stability of *Mycobacterium tuberculosis* DNA genotypes. J Infect Dis 1998;177:1107–11.
32. Cave MD, Eisenach KD, Templeton G, Salfinger M, Muzerek G, Bates JH, et al. Stability of the DNA fingerprint patterns produced with IS*6110* in strains of *Mycobacterium tuberculosis*. J Clin Microbiol 1994;32:262–6.
33. Dunlap NE. The use of RFLP as a tool for tuberculosis control: utility or futility? Int J Tuberc Lung Dis 2000;4:134–8.

Address for correspondence: Sharon Sharnprapai, Massachusetts Department of Public Health, Division of Tuberculosis Prevention and Control, 305 South St., Jamaica Plain, MA 02130, USA; fax: 617-983-6990; e-mail: Sharon.Sharnprapai@state.ma.us

Human Exposure following *Mycobacterium tuberculosis* Infection of Multiple Animal Species in a Metropolitan Zoo

Peter Oh,* Reuben Granich,*† Jim Scott,* Ben Sun,‡ Michael Joseph,* Cynthia Stringfield,§ Susan Thisdell,§ Jothan Staley,¶ Donna Workman-Malcolm,¶ Lee Borenstein,# Eleanor Lehnkering,# Patrick Ryan,# Jeanne Soukup,# Annette Nitta,# and Jennifer Flood*

From 1997 to 2000, *Mycobacterium tuberculosis* was diagnosed in two Asian elephants (*Elephas maximus*), three Rocky Mountain goats (*Oreamnos americanus*), and one black rhinoceros (*Diceros bicornis*) in the Los Angeles Zoo. DNA fingerprint patterns suggested recent transmission. An investigation found no active cases of tuberculosis in humans; however, tuberculin skin-test conversions in humans were associated with training elephants and attending an elephant necropsy.

Outbreaks of *Mycobacterium tuberculosis* have been documented in environments such as hospitals, schools, factories, homeless shelters, and prisons (1–5). In more unconventional settings, such as circuses and exotic animal facilities, outbreaks pose unique tuberculosis (TB) control challenges because transmission may involve animals as well as humans (6,7). Zoos are a particular public health concern because of the close contact between TB-susceptible animals and humans, specifically animal handlers and visitors to the facility or exhibit. Infection and disease related to *M. tuberculosis* have been reported for a variety of species ranging from birds to primates (8–10). Although evidence for human-to-animal transmission of *M. tuberculosis* has been described (11), little documentation of zoonotic transmission to humans exists (7). We describe the first reported multispecies epizootic of genotypically identical strains of *M. tuberculosis* in a zoo and the results of an investigation of exposed zoo employees.

Synopsis of Animal TB Cases

From 1997 to 2000, *M. tuberculosis* was identified in six animals at the Los Angeles Zoo. In March 1997, an Asian elephant (*Elephas maximus*) (elephant 1) died of salmonellosis. During the necropsy, pulmonary lesions were discovered, and a lymph node specimen showed *M. tuberculosis*. In April 1997, a positive trunk wash culture of *M. tuberculosis* was obtained from a second Asian elephant (elephant 2), which had resided in the same barn as elephant 1. In July 1998, a Rocky Mountain goat (*Oreamnos americanus*) (goat 1) suffered deterioration associated with worsening pneumonia; the pathologic examination was consistent with TB, and culture confirmed *M. tuberculosis.* Tuberculin skin tests of two cohabiting goats (goats 2 and 3) were positive, but their cultures were negative. In September 1998, a black rhinoceros (*Diceros bicornis*) had a positive *M. tuberculosis* culture. In February 2000, routine chest radiographs of goats 2 and 3 showed abnormalities consistent with TB. We isolated *M. tuberculosis* from both animals.

Veterinary Epidemiologic Investigation

We examined medical and location histories of the affected animals as well as handling practices, health-care procedures, and other means of potential exposure to *M. tuberculosis.* An epidemiologic link was defined as documented exposure to an infectious human or animal with TB. We conducted an infection control assessment of the animal compounds and health-care facilities and measured air flow in the compounds by smoke testing (12).

Examination of Animal Isolates

Elephant isolates (e.g., trunk washes) were obtained according to United States Department of Agriculture guidelines (13). We used saline nasal washes to gather rhinoceros isolates and tracheal washes to gather isolates from goats, as well as specimens for pathologic examination. The Los Angeles County Public Health Laboratory performed restriction fragment length polymorphism (RFLP) analyses on the isolates. Southern blots of *Pvu*II–restricted chromosomal DNA were run in 0.08% agarose gels, probed with a DNA fragment corresponding to IS*6110*, and detected by chemiluminescence (14).

*California Department of Health Services, Berkeley, California, USA; †Centers for Disease Control and Prevention, Atlanta, Georgia, USA; ‡California Department of Health Services, Sacramento, California, USA; §Los Angeles Zoo, Los Angeles, California, USA; ¶City of Los Angeles Occupational Health Services Division, Los Angeles, California, USA; and #Los Angeles County Department of Health Services, Los Angeles, California, USA

Employee TB Screening

Medical records of zoo employees were reviewed for evidence of TB symptoms (i.e., persistent cough, hemoptysis, night sweats, difficulty in breathing, and weight loss), chest radiograph information, and tuberculin skin-test results. In addition, a list of current and former employees was confidentially matched against reported TB cases in the California state registry from 1985 to 2000 (15). During an annual occupational health screening in June 2000, employees participated in TB symptom reviews and received tuberculin skin tests; they also completed a questionnaire on medical history, job type, and history of contact with the infected animals.

Tuberculin Skin-Test Conversion Categories and Statistical Analyses

A positive tuberculin skin test was defined with a documented induration of ≥5 mm. We categorized employees with positive tuberculin skin tests as true, probable, or possible converters or as nonconverters. True converters were patients with a negative two-step test (within a 3-week period), followed by an increase in induration of ≥10 mm within 2 years. Probable converters had no two-step test but had either an induration increase of ≥10 mm within a 2-year period or two negative (<5 mm) results within 1 year followed by a positive result of ≥10 mm. Possible converters had an initial negative result followed by a positive test. Nonconverters were patients with positive tuberculin skin tests who did not fit these conversion categories (e.g., one positive tuberculin skin test without a previous test). The questionnaire responses of converters (true, probable, and possible combined) were compared to those of employees with negative tuberculin skin tests.

Relative risk (RR) ratios were calculated by chi-square or Fisher's exact test by using Epi Info 6 (Centers for Disease Control and Prevention, Atlanta, GA). Statistical significance was considered to be $p<0.05$.

Epidemiologic Findings and Genotyping of Animal Isolates

Both elephants with TB had resided at the same exotic animal facility in the United States before arriving at the Los Angeles Zoo in 1994. In 1997, *M. tuberculosis* was found in four other elephants at the exotic animal facility; however, the RFLP pattern differed from that of elephants 1 and 2 (unpub. data). The only documented epidemiologic links among the affected animals were between the two elephants and among the three goats. No common contact outside the animal compounds and no contact with an infectious human was found to account for TB transmission among multiple species.

Standard operating procedures at the zoo included guidelines for animal quarantine and the use of N95 respirators during medical procedures. The elephant compound was 27 m from the rhinoceros compound, and the goat compound was 90 m from both. Smoke tests of the animal compounds showed adequate air movement of 0.3–0.9 m/s and winds of 4.8–8.0 km/hr in ambient conditions.

RFLP analysis showed that five of six animal isolates shared an identical IS*6110* pattern (Figure). The isolate of goat 3 differed by one additional band.

Employee Screening Findings

No active TB cases in humans were found during employee TB screening, medical records review, or query of the state case registry. Of 1,088 employees, no matches were identified in the database of cases reported from 1985 to 2000.

Of 336 employees screened for TB during this investigation, 332 (99%) completed the questionnaire, and 307 (91%) had a tuberculin skin test. Of the 323 employees who responded to the job category question, most were veterinarians or animal keepers (112 [35%]) and administrative staff (73 [23%]); other job categories included maintenance workers (38 [12%]), custodians (29 [9%]), and groundskeepers (24 [7%]). Sixty-two percent of the animal handlers were Caucasian, and 61% of the groundskeepers were Hispanic. Sixty percent of all employees reported contact with animals.

Of the 307 employees who had tuberculin skin tests, 55 (18%) reported a positive result. Of these, none reported TB symptoms, and chest radiographs showed no abnormalities suggesting active TB. Persons with positive tuberculin skin tests were more likely than persons with negative tests to be men (RR 3.7, 95% confidence interval [CI] 2.0 to 6.8), groundskeepers (RR 2.6, 95% CI 1.5 to 4.7), construction workers (RR 2.5, 95 CI 1.3 to 4.8), or attendees at the elephant necropsy (RR 2.9, 95% CI 1.5 to 5.5). However, animal caretaking and animal contact were not associated with a positive tuberculin skin test. In this group of employees, we found no true converters, 10 (18%) probable converters, 5 (9) possible converters, and 40 (73%) nonconverters.

Risk factors for tuberculin skin-test conversion are described in the Table. Employees reporting attendance at elephant 1's necropsy were more likely to have documented tuberculin skin-test conversions than those not present (RR 6.3, 95% CI 2.1 to 18.9). Furthermore, employees who trained elephants were more likely to have tuberculin skin-test conversions than those who did not train elephants (RR 4.1, 95% CI

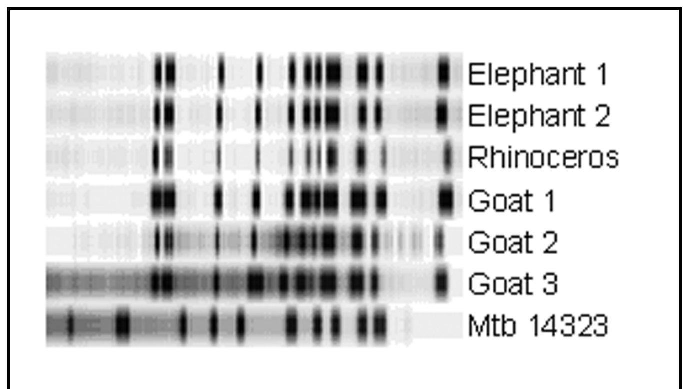

Figure. IS*6110* restriction fragment length polymorphism results of the six animal isolates and the *Mycobacterium tuberculosis* reference strain Mtb14323. Molecular weights of IS*6110*-containing *Pvu*II fragments of the reference strain are approximately 17, 7.4, 7.1, 4.5, 3.6, 3.1, 2.1, 1.9, 1.7, 1.5, and 1.4 kb.

Table. Relative risks for tuberculin skin-test conversion based on answers reported on employee exposure questionnaires[a]

Characteristics reported in questionnaire	Converters[b] (%)	TST-negatives[c] (%)	Risk ratio (95% CI)	p value
Male	13 (87)	112 (45)	7.3 (1.7 to 31.9)	0.002
Ethnicity				
Hispanic	10 (67)	109 (44)	2.4 (0.8 to 6.8)	0.092
White	2 (13)	105 (43)[d]	0.2 (0.1 to 1.0)	0.025
Black	2 (13)	22 (9)	1.5 (0.4 to 6.3)	0.41
Asian	1 (7)	9 (4)	1.8 (0.3 to 12.3)	0.45
U.S.-born	11 (73)	226 (90)	0.4 (0.1 to 1.0)	0.052
BCG vaccine history	0 (0)	15 (7)[e]	0.0 (0.0 to 6.0)[f]	0.40
Contact with any animal	7 (50)[g]	147 (60)[h]	0.7 (0.2 to 1.9)	0.45
Job type				
Animal care	4 (29)[g]	91 (37)[i]	0.7 (0.2 to 2.2)	0.54
Groundskeeping	5 (36)[g]	14 (6)	7.1 (2.6 to 19.1)	<0.001
Custodial	2 (14)[g]	19 (8)	1.9 (0.5 to 8.0)	0.31
Maintenance	1 (7)	15 (6)	1.2 (0.2 to 8.5)	0.60
Construction	1 (7)	11 (4)	1.6 (0.2 to11.3)	0.49
Administrative	1 (7)	67 (27)	0.2 (0.0 to 1.7)	0.082
Animal health center exposure	5 (33)	152 (63)[j]	0.3 (0.1 to 0.9)	0.025
Elephant compound exposure	5 (33)	130 (52)	0.5 (0.2 to 1.4)	0.17
Trained elephants	3 (20)	12 (5)	4.1 (1.3 to 13.1)	0.045
Visited elephants	2 (13)	77 (31)	0.4 (0.1 to 1.6)	0.12
Attended necropsy	3 (20)	7 (3)	6.3 (2.1 to 18.9)	0.014
Goat compound exposure	3 (20)	72 (29)	0.6 (0.2 to 2.2)	0.35
Rhino compound exposure	2 (13)	82 (33)	0.3 (0.1 to 1.5)	0.097
Animal nursery exposure	2 (13)	124 (50)[i]	0.2 (0.0 to 0.7)	0.007

[a]TST, tuberculin skin test; CI, confidence interval.
[b]Employee with a positive TST in any of the converter categories as described (true, probable, or possible). N=15, except where noted.
[c]N=251, except where noted.
[d]N=245.
[e]N=232.
[f]Odds ratio reported because of a zero value.
[g]N=14.
[h]N=244.
[i]N=248.
[j]N=243.

1.3 to 13.1). Groundskeepers (n=24) had an increased risk of tuberculin skin-test conversion compared with other job categories (RR 7.1, 95% CI 2.6 to 19.1). Four of five groundskeepers with tuberculin skin-test conversions were born in the United States; 11 of 14 employees with negative skin tests and none of the 5 groundskeepers with positive tuberculin skin tests (in the nonconverter group) were born in the United States. A lower likelihood of tuberculin skin-test conversion was associated with visiting the animal nursery (RR 0.2, 95% CI 0.0 to 0.7) and the health center (RR 0.3, 95% CI 0.1 to 0.9).

Conclusions

Although 55 zoo employees showed evidence of *M. tuberculosis* infection, no person with active TB disease was identified. Given the public's distance from the animals and the absence of active TB among employees with closer contact with these animals, *M. tuberculosis* was likely not transmitted from humans to animals at this zoo.

The finding that groundskeepers and not animal handlers were associated with a higher risk of tuberculin skin-test conversion was unexpected. Because groundskeepers as a group were more likely to be born outside of the United States than animal keepers, we hypothesized that tuberculin skin-test conversion may have resulted from infections acquired outside of the zoo. However, within this group, only one of five groundskeepers with a tuberculin skin-test conversion was born outside of the country. This finding suggests that a recent exposure may have been responsible for tuberculin skin-test conversion in this occupational category, although small numbers limit the inference.

Genotyping evidence strongly suggested transmission from one species to another, although corroborating epidemiologic evidence of transmission was not discovered. One expla-

nation for transmission is that the elephants may have been exposed to TB at the animal facility in which they resided before their arrival at the zoo. The distances to other animal compounds at the zoo make airborne spread unlikely. In addition, we found no employees with active TB. Since interspecies transmission routes were not found, we suggest that continued vigilance for sources of ongoing transmission is warranted.

Finally, we did discover a significant association between tuberculin skin-test conversion and attending the elephant necropsy and training elephants in the compound. This report emphasizes the importance of adhering to strict infection control measures during large animal necropsies and medical procedures, even when TB is not suspected, because of potentially large bacillary loads.

Acknowledgments

We thank Charles Sedgwick and the zoo administrative office, staff veterinarians, animal health and animal care staff, occupational health nurses, and zoo employees for their participation in the questionnaire. We also thank the TB Control Program staff and David Sasai for conducting the ventilation assessment.

This investigation was supported by the Centers for Disease Control and Prevention cooperative agreement funds for the National Tuberculosis Genotyping and Surveillance Network.

Mr. Oh is an epidemiologist in the TB Control Branch, Division of Communicable Disease Control at the California Department of Health Services. His main interests are TB surveillance, outbreak investigations, and international health.

References

1. Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams J, Sordillo EM, et al. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
2. Hoge CW, Fisher L, Donnell HD, Dodson DR, Tomlinson GV, Breiman RF, et al. Risk factors for transmission of *Mycobacterium tuberculosis* in a primary school outbreak: lack of racial difference in susceptibility to infection. Am J Epidemiol 1994;139:520–30.
3. Mosher CB, Derebery VJ, Young BJ, Adams RA. Unusually aggressive transmission of tuberculosis in a factory. Journal of Occupational Medicine 1987;29:29–31.
4. Curtis AB, Ridzon R, Novick LF, Driscoll J, Blair D, Oxtoby M, et al. Analysis of *Mycobacterium tuberculosis* transmission patterns in a homeless shelter outbreak. Int J Tuberc Lung Dis 2000;4:308–13.
5. Valway SE, Richards SB, Kovacovich J, Greifinger RB, Crawford JT, Dooley SW. Outbreak of multi-drug-resistant tuberculosis in a New York State prison, 1991. Am J Epidemiol 1994;140:113–22.
6. Darney PD, Greene JE. Tuberculosis outbreak in a circus: report of a cooperative investigation. Am J Public Health 1973;63:43–5.
7. Michalak K, Austin C, Diesel S, Bacon MJ, Zimmerman P, Maslow JN. *Mycobacterium tuberculosis* infection as a zoonotic disease: transmission between humans and elephants. Emerg Infect Dis 1998;4:283–7.
8. Ackerman LJ, Benbrook SC, Walton BC. *Mycobacterium tuberculosis* infection in a parrot (*Amazona farinosa*). American Review of Respiratory Diseases 1974;109:388–90.
9. Shin NS, Kwon SW, Han DH, Bai GH, Yoon J, Cheon DS, et al. *Mycobacterium tuberculosis* infection in an orangutan (*Pongo pygmaeus*). J Vet Med Sci 1995;57:951–3.
10. Montali RJ, Mikota SK, Cheng LI. *Mycobacterium tuberculosis* in zoo and wildlife species. Rev Sci Tech 2001;20:291–303.
11. Michel AL, Huchzermeyer HF. The zoonotic importance of *Mycobacterium tuberculosis*: transmission from human to monkey. J S Afr Vet Assoc 1998;69:64–5.
12. Schwartzman K, Loo V, Pasztor J, Menzies D. Tuberculosis infection among health care workers in Montreal. Am J Respir Crit Care Med 1996;154:1006–12.
13. Department of Agriculture (US). Animal and Plant Health Inspection Service, Guidelines for the control of tuberculosis in elephants. Washington: The Department; 1997.
14. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
15. Yokoe DS, Subramanyan GS, Nardell E, Sharnprapai S, McCray E, Platt R. Supplementing tuberculosis surveillance with automated data from health maintenance organizations. Emerg Infec Dis 1999;5:779–87.

Address for correspondence: Jennifer Flood, California Department of Health Services, Tuberculosis Control Branch, 2151 Berkeley Way, Room 608, Berkeley, CA 94704, USA; fax: 510-849-5269; e-mail: jflood@dhs.ca.gov

Antique poster from the collection of Joel and Jahan Montague. Joel Montague is Board Chairman of Partners for Development, a nonprofit organization. Jahan Montague is associate professor of nephrology, University of Massachusetts Medical School in Worcester.

Genotype Analysis of *Mycobacterium tuberculosis* Isolates from a Sentinel Surveillance Population

Lauren S. Cowan* and Jack T. Crawford*

As part of the National Tuberculosis and Genotyping Surveillance Network, isolates obtained from all new cases of tuberculosis occurring in seven geographically separate surveillance sites from 1996 through 2000 were genotyped. A total of 10,883 isolates were fingerprinted by the IS*6110*-restriction fragment length polymorphism method, yielding 6,128 distinct patterns. Low-copy isolates (those with six or fewer bands) were also spoligotyped. The distribution of specific genotype clusters was examined. Databases were also examined for families of related genotypes. Analysis of IS*6110* patterns showed 497 patterns related to the W-Beijing family; these pattens represent 946 (9%) of all isolates in the study. Six new sets of related fingerprint patterns were also proposed for isolates containing 6–15 copies of IS*6110*. These fingerprint sets contain up to 251 patterns and 414 isolates; together, they contain 21% of isolates in this copy number range. These sets of fingerprints may represent endemic strains distributed across the United States.

The National Tuberculosis Genotyping and Surveillance Network was created by the Centers for Disease Control and Prevention (CDC) to determine the relative frequency of *Mycobacterium tuberculosis* strains in specific geographic areas, the extent of spread of related strains in communities, and the impact of IS*6110* fingerprinting on tuberculosis (TB) control. From 1996 through 2000, the TB genotyping network laboratories fingerprinted 10,883 isolates (one isolate per newly diagnosed case of TB) from seven sentinel surveillance sites in the United States: the states of Arkansas, Maryland, Massachusetts, Michigan, and New Jersey, along with four counties in Texas and six counties in California. Key components of the project included the establishment of standard methods and use of specialized software, the BioImage Whole Band Analyzer version 3.4 (BioImage, Ann Arbor, MI), for pattern analysis. The following were created as part of this study: databases containing the images of IS*6110* patterns for all isolates at each sentinel surveillance site; a network database that includes all distinct spoligotype patterns; and an epidemiologic database (EpiInfo) for information about sentinel surveillance site, case report date, IS*6110* pattern designation, and secondary typing results for each patient. The final network database of fingerprints contains 6,128 patterns.

We report here an overview of the contents of the TB genotyping network fingerprint database, including the distribution of isolates at sentinel surveillance sites, genotype patterns that occurred with high frequency, the extent of previously described genotype families, and new families of related fingerprints. This analysis should not be considered exhaustive but rather a summary of our observations and an introduction to the types of data that can be derived.

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Methods

Methods for IS*6110* fingerprinting, spoligotyping, and compiling the network databases are described elsewhere (1). The distribution of isolates by sentinel surveillance site, fingerprint pattern, and spoligotype (for isolates with six or fewer copies of IS*6110*) was derived from the Epi Info database. The spoligotype patterns are reported in octal code by the convention previously described (2). IS*6110* fingerprint patterns start with the prefix FP and spoligotype patterns with SP.

We used the BioImage Whole Band Analyzer software package version 3.4 to analyze fingerprint patterns in the genotyping network fingerprint database. The bands in two patterns were compared at two levels. First, bands in two patterns were identified as matched bands if the size of the bands differed by <2.5%. Next, the interband spacing between matched bands in the two patterns was compared; a limit of 95% for variation in interband spacing was used. The Jaccard coefficient of similarity between two patterns, A and B, was used to calculate the percentage match between two patterns: 100 x number of matched bands /(number bands in A + number bands in B - number of matched bands).

Results and Discussion

Isolates from 10,883 patients from seven sentinel surveillance sites were fingerprinted by using the IS*6110* restriction fragment length polymorphism (RFLP) method: Arkansas, 709; California, 2,514; Massachusetts, 986; Maryland, 1,180; Michigan, 1,471; New Jersey, 2,113; and Texas, 1,910. From

these isolates, 6,128 distinct fingerprint patterns were identified and included in the final genotyping network database. Analysis of the IS*6110* copy number of the isolates confirmed the previously described bimodal distribution (Figure 1) (3). This distribution has been used to separate isolates of *M. tuberculosis* into two groups: isolates with six or fewer copies of IS*6110* are classified as low-copy isolates and those with more than six copies as high-copy isolates. The greatest numbers of patterns occurred in the 9–14 copy-number range (Figure 1). Clustering of isolates on the basis of matching fingerprint patterns is summarized in Table 1**.** Clustering was very high among the low-copy isolates, which supports the requirement for secondary typing of these isolates. Clustering decreased with increasing copy number; copy numbers 21 and 22, which included large outbreaks, were the exceptions.

High-Copy Isolates

The 8,245 isolates with more than six copies of IS*6110* yielded 5,640 fingerprint patterns. Of these patterns, 4,846 (85.9%) were identified for a single isolate, and 3,399 isolates were grouped into 794 fingerprint-defined clusters. Of these clusters, 557 contained isolates from a single site. The clusters contained up to 105 isolates, but 683 (86.0%) of the clusters contained only two to five isolates. In fact, only 18 clusters contained 20 or more isolates. The distribution of isolates in these 18 clusters is shown in Table 2, and the fingerprint patterns are shown in Figure 2. For 11 of the 18 largest clusters, ≥ 90% of the isolates in the cluster were from a single site.

One of the largest clusters (FP 00237, 100 isolates) corresponds to *M. tuberculosis* strain 210, a member of the W-Beijing family that was shown in previous studies to be disseminated across the United States (3). In the network, FP 00237 was associated with large clusters in Arkansas and Texas and was also reported by Maryland and New Jersey. Two additional patterns associated with large clusters, FP 00027 (102 isolates in Michigan) and FP 01284 (46 isolates in Texas), were similar to FP 00237. The Beijing family of strains has received considerable attention because of its association with several large outbreaks, frequent association with

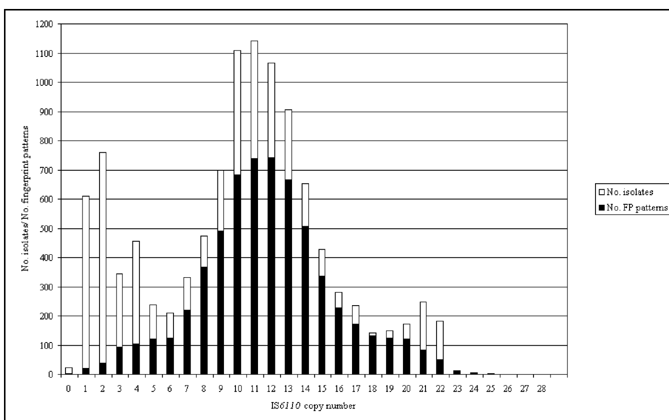

Figure 1. Distribution of all isolates and fingerprint patterns by number of copies of IS*6110*. The light bars show the distribution of isolates; the dark bars show the distribution of fingerprint patterns.

Table 1. Distribution of isolates and fingerprint patterns by number of copies of IS*6110*

IS*6110* copy no.	No. patterns	No. isolates	No. clusters[a]	No. clustered isolates (%)	Average cluster size (range)
0	1	22	1	22 (100)	22 (22)
1	17	610	9	602 (99)	67 (2–291)
2	36	759	16	739 (97)	46 (2–456)
3	92	345	40	293 (85)	7.3 (2–49)
4	102	456	28	382 (84)	14 (2–212)
5	119	237	37	155 (65)	4.2 (2–13)
6	121	209	21	109 (52)	5.2 (2–20)
7	217	332	40	155 (47)	3.9 (2–29)
8	364	474	56	166 (35)	3.0 (2–12)
9	489	699	78	288 (41)	3.7 (2–20)
10	681	1,108	112	539 (49)	4.8 (2–105)
11	737	1,143	116	522 (46)	4.5 (2–70)
12	738	1,067	114	443 (42)	3.9 (2–27)
13	663	906	83	326 (36)	3.9 (2–23)
14	505	653	51	199 (30)	3.9 (2–27)
15	333	428	35	130 (30)	3.7 (2–15)
16	225	282	17	74 (26)	4.4 (2–21)
17	169	236	16	83 (35)	5.2 (2–46)
18	128	143	9	24 (17)	2.7 (2–5)
19	121	149	17	45 (30)	2.6 (2–7)
20	118	172	23	77 (45)	3.3 (2–14)
21	81	248	15	182 (73)	12 (2–100)
22	48	182	12	146 (80)	12 (2–102)
23	13	13	0		
24	5	5	0		
25	2	2	0		
26	1	1	0		
27	1	1	0		
28	1	1	0		
All	6,128	10,883	946	5,701 (52)	6.0 (2–456)

[a]Number of fingerprint patterns reported for more than one isolate.

multidrug resistance, and emergence in selected populations, particularly in the former Soviet Union (4,5). All Beijing isolates share a characteristic spoligotype (000000000003771); however, in this study, spoligotyping was not performed for high-copy isolates. Other molecular criteria that define W-Beijing strains include insertions of IS*6110* in the dnaA-dnaN region (A1 insertion) and in the NTF region and an empirical fingerprint pattern that contains 15 to 24 bands and is similar to that of strain W (4). To estimate the occurrence of Beijing isolates in our study, all patterns with 16 to 24 bands were visually compared to FP 00237. The W fingerprint was easily

Table 2. Distribution of isolates with high-copy fingerprint patterns reported with high frequency[a, b]

FP	No. bands	No. isolates	No. isolates/site						
			AR	CA	MA	MD	MI	NJ	TX
00015	7	29	0	0	0	2	0	27	0
00019	12	27	4	7	0	2	3	2	9
00027	22	102	0	0	0	0	102	0	0
00028	11	70	0	0	0	0	70	0	0
00035	13	33	0	0	0	0	0	0	33
00159	11	24	0	0	0	0	0	0	24
00237	21	100	12	0	0	2	0	1	85
00242	10	105	6	1	1	0	1	0	96
00316	14	27	3	22	2	0	0	0	0
00325	11	20	15	1	0	0	0	4	0
00372	12	20	11	0	0	0	0	0	9
00469	16	21	1	0	0	0	0	0	20
00673	11	25	0	19	0	0	0	4	2
00757	11	20	0	0	0	17	0	3	0
00768	9	20	0	0	0	0	0	20	0
00867	14	24	0	24	0	0	0	0	0
01284	17	46	0	0	0	0	0	0	46
01693	21	40	0	0	1	0	0	39	0

[a]FP, fingerprint; AR, Arkansas, CA, California; MA, Massachusetts; MD, Maryland; MI, Michigan; NJ, New Jersey, TX, Texas.
[b]Patterns reported for ≥20 isolates.

identified among the patterns with 17 or more bands; however, we were less confident about identifying it in those with 16 bands and did not include them in this analysis. Of the 688 patterns analyzed, 497 (72.2%) were similar to FP 00237. Examples can be seen in Figure 3. Nearly all of the individual patterns, 480 (97%), were reported by a single site. These 497 patterns represent 946 isolates, 82% of all isolates with ≥17 copies of IS*6110* and 9% of all isolates in the study (Arkansas 3%; Maryland, 4%; New Jersey, 7%; Massachusetts, 9%; and California, Michigan, and Texas, 11%). The distribution of these isolates by site is reported in Table 3.

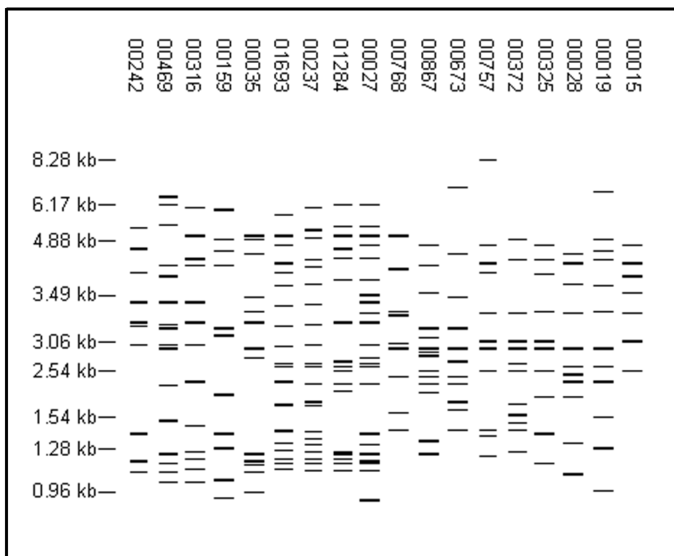

Figure 2. High-frequency, high-copy fingerprint patterns. Each pattern was reported for ≥20 isolates. The distribution of isolates with these patterns by sentinel surveillance site is shown in Table 2.

Because only one of the molecular criteria (overall fingerprint pattern) could be applied, isolates with these patterns cannot be definitively called W-Beijing. All of the insertion sites in strain 210, FP 00237, have been defined by sequencing (6). To identify conserved insertion sites, we determined the percentage of the 497 patterns that contained each of the bands in FP 00237 (Figure 4). Nine bands were found in >50% of the patterns, and two were present in >85%. A common feature of these fingerprint patterns is a group of smaller bands (1.0 to 1.5 kb) that are difficult to resolve. Variation in band identification resulted in some of the heterogeneity of the patterns in the database. W-Beijing strains likely account for a large portion of Beijing isolates, but other Beijing strains exist. FP 00242 (reported for 96 isolates in Texas; fingerprint pattern shown in Figure 2) shares only a few bands with FP 00237, but isolates with this pattern have the Beijing spoligotype (Teresa Quitugua, pers. comm.).

To identify other large families in the database, we analyzed all patterns having 6 to 15 bands (4,846 patterns). Since the BioImage software cannot create a dendrogram for more than 1,250 patterns, patterns were compared to each other by

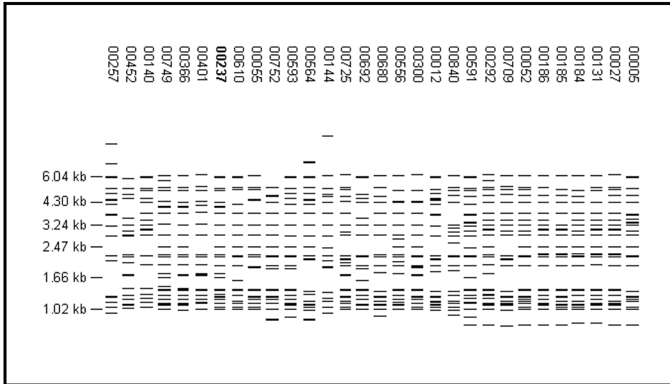

Figure 3. Examples of fingerprint patterns in the W-Beijing family that were visually identified as being similar to the prototype pattern, FP 00237.

using an arbitrarily chosen matching threshold of 50% to identify those that matched a large number of other patterns. For a 50% match, two thirds of the bands in two patterns with equal band number must match. Six sets of related fingerprints, designated A through F, were defined; each consisted of six prototype patterns (Figure 5) along with all of the patterns that matched the prototypes. Data on these sets are summarized in Table 3. The isolates in each set appear widely dispersed across the sites, and the patterns likely represent endemic strains in the United States. Key bands in each set were determined in comparison to the common bands in the prototype patterns as described above for FP 00237 (Figure 4). Sequencing the IS*6110* insertion sites corresponding to these key bands would allow isolates belonging to these sets to be rapidly identified with microarray techniques or the reverse dot-blot ("insite") assay we have described previously (7).

The sets described here are certainly not the only sets of related patterns in the database, nor are they necessarily novel. The patterns in set A are similar to the patterns for *M. tuberculosis* strains H37Rv and H37Ra (8); the eight common bands in the prototype patterns for set A are also found in the patterns for these two laboratory strains. The patterns in set D appear similar to those of the Haarlem family (9). Interestingly, 1,404 (29.0%) of the patterns with 6 to 15 bands did not match any other pattern at the 50% matching threshold, suggesting a substantial number of orphan strains in this study.

Low-Copy Isolates

Of the 457 fingerprint patterns identified among the 2,507 low-copy isolates, 314 (68.7%) were reported for a single isolate, and 143 patterns grouped 2,193 isolates into clusters. Clustering was much higher in low-copy isolates (87.5%) than in high-copy isolates (41.2%). Most isolates were in a few large clusters; 14 clusters contained 1,601 (63.9%) low-copy isolates. The distribution of isolates in the largest clusters across the sentinel surveillance sites is shown in Table 4, and the fingerprint patterns are shown in Figure 6.

Table 3. Distribution of isolates in genotype families[a]

			No. isolates/site								
	No. patterns	No. isolates	AR	CA	MA	MD	MI	NJ	TX	Clustered isolates (%)	Average no. copies of IS*6110*/isolate (range)
FP sets											
W-Beijing	497	946	22	279	88	49	162	144	202	56	19.9 (17–27)
Set A	141	190	37	18	14	19	43	29	30	39	13.1 (10–15)
Set B	97	162	14	33	20	4	15	33	43	52	11.0 (7–15)
Set C	251	414	4	295	9	15	20	53	18	52	11.9 (8–15)
Set D	181	275	11	37	44	32	27	102	22	48	8.9 (6–13)
Set E	119	137	0	37	5	16	16	56	7	20	12.9 (9–15)
Set F	177	321	24	44	31	47	47	71	57	55	9.4 (6–14)
FP 17	54	411	31	59	31	64	65	72	89	70	4.5 (3–6)
Spoligotype family											
EA-I	161	558	7	247	71	46	62	51	74	56	1.8 (1–6)
X	113	1,291	61	267	73	98	232	270	290	83	3.1 (1–6)
Haarlem	11	47	1	1	5	23	2	5	10	45	4.9 (1–6)
LAM-1	5	6	0	1	2	2	0	1	0	33	3.7 (1–6)
LAM-2	4	54	0	0	0	51	0	2	1	81	2.6 (1–6)
bovis	19	32	2	8	1	2	5	8	6	41	1.4 (1–5)
africanum	10	19	0	1	1	11	2	3	1	47	4.4 (3–6)

[a]FP, fingerprint; AR, Arkansas; CA, California; MA, Massachusetts; MD, Maryland; MI, Michigan; NJ, New Jersey; TX, Texas; EA-I, East African-India; LAM, Latin American-Mediterranean.

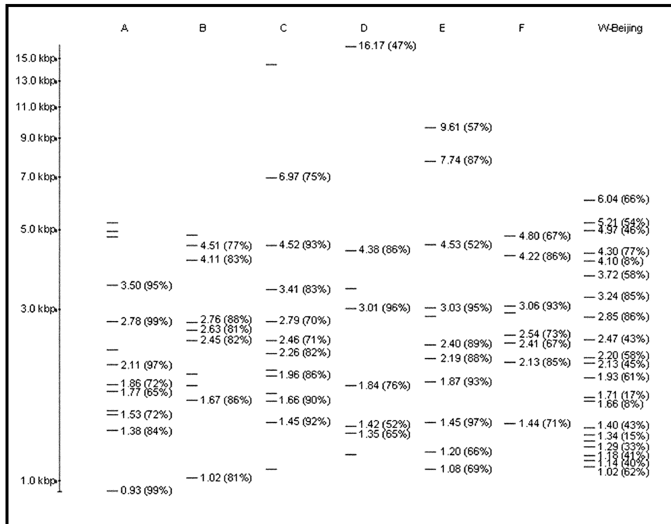

Figure 4. Prototype patterns for genotype sets. Set A: FP 00102; Set B: FP 04924; Set C: FP 02789; Set D: FP 02646; Set E: FP 02170; Set F: FP 04666; and W-Beijing family: FP 00237. The size of each band and the percentage of patterns in the set with each band are indicated on the pattern obtained by restriction fragment length polymorphism typing. For sets A–F, only the bands common to the six prototype patterns were analyzed.

Spoligotype results were available for most low-copy isolates (2,507 of 2,638 isolates). Isolates collected in Arkansas before 1998 were not spoligotyped (97 of 210 isolates) nor were most isolates collected in Maryland before 1998 that had unique fingerprint patterns (23 of 323 isolates). Of the 495 spoligotypes identified among the low-copy isolates, 322 were reported for a single isolate, and 173 grouped 2,185 isolates into clusters. In this study, the clustering of low-copy isolates by spoligotyping (87.2%) was only slightly lower than clustering by fingerprinting (87.5%). Analysis of the isolates by IS*6110* copy number showed that spoligotyping performed better than fingerprinting only for those isolates with fewer than four copies of IS*6110* (data not shown). Similar to the results obtained with fingerprinting, most isolates are in large clusters; 1,481 isolates are in the 20 largest clusters. The spoligotypes for these clusters as well as the distribution of these isolates by site and IS*6110* copy number are listed in Table 5.

Neither fingerprinting nor spoligotyping provided great discriminatory power among low-copy isolates, but the combination of the two methods gave slightly better results. The number of spoligotypes identified per fingerprint pattern ranged from 1–92 spoligotypes, and the number of fingerprint patterns identified per spoligotype ranged from 1–77 patterns. Combining the fingerprinting and spoligotyping data resulted in the identification of 987 distinct genotypes; 745 genotypes were unique, and 242 grouped 1,762 isolates into clusters. These genotype clusters contained up to 167 isolates. Performing the secondary typing method decreased the number of clustered isolates by nearly 20%, but clustering was still much higher among the low-copy isolates (70.2%) than among the high-copy isolates (41.2%).

In our recent study of low-copy isolates from Michigan, we noted numerous patterns with similarities to FP 00017 (10). In this study, 201 isolates from all seven sites had FP 00017. When the three lower bands (1.39, 2.32, and 3.03 kb) in FP 00017 were matched to all patterns having three to six bands, 54 patterns representing 411 isolates were identified. The distribution of these isolates by site can be seen in Table 3. Of note, *M. tuberculosis* strain CDC1551 has FP 00017 (11), but none of the study isolates with this fingerprint had the spoligotype corresponding to strain CDC1551 (7000767577 60771).

Spoligotypes have been divided into clades or families on the basis of commonly observed motifs (Figure 7) (12). First, spoligotypes can be subdivided on the basis of spacers 33–36. Only *M. tuberculosis* complex genotypic group 1 strains (*M. bovis*, *M. africanum,* and some *M. tuberculosis* strains) have spacers 33–36 (13,14). Four spoligotype motifs have been identified among group 1 isolates: bovis (15), africanum (16), Beijing (5), and East African-Indian (EA-I) (12,17) (Figure 7). The remaining spoligotypes that lack spacers 33–36 can be subdivided into two subgroups on the basis of spacers 29–32. Isolates with at least one of spacers 29–32 are likely to be isolates in *M. tuberculosis* genotypic groups 2 or 3. Isolates without spacers 29–32 have a deletion in the direct repeat locus that is too large to definitively assign to a genotypic group. Four specific motifs have been identified among the spoligotypes associated with non–genotypic group 1 isolates: Haarlem (9), Latin American and Mediterranean 1 and 2 (12,17), and X (12) (Figure 7). Of the 495 spoligotypes observed for low-copy isolates, 323 contained one of the eight defined motifs. This allowed 2,007 (80.1%) low-copy isolates to be assigned to a spoligotype family; the data for each family are summarized in Table 3. The majority (51.5%) of the low-copy isolates belonged to family X. The only published information

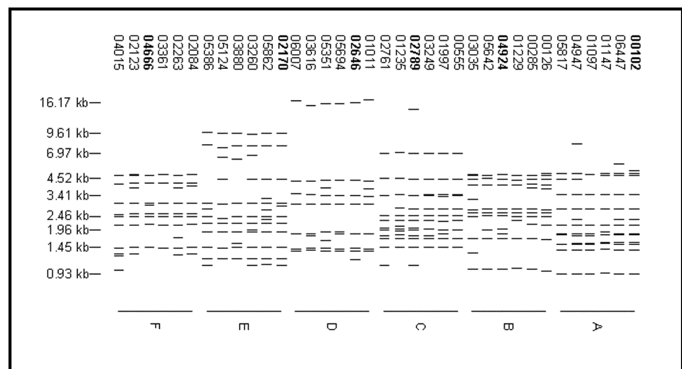

Figure 5. Prototype fingerprint patterns used to define genotype sets A–F. The prototype patterns were identified by first matching all patterns with 6–15 bands to each other at a matching threshold of 50%. The pattern that matched the greatest number of other patterns and the five patterns that matched this pattern and the greatest number of other patterns were selected as the prototype patterns for a set. All patterns that matched one of the six prototype patterns were then assigned to this set. The process was repeated by using the remaining unassigned patterns to create the six sets of fingerprint patterns. For each set, the six prototype patterns were visually examined for common bands. The mean, standard deviation, and coefficient of variation of each band in the six patterns were calculated, and bands with a coefficient of variation >2% were eliminated. The mean band size of the selected bands was used to identify which of the six patterns contained the common bands closest to the average band size (pattern identified with bold font).

Table 4. Distribution of isolates with low-copy fingerprint patterns reported with high frequency[a,b]

FP	No. bands	No. isolates	No. isolates/site[a]							No. spoligotypes[c]
			AR	CA	MA	MD	MI	NJ	TX	
00000	0	21	0	11	0	7	1	0	2	11
00003	1	87	2	32	13	12	11	12	5	26
00016	2	429	47	0	22	67	116	56	121	70
00017	4	201	12	15	13	39	29	48	45	45
00077	3	49	0	0	7	12	16	0	14	14
00129	1	289	1	84	60	35	29	66	14	92
00143	4	28	2	1	0	9	5	10	1	5
00195	1	148	5	76	0	5	13	3	46	52
00256	1	28	2	7	1	2	4	7	5	16
00370	3	38	0	9	0	0	0	26	3	9
00434	3	21	1	4	1	0	10	2	3	9
00456	1	32	0	0	0	0	0	32	0	18
00708	2	207	0	184	0	0	3	20	0	19
01285	4	23	0	0	0	0	0	1	22	4

[a]FP, fingerprint, AR, Arkansas; CA, California; MA, Massachusetts; MD, Maryland; MI, Michigan; NJ, New Jersey, TX, Texas.
[b]Patterns reported for ≥20 isolates.
[c]Number of different spoligotypes reported for isolates with this pattern.

regarding this motif indicated that it is highly prevalent in some English-speaking countries (12). In our study, 1,036 (70.3%) of isolates with two to four copies of IS*6110* belonged to family X. The second largest spoligotype family was family EA-I. Isolates with this motif belong to group 1 (13) and have up to nine copies of IS*6110* (17). Our isolates that belonged to this family had one to six copies of IS*6110*, but 378 (67.7%) possessed a single copy. In fact, 62.7% of isolates with a single copy of IS*6110* belonged to the EA-I family. The remaining spoligotype families grouped only a few isolates, probably because isolates in these families are mostly high copy (9,17), and this occurrence should not suggest that these spoligotype families are uncommon in the United States. Thirty-two isolates were classified as *M. bovis* and 19 as *M. africanum,* solely by spoligotype motifs; no additional tests were conducted to confirm this classification.

After isolates were assigned to a spoligotype family, fingerprint clusters of isolates were examined for consistency with the spoligotype family assignment. We were surprised to identify several fingerprint patterns that have isolates with very different spoligotype patterns. For example, FP 00017 (Figure 6) and FP 00104 (a five-band pattern) share four bands in common with a size difference of <1% and also have two spoligopatterns in common. SP 3 (777776777760771) is a very common pattern among *M. tuberculosis* group 2 and 3 isolates (13), whereas SP 290 (330777777767671) has a motif associated with *M. africanum* isolates (group 1). The spoligotype patterns are clearly divergent, indicating either that the strains independently acquired three copies of IS*6110* at the same insertion sites or that they have different IS*6110* insertions that coincidentally yield *Pvu*II fragments of the same length.

Most of the other examples of isolates clustered by IS*6110* with divergent spoligotypes are among isolates with one or two copies of IS*6110*. Mathema et al. (18) investigated differences among 66 isolates with FP 00129 (one band of 1.40 kb); 26 had group 1 spoligotypes, and 40 had group 2 or 3 spoligotypes. In most isolates with a single copy of IS*6110*, the IS*6110* is inserted in the direct repeat locus in the repeat located between spacers 24 and 25. The predicted fragment size for this insertion in isolates with group 1 spoligotypes is

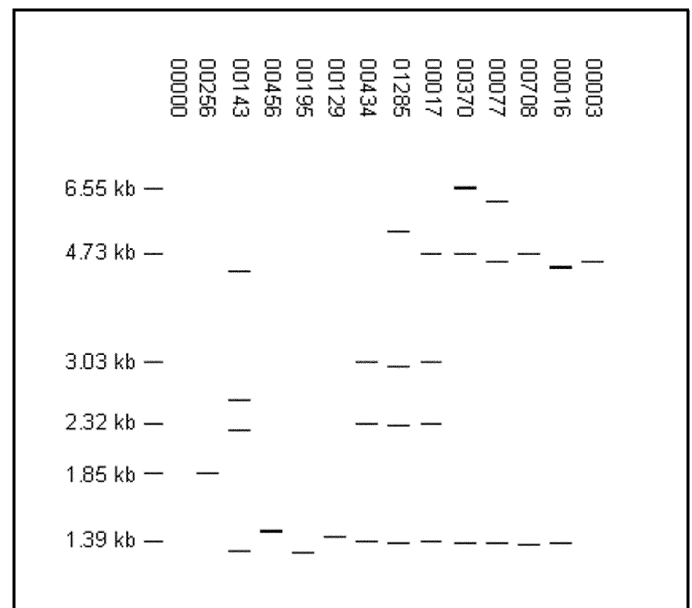

Figure 6. High-frequency, low-copy fingerprint patterns. Each pattern was reported for ≥20 isolates. The distribution of isolates with these patterns by sentinel surveillance site is shown in Table 4.

Table 5. Distribution of isolates with spoligotypes reported with high frequency[a,b]

SP	Octal code[c]	No. isolates	No. isolates/site							No. isolates/IS*6110* copy number						No. fingerprint patterns[d]
			AR	CA	MA	MD	MI	NJ	TX	1	2	3	4	5	6	
2	777777777760771	84	8	25	10	2	8	14	17	40	11	5	15	3	10	30
3	777776777760771	331	16	57	16	3	40	48	151	5	82	25	122	61	36	77
9	777776777760601	288	20	35	18	3	89	51	72	3	190	72	17	6	0	44
15	777777777413771	50	0	19	11	2	5	3	10	26	6	3	7	4	4	21
16	777777777416761	21	0	0	16	0	0	2	3	10	2	0	2	6	1	10
19	777777774413771	99	0	69	18	0	7	4	1	83	5	6	3	1	1	17
27	701776777760601	131	0	130	1	0	0	0	0	0	129	1	0	1	0	4
28	700036777760771	34	0	4	7	1	2	18	2	1	6	19	0	8	0	12
29	700076777760771	46	2	7	5	1	7	13	11	0	0	0	27	11	8	13
30	700036777760731	44	0	3	3	0	10	16	12	0	2	42	0	0	0	4
72	700076777760671	38	4	2	2	0	13	14	3	0	0	0	28	9	1	7
75	777776407760601	57	0	1	0	0	0	55	1	41	14	2	0	0	0	4
91	477777777741071	24	0	0	1	0	0	23	0	18	3	0	0	1	2	8
300	777756777760601	41	0	0	1	38	2	0	0	0	4	9	14	0	14	9
540	477777777413071	44	1	17	5	9	10	1	1	25	4	5	3	1	6	21
545	037776777760601	31	0	0	0	0	30	0	1	0	30	1	0	0	0	2
546	777777777413731	26	0	17	0	3	4	0	2	20	3	0	2	0	1	7
560	777777777760601	20	8	0	0	0	0	0	12	0	17	2	1	0	0	4
562	777777776413771	21	4	0	0	0	0	0	17	18	2	1	0	0	0	4
900	776377777740731	51	0	0	0	51	0	0	0	0	36	11	2	0	2	11

[a]AR, Arkansas; CA, California; MA, Massachusetts; MD, Maryland; MI, Michigan; NJ, New Jersey; TX, Texas.
[b]Spoligotype; reported for ≥ 20 isolates.
[c]The 43-digit spoligotype pattern is reported in the standard octal code format (2).
[d]Number of different fingerprint patterns reported for isolates with this spoligotype.

1.30–1.45 kb, depending on the number of spacers between spacers 25 and 36, where the *Pvu*II site is located. The predicted fragment size for this insertion in isolates with group 2 or 3 spoligotypes is 4.51–4.58 kb, depending on the number of spacers between spacers 25 and 43 (the next *Pvu*II site occurs outside of the direct repeat locus). Since the predicted fragment size for these 40 isolates was not consistent with the observed size, the insertion site in these isolates was sequenced. Sequencing showed that the isolates had a different insertion site (DK1) (19), which is very common among isolates with two copies of IS*6110*. The predicted fragment size for this insertion is 1.38 kb. This size is the one predicted for group 1 isolates and is a clear example of two isolates with different IS*6110* insertions yielding *Pvu*II fragments that are indistinguishable by the standard RFLP method.

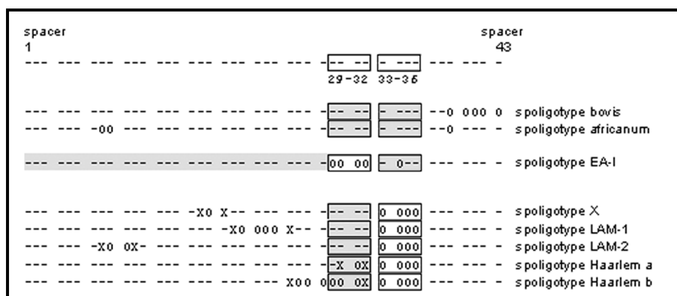

Figure 7. Motifs used to assign spoligotype patterns to spoligotype families. Each spoligotype was analyzed for the bovis (15), africanum (16), East-African-Indian (EA-I) (12,17), X (12), Latin American-Mediterranean 1 and 2 (12,17), and Haarlem a and b spoligotype motifs (9). Each motif definition was modified from the original references to ensure that motifs were not identified in a spoligotype pattern due to an unrelated deletion at the spacers of interest; each of the motif-defining absent spacers must be flanked on both sides by the adjacent spacer. The 43 spacers in the spoligotype pattern are classified with symbols: X: spacer must be present; 0: spacer must be absent; -: spacer may or may not be present; spacers in shaded boxes: at least one of the spacers in the box must be present.

Summary

The TB genotyping network database demonstrates the diversity of strains that cause TB in the United States. The 10,883 patients in the study represented approximately 11.6% of all new cases of TB in the United States from 1996 through 2000. The sentinel sites were reasonably representative of the geographic and demographic diversity in the United States. Compiling this database from results submitted from seven laboratories was a considerable undertaking, and analyzing such a large collection of fingerprint patterns is difficult. From

our quality assurance program and personal experience, we know that, even under the most carefully controlled conditions, IS*6110* fingerprinting results are not 100% reproducible. We are certain that some of the fingerprint patterns, which were classified as different and received different designations, would have been identical had they been run side by side on the same gel. Also, as we have described, some fingerprint patterns for low-copy isolates appear identical but do not represent the same IS*6110* insertions and thus do not represent closely related strains. Some of these difficulties resulted from the application of a rigid standard for defining distinct patterns, a process that is often subjective.

Even though some individual results may have been questionable, several clear conclusions emerged. Large sets of strains with related fingerprint patterns, not previously recognized, are spread across the United States. Given the rather slow rate of change in fingerprints, these must represent endemic strains that have circulated in the United States for decades. Consistent with this conclusion is the presence in the database of fingerprint patterns resembling the pattern of the laboratory strain H37 that was originally isolated in New York in 1905 (8). Many of the patterns in these sets represent single isolates, which suggests that they are the result of reactivation of remote infections acquired years or decades earlier. Analysis of the demographic characteristics of the patients will be required to confirm this observation. Among these large sets, outbreak strains (patterns) were generally restricted to a single sentinel site, as were clustered isolates in general.

We conclude that a large-scale, prospective comparison of fingerprint patters from wide geographic regions is useful for research studies but is of limited value for TB control purposes. Comparisons of isolates from smaller areas are not only more meaningful but also more feasible. This limitation does not mean that searching multiple databases for specific fingerprint patterns, for example the “W” strain, is not useful in some circumstances.

The difficulties in analyzing IS*6110* fingerprint patterns and the often slow turnaround time for obtaining results limit the value of this procedure to TB control programs. As an alternative, rapid, polymerase chain reaction–based testing, such as spoligotyping or mycobacterial interspersed repetitive units variable number of tandem repeats (MIRU-VNTR) analysis, would be a logical first step for universal genotyping of isolates. These methods provide greater reproducibility and give digital results, which simplify analysis. However, this approach has the following limitations. Many common spoligotypes were seen among low-copy-number isolates, although even IS*6110* fingerprinting does not greatly improve resolution with these isolates. We also found that 9% of the isolates have W-Beijing fingerprint patterns that are known to have the same spoligotype; all isolates yielding the Beijing spoligotype would require IS*6110* typing. Sufficient data are not available to predict the discriminatory power of MIRU-VTNR. However, preliminary results suggest that the combination of spoligotyping and VNTR typing will provide adequate resolution for most uses, thus limiting the need for additional typing by IS*6110* fingerprinting.

Acknowledgments

We thank Steve Kammerer, and Charles L. Woodley for assistance in database analysis, and Barbara A. Schable, Christopher R. Braden, and the participants in the National Tuberculosis Genotyping and Surveillance project for their extensive efforts in compiling the databases on which this article is based.

Dr. Cowan is a service fellow in the Tuberculosis/Mycobacteriology Branch, Division of AIDS, STD, and TB Laboratory Research, National Center for Infectious Diseases, Centers for Disease Control and Prevention. Her current research interests include DNA fingerprinting and the molecular epidemiology of *Mycobacterium tuberculosis*.

Dr. Crawford is Chief of the Immunology and Molecular Pathogenesis Section, Tuberculosis/Mycobacteriology Branch, Division of AIDS, STC, and TB Laboratory Research, Centers for Disease Control and Prevention. His research interests include application of molecular methods to epidemiology and diagnostics of mycobacterial diseases.

References

1. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192-6.
2. Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, et al. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardized nomenclature. Int J Tuberc Lung Dis 2001;5:216–9.
3. Yang Z, Barnes PF, Chaves F, Eisenach KD, Weis SE, Bates JH, et al. Diversity of DNA fingerprints of *Mycobacterium tuberculosis* isolates in the United States. J Clin Microbiol 1998;36:1003–7.
4. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol 2002;10:45–52.
5. van Soolingen D, Qian, L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. J Clin Microbiol 1995;33:3234–8.
6. Beggs ML, Eisenach KD, Cave MD. Mapping of IS*6110* insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J Clin Microbiol 2000;38:2923–8.
7. Steinlein LM, Crawford JT. Reverse dot blot assay (insertion site typing) for precise detection of sites of IS*6110* insertion in the *Mycobacterium tuberculosis* genome. J Clin Microbiol 2001;39:871–8.
8. Bifani P, Moghazeh S, Shopsin B, Driscoll J, Ravikovitch A, Kreiswirth, BN. Molecular characterization of *Mycobacterium tuberculosis* H37Rv/Ra variants: distinguishing the mycobacterial laboratory strain. J Clin Microbiol 2000;38:3200–4.
9. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PWM, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
10. Cowan LS, Mosher L, Diem L, Massey JP, Crawford JT. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low-copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J Clin Microbiol 2002;40:1592–602.
11. Plikaytis BB, Kurepina N, Woodley CL, Butler WR, Shinnick TM. Multiplex PCR assay to aid in the identification of the highly transmissible *Mycobacterium tuberculosis* strain CDC1551. Tuber Lung Dis 1999;79:273–8.

12. Sebban M, Mokrousov I, Rastogi N, Sola C. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. Bioinformatics 2002;18:235–43.
13. Soini H, Pan X, Amin A, Graviss EA, Siddiqui A, Musser JM. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J Clin Microbiol 2000;38:669–76.
14. Sreevatsan S, Pan X, Stockbauer K, Connell ND, Kreiswirth BN, Whittam TS, 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.
15. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
16. Viana-Niero C, Gutierrez C, Sola C, Filliol I, Boulahbal F, Vincent V, et al. Genetic diversity of *Mycobacterium africanum* clinical isolates based on IS*6110*-restriction fragment length polymorphism analysis, spoligotyping, and variable number of tandem DNA repeats. J Clin Microbiol 2001;39:57–65.
17. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J Mol Evol 2001;53:680–9.
18. Mathema B, Bifani PJ, Driscoll J, Steinlein L, Kurepina N, Moghazeh SL, et al. Identification and evolution of an IS*6110* low-copy *Mycobacterium tuberculosis* cluster. J Infect Dis 2002;185:641–9.
19. Fomukong N, Beggs M, El Hajj H, Templeton G, Eisenach K, Cave MD. Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high-copy numbers of IS*6110*. Tuber Lung Dis 1998;78:109-16.

Address for correspondence: Jack T. Crawford, Centers for Disease Control and Prevention, Mailstop F08, 1600 Clifton Rd., NE, Atlanta, GA 30333 USA; fax: 404-639-1287; e-mail: jcrawford@cdc.gov.

Molecular Typing of *Mycobacterium tuberculosis* Strains with a Common Two-Band IS*6110* Pattern

Kerry H. Lok,* William H. Benjamin, Jr.,* Michael E. Kimerling,* Virginia Pruitt,† Donna Mulcahy,† Nancy Robinson,† Nancy B. Keenan,† and Nancy E. Dunlap*

We conducted a progam of population-based molecular typing of all *Mycobacterium tuberculosis* isolates obtained in Alabama since 1994. Of 2,452 isolates, 1,013 (41%) had fewer than 6 bands of IS*6110*; 348 (14%) had a single two-band pattern (JH2). With conventional epidemiologic methods, we identified three groups of related patients with JH2 isolates. Spoligotyping and pattern of variable number of tandem repeats identified 10 molecular groups; two found by conventional methods were subdivided.

To achieve the national goal of tuberculosis (TB) elimination, disease control efforts must be based on a thorough understanding of transmission patterns among the general population and among groups at particular risk (1,2). In close coordination with the Alabama Department of Public Health, we conducted a program of population-based molecular typing of all *Mycobacterium tuberculosis* isolates identified since 1994 in Alabama. The state contains a mixture of rural and urban settings with a stable population, and the proportion of tuberculosis cases among the foreign-born and HIV-positive groups is low (2).

The Study

Through May 2001, we typed 2,452 isolates using the IS*6110* restriction fragment length polymorphism (RFLP) method and have combined the results of molecular analysis with an aggressive contact investigation program. We found that 1,013 (41%) of the IS*6110* patterns had ≤6 bands, including 348 (14%) with a two-band RFLP pattern called JH2. Conventional epidemiologic techniques identified three groups with related JH2. Resistance to a single drug (isoniazid or streptomycin) was associated with two groups, and homelessness was associated with a large group. Results from the latter group have been previously reported (3,4).

To further characterize JH2, the most common RFLP pattern statewide, we used two polymerase chain reaction (PCR)–based methods for secondary typing methods, spacer oligonucleotide type (spoligotyping) and variable number of tandem repeats (VNTR). Isolates were selected for further molecular typing to determine which method or combination of methods could help us understand the epidemiology of this important RFLP pattern in Alabama and the extent of molecular relatedness in the JH2 cluster.

*University of Alabama School of Medicine, Birmingham, Alabama, USA; and †Alabama Department of Public Health, Montgomery, Alabama, USA

We selected 102 (29%) of the 348 JH2 pattern isolates from the Alabama genotype database that represented two-band IS*6110* RFLP strains. Groups of isolates identified through conventional epidemiologic links were selected for evaluating the secondary molecular typing techniques. Fifteen isolates from northeast Alabama were obtained from a conventional epidemiologic investigation that showed a convenience store as the site of transmission (5). Twenty-four isolates from central Alabama came from a previous investigation (3,4) of a TB outbreak in a large community of homeless persons, and four isolates were from a known outbreak on a school bus in southeast Alabama. The remaining 59 isolates were selected from the general strain collection for comparison. The selection was based on availability of viable cultures for DNA extraction and the year that these outbreaks occurred. Two isolates from the nonclustered strain collection did not generate an adequate VNTR pattern and were excluded from the final analysis. Therefore, 100 isolates were available for the combined analysis.

Isolates of *M. tuberculosis* were cultured on Lowenstein-Jensen or 7H11 Middlebrook plates for at least 4 weeks before DNA extraction. Chloroform-isoamyl alcohol was used to extract chromosomal DNA from isolates, and IS*6110* RFLP typing was performed according to international standards (6,7). All fingerprint images were scanned into a computerized database and Whole Band Analyzer, version 4.01 (Genomic Solutions Inc., Ann Arbor, MI, USA) was used. Spoligotyping was used as a secondary typing method on the basis of the presence or absence of 43 variable spacers in the direct repeat region of *M. tuberculosis* (3,8–12). Spoligotyping membranes were used according to manufacturer's recommendations (Isogen Bioscience BV, Bilthoven, the Netherlands), and the spacer regions were numbered as reported previously (8). A spreadsheet was used to analyze the spoligotyping results. For national database reporting purposes, we converted the spoligotyping image into an octal digital format (13).

VNTR typing was used to further investigate strains that matched by both IS*6110* and spoligotyping methods or had a known epidemiologic link. This typing yielded 86 samples. As previously described (14), we used seven sets of primers and individually amplified regions with known informative tandem repeats. Electrophoresis was conducted on the PCR products in agarose. The ethidium bromide–stained gel was imaged to determine the size of the amplicon and the number of tandem repeats in each locus. We used a DNA ladder and amplicons from strain H37Rv to determine fragment sizes, according to Frothingham and Meeker-O'Connell (14).

Results

We used spoligotyping to separate the JH2 isolates into 20 different subtypes. Four of the 20 patterns contained >2 isolates. Eighty-four of the 100 isolates remained clustered after spoligotyping. The largest spoligotyping group (designated CDC spoligo 9) contained 55 isolates. The other 3 clustered patterns were designated CDC spoligotype 3 (n=12 isolates), CDC spoligotype 545 (n=14 isolates), and CDC spoligotype 550 (n=3 isolates). Two of the clusters previously identified through contact epidemiology were subdivided.

Eighty-six strains were typed with VNTR, which generated 15 different VNTR profiles (Table), of which the largest group contained 25 isolates. Nine profiles contained more than one isolate, and six isolates were unique. The number of isolates in the groups varied from 2 to 25. The largest cluster (homeless, n=24), which was identified previously through contact epidemiology, was subdivided into 10 subtypes.

Among those isolates with both secondary typing results, we identified 10 clusters plus 9 unique isolates. Clustered cases accounted for 77 of 100 cases. The largest spoligotype cluster of 55 isolates (CDC spoligotype 9), after combination with the VNTR results, was reclassified into nine unique profiles, including seven different clusters. The largest cluster had 22 isolates. CDC spoligotype 3 was divided into a single cluster of seven isolates and five unique profiles. CDC spoligotype 545 and CDC spoligotype 550 each represented single clusters.

The two-band JH2 pattern was the most common IS*6110* fingerprint found during nearly 8 years of TB genotyping surveillance in Alabama. This pattern matches National Tuberculosis Genotyping Fingerprint Pattern (NTGFP) 00016 in the database of the National Tuberculosis Genotyping Surveillance Network. 00016 was also the most common fingerprint pattern found during 5 years of the TB genotype project and was reported from all seven sentinel surveillance sites with a final frequency of 5% of all isolates.

Conclusion

Pattern JH2 represents a group of similar IS*6110* patterns.

Table. Comparison of a common two-band IS*6110* RFLP pattern with spoligotyping and VNTR[a]

Spoligotyping		VNTR		Related to conventional outbreak		
Octal (CDC designation)	No. of isolates	VNTR results (AL designation)	No. of isolates	Store	Bus	Homeless
777776777760601(0009)	55	6,3,2,4,2,3,2 (01)	22			2
		6,3,2,3,2,2,2 (02)	12		4	4
		6,3,2,4,2,3,3 (03)	8			2
		6,3,2,3,2,2,3 (04)	4			3
		6,3,2,2,2,2,2 (05)	3			3
		6,3,2,3,2,3,2 (06)	2			
		6,3,2,3,2,3,3 (07)	2			1
		6,3,2,4,2,2,2 (08)	1			
		6,2,2,3,2,2,2 (09)	1			1
037776777760601(0545)	14	6,3,2,3,2,4,2 (10)	14	14		
777776777760771(0003)	12	6,3,2,2,2,1,2 (11)	7			5
		6,3,2,3,2,3,2.5 (12)	1			
		6,3,2,2,2,1,3 (13)	1			
		6,3,1,2,2,2,3 (14)	1			1
		6,3,2,3,2,2,2 (02)	1			1
		6,3,2,3,2,3,3 (07)	1			
001776777760601(0550)	3	6,3,2,4,2,3,2 (01)	3			
777776374360711(1445)	1	6,3,2,1,2,3,2.5 (15)	1			1
037766677760601(n/a)	1	6,3,2,3,2,4,2 (10)	1	1		
Total	86		86	15	4	24

[a]RFLP, restriction fragment length polymorphism; VNTR, variable number of tandem repeats; CDC, Centers for Disease Control and Prevention.

The smaller band at 1.46 kb is conserved, and the larger band is located in direct repeat 24, a common insertion site. Consequently, the size of the larger band varies, depending on the number of direct repeats located upstream of spacer 24. A strain's spoligotype can predict the size of the larger band; the larger bands analyzed in our study were 4.5 kb–4.8 kb. The resolution of the gel and the Whole Band Analyzer (Genomic Solutions, Inc.) do not discriminate well between these sizes, which could result in different spoligotypes of a single IS*6110* insertion pattern being considered as different RFLP patterns.

The prominence of pattern JH2 (NTGFP 00016) statewide and throughout the network suggests an older, more stable IS*6110* pattern. A group of strains likely spread throughout the general population early in the TB epidemic (19th and early 20th centuries) and remain endemic in the 21st century in the United States. Although this pattern is found rarely in European isolates, some countries in Africa report a high prevalence (15).

Despite large numbers of the IS*6110* pattern found in Alabama, we could define only a few distinct groups through conventional epidemiologic methods. Using VNTR and spoligotyping techniques to reexamine some of these clusters, we gained a better understanding of disease transmission among community groups. Two clusters previously classified by conventional epidemiologic methods were confirmed by this study. Both clusters were associated with unique drug resistance patterns in outbreaks involving a school bus (isoniazid resistance) and a neighborhood convenience store (streptomycin resistance). The largest cluster, which involved the homeless community, revealed multiple molecular-based subclusters not identifiable by routine epidemiologic study, drug susceptibility data, or both. These subclusters signify the actual genetic diversity of this most prevalent IS*6110* pattern.

The importance of secondary typing for low-copy IS*6110* strains is well accepted (9–12), as is the necessity of combining conventional methods with those of molecular epidemiology (2–4). We found that VNTR is particularly well suited as a secondary typing method for the common JH2 pattern. However, the VNTR method is more expensive and time-consuming because seven PCRs are required. VTNR may be a preferred method for TB research, but not for widespread TB surveillance. Automated systems could reduce overall costs, however. While spoligotyping was less discriminatory in this study, we believe it should be evaluated carefully for use in other settings.

Support for this project was provided by a cooperative agreement between the Centers for Disease Control and Prevention and the Alabama Department of Public Health (Agreement C10119105) and by all sites of the National Tuberculosis Genotyping and Surveillance Network. Support was also provided through the Pulmonary Disseminated Infectious Diseases in HIV training grant from (No. T32AI07441) the National Institutes of Health.

Dr. Lok is a postdoctoral trainee in the Pulmonary, Allergy, and Critical Care Division, of University of Alabama School of Medicine, Birmingham, Alabama. His research interests include the molecular epidemiology of tuberculosis.

References

1. Jasmer R, Hahn J, Small P, Daley C, Behr M, Moss A, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 1999;130:971–8.
2. Dunlap NE. The use of RFLP as a tool for tuberculosis control: utility or futility? Int J Tuberc Lung Dis 2000;4:S134–8.
3. Kimerling M, Benjamin W, Lok K, Curtis G, Dunlap N. Restriction fragment length polymorphism screening of *Mycobacterium tuberculosis* isolates: population surveillance for targeting disease transmission in a community. Int J Tuberc Lung Dis 1998;2:655–62.
4. Kimerling ME, Shakes CF, Carlisle R, Lok KH, Benjamin WH, Dunlap NE. Spot sputum screening: evaluation of an intervention in two homeless shelters. Int J Tuberc Lung Dis 1999;3:613–9.
5. Kimerling ME, Brook NL, Lok KH, Benjamin WH Jr., Dunlap NE. *Mycobacterium tuberculosis* at a neighborhood convenience store [abstract]. Am J Respir Crit Care Med 1998;157:708.
6. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
7. Yang ZH, Ijaz K, Bates JH, Eisenach KD, Cave MD. Spoligotyping and polymorphic GC-rich repetitive sequence fingerprinting of *Mycobacterium tuberculosis* strains having few copies of IS*6110*. J Clin Microbiol 2000;38:3572–6.
8. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol 1993;10:1057–65.
9. Goyal M, Saunders NA, van Embden JD, Young DB, Shaw RJ. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J Clin Microbiol. 1997;35:647–51.
10. Soini H, Pan X, Amin A, Graviss EA, Siddiqui A, Musser JM. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J Clin Microbiol 2000;38:669–76.
11. Soini H, Pan X, Teeter L, Musser JM, Graviss EA. Transmission dynamics and molecular characterization of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J Clin Microbiol 2001;39:217–21.
12. Cronin WA, Golub JE, Magder LS, Baruch NG, Lathan MJ, Mukasa LN, et al. Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J Clin Microbiol 2001;39:3709–11.
13. Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, et al. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardized nomenclature. Int J Tuberc Lung Dis 2001;5:216–9.
14. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–96.
15. Hermans PW, Messadi F, Guebrexabher H, van Soolingen D, de Haas PE, Heersma H, et al. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J Infect Dis 1995;171:1504–13.

Address for correspondence: Michael E. Kimerling, University of Alabama at Birmingham, RPHB 217, 1530 3rd Ave. South; Birmingham, AL 35294-0022 USA; fax: 205-934-1746; e-mail: kimerlin@uab.edu

Spoligologos: A Bioinformatic Approach to Displaying and Analyzing *Mycobacterium tuberculosis* Data

Jeffrey R. Driscoll,* Pablo J. Bifani,† Barun Mathema,‡ Michael A. McGarry,* Genét M. Zickas,* Barry N. Kreiswirth,‡ and Harry W. Taber*

Spacer oligonucleotide (spoligotyping) analysis is a rapid polymerase chain reaction–based method of DNA fingerprinting the *Mycobacterium tuberculosis* complex. We examined spoligotype data using a bioinformatic tool (sequence logo analysis) to elucidate undisclosed phylogenetic relationships and gain insights into the global dissemination of strains of tuberculosis. Logo analysis of spoligotyping data provides a simple way to describe a fingerprint signature and may be useful in categorizing unique spoligotypes patterns as they are discovered. Large databases of DNA fingerprint information, such as those from the U.S. National Tuberculosis Genotyping and Surveillance Network and the European Concerted Action on Tuberculosis, contain information on thousands of strains from diverse regions. The description of related spoligotypes has depended on exhaustive listings of the individual spoligotyping patterns. Logo analysis may become another useful graphic method of visualizing and presenting spoligotyping clusters from these databases.

The process of DNA fingerprinting *Mycobacterium tuberculosis* isolates provides epidemiologists with data for investigating transmission and confirming laboratory cross-contamination. When personnel and resources are available, strains from all newly diagnosed cases of tuberculosis (TB) are fingerprinted. Interpreting the data from these sentinel surveillance studies can be challenging. TB can be latent for decades before resulting in active disease. The best method for examining *M. tuberculosis* complex genotyping data gathered over just a few years is still in question.

The genotyping of *M. tuberculosis* complex has been undertaken in the United States in directed studies and in sentinel surveillance. Since 1992, *M. tuberculosis* complex DNA fingerprinting for isolates has been available to U.S. Departments of Health and TB Control Offices to investigate cases of suspected TB transmission and suspected laboratory cross-contamination. At the same time, sentinel surveillance of select regions was used to evaluate if genotyping every new patient isolate from a particular region was useful. Although direct DNA fingerprinting studies are relatively simple to design, we are still learning how best to use in toto the large and diverse databases generated by sentinel surveillance studies from multiple laboratories (1–7). Previous studies have used DNA fingerprinting methods to understand the development and spread of subspecies of the *M. tuberculosis* complex, such as *Mycobacterium africanum* (8,9), *M. bovis* (10), and the W-Beijing family (11,12).

*New York State Department of Health, Albany, New York, USA; †Institut Pasteur de Lille, Lille, France; and ‡Public Health Research Institute, New York, New York, USA

Analyzing the spread and evolution of *M. tuberculosis* complex strains is more complicated because of the long incubation of disease and relatively short-term collection of data (approximately 10 years). We examine whether bioinformatic tools can help in analyzing the data collected. Bioinformatics uses sophisticated analyses of large amounts of genetic information to clarify the relationships between species, explain the evolution of groups of genes, and assemble information from genome sequencing projects. Bioinformatic analysis involves searching nucleic acid or protein sequence information for previously unrecognized motifs that may signal previously unrecognized regions of interest. Spoligotype analysis (13) is a form of DNA sequencing by hybridization. Several groups have used some of these novel analytical tools to examine *M. tuberculosis* complex genotyping (4,6,14). Sequence logo analysis can find motifs in potentially related nucleic acid or protein sequences (15). Logo analysis combines these data on a graphic that illustrates the location and degree of sequence conservation in the selected sequences. We applied sequence logo analysis to find motifs based on the presence or absence of specific spacer sequences. We also evaluated the usefulness of logo analysis in examining phylogenetic relationships of the *M. tuberculosis* complex direct repeat (DR) (7) locus and its potential as a simple graphic method presenting grouped spoligotyping data.

We suggest using the sequence logo technique to understand the distribution of each spacer sequence used in the spoligotype assay. This information is useful in improving or redesigning the spoligotype assay by showing the degree of differentiation achieved with each spacer sequence.

Methods

Isolates

Approximately 5,100 isolates of *M. tuberculosis* complex, predominately from patients in TB control programs in the northeast United States, are part of the Wadsworth Center spoligotype database. Most of these strains have been collected through ongoing sentinel surveillance projects in Massachusetts and New York City from 1996 to the present. In our strain collection, 920 different spoligotype patterns were identified.

DNA Fingerprinting Analyses

Spoligotype analysis was performed according to a standard method (13). PCR amplifications were performed on extracted DNA or cell suspensions, which were heat-killed at 80°C for 1 hr in an oven. Spoligotype patterns were analyzed with BioImage Whole Band Analysis v3.4 software (Genomic Solutions, Ann Arbor, MI) on a Sun Ultra10 workstation (Sun Microsystems Inc., Santa Clara, CA). Spoligotype patterns were given descriptive nomenclature according to the standard method (16), along with a unique arbitrary numeric designation by the Centers for Disease Control and Prevention (CDC).

IS*6110*-based restriction fragment length polymorphism (RFLP) analysis was performed, according to standard protocol (17), at the Wadsworth Center (Albany, NY) and the Public Health Research Institute (New York City, NY), on approximately 3,700 of these isolates. RFLP patterns were analyzed with BioImage Whole Band Analysis v3.4 software (Genomic Solutions). RFLP pattern designations for sentinel surveillance isolates were assigned a unique arbitrary numerical designation by CDC.

Spoligologo Analysis

Spoligologo denotes the application of sequence logo analysis to spoligotype assay data. Sequence logo analysis was originally devised as a method to find blocks of related amino acids between protein sequences and display the information in an intuitive visual description that illustrates both the residue and the degree of conservation at each position (18). Logo analysis has been used to look for functional and evolutionary relationships among groups of proteins and nucleic acids (19). Spoligologo analysis (15) was accomplished by using WebLogo software from the School of the Biological Sciences at the University of Cambridge (available from: URL: http://www.bio.cam.ac.uk/seqlogo/). To be compatible with WebLogo, letter designations (x=hybridization, o=no hybridization signal) were used to denote the pattern of hybridization observed for each spoligotype pattern. Spoligotype patterns to be compared were entered directly into WebLogo and a postscript file of the results was generated. Logo analysis compares each selected pattern against the other at the same position. Thus, we compared hybridization to spacer 1 in the group of selected spoligotypes, followed by analysis of spacer 2, and so on for all 43 spacers.

For convenience of illustration, two groups of isolates were chosen for spoligologo analysis. The first set consisted of 43 strains of *M. bovis* (13) in our collection. The second set was 12 low-band (exhibiting fewer than six copies of IS*6110* by RFLP analysis) *M. tuberculosis* isolates from Vietnam-born, Massachusetts sentinel surveillance case-patients. Vietnamese patients are the largest foreign-born group represented in the low-band data from Massachusetts.

Results

Figure 1 illustrates logo analysis with spoligotyping data. Spoligotyping identified 28 different spoligotypes associated with *M. bovis* from 43 isolates in our collection. The spoligotypes were then coded for sequence logo analysis (Figure 1A). Letter designations were chosen for compatibility with the WebLogo program as described in Methods. The tallest x and o characters represent areas of absolute concordance between the patterns chosen for analysis. Where differences occur, the

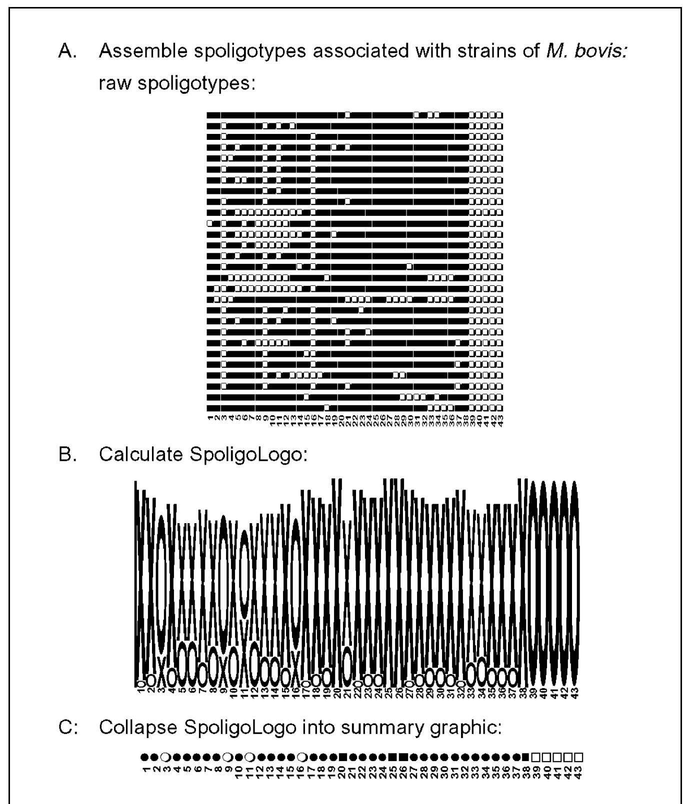

Figure 1. Logo analysis on spoligotypes associated with *Mycobacterium bovis*. The Wadsworth Center database contains 28 unique spoligotypes from strains of *M. bovis*. Panel A illustrates the raw hybridization data followed by the same patterns coded for logo analysis. To be compatible with WebLogo analysis, patterns were converted to a 43-character–long string consisting of the letters x and o. The letter x represents a positive hybridization, and o represents no hybridization detected for each of the 43 spacer sequences. Panel B is the graphic output from WebLogo. Numbers in each panel represent the spoligotype assay spacer sequences 1–43. Panel C shows the summary graphic of the spoligotypes by collapsing the data into a single row. Legend: x=hybridization observed to spacer, o= no hybridization observed to spacer, ■= positive hybridization in every spoligotype pattern for that individual spacer sequence, □= no hybridization, ●=positive hybridization in ≥50% of the patterns, ❍= no hybridization in >50% of the patterns.

ratio of those spoligotypes showing hybridization to those that do not is represented by relative height differences of the characters in that column. The resulting spoligologo (Figure 1B) shows that spacers 20, 25, 26, and 38 are present in all 28 spoligotype patterns. The absence of spacers 39–43 in these spoligotype patterns is consistent with an identification of *M. bovis* (13). The greatest polymorphism between the patterns appears in spacers 3 through 16 (Figure 1B). Spoligotype patterns probably evolve through the loss of spacer sequences through a variety of mechanisms (7,20). We try to extrapolate back to the hypothetical founder of these *M. bovis* isolates, which we believe had spacers 1–38 and not 39–43. No isolate with this probable *M. bovis* founder spoligotype has yet been observed in the Wadsworth Center spoligotype database.

Foreign-born persons, especially those from regions with high TB case rates, are of concern for TB transmission in the United States (21,22). The low-band *M. tuberculosis* strains from Vietnam-born patients, which we selected, produced another spoligologo pattern (Figure 2). The ability of spoligologo analysis to collapse even a small selection of spoligotype patterns from a select group of strains (Figure 2A) into the possible founder spoligotype can be observed in the summary (Figure 2C). However, additional typing methods would be required to verify that the strains are related, rather than exhibiting convergent evolution of their respective spoligotypes.

As previously suggested (5,7,23), *M. tuberculosis* strains that generally contain spacers 33–43 may form a family that is an intermediate lineage between the Beijing (11,12) and non-Beijing *M. tuberculosis* strain families, such as Haarlem (2).

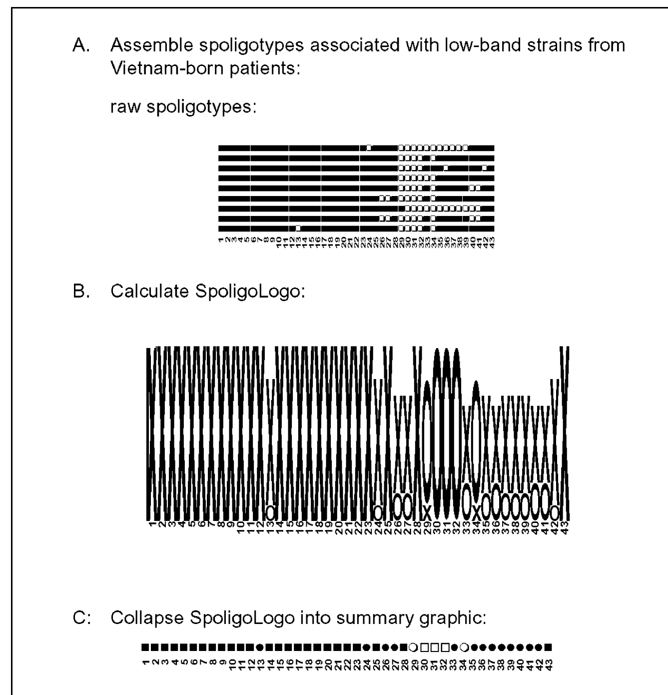

Figure 2. Logo analysis on nine different spoligotypes observed for *Mycobacterium tuberculosis* isolates from Vietnam-born patients in Massachusetts demonstrating fewer than six copies of IS*6110* by RFLP analysis. Legend: x= hybridization observed to spacer, o= no hybridization observed to spacer, ■= positive hybridization in every spoligotype pattern for that individual spacer sequence, □= no hybridization, ●=positive hybridization in ≥50% of the patterns, ❍= no hybridization in >50% of the patterns.

Discussion

Spoligotyping, microarrays, and DNA-chips are all examples of reverse-hybridization array-based assays. Although microarrays and DNA-chips can contain thousands of bits of data, the principle behind them is similar to that of the spoligotype assay, which uses a simple 1 x 43 array. Array-based assays use reverse hybridization in which a labeled sample is probed against a series of proteins or nucleic acids that are bound to a solid support, such as a nylon membrane or silica. The result for each potential binding event can be recorded as yes or no. The binary nature of array-based assays allows the data to be analyzed usefully with algorithms associated with motif recognition, such as sequence logo analysis. The relative low cost and simplicity of the spoligotype assay means it can be performed by many laboratories and the digital nature of the data facilitates the exchange of information among researchers.

The growth in the availability of array-based assays has outpaced the ability of conventional software analysis packages to provide every possible method of analysis. Customized versions of software are extremely expensive, and researchers, who want to implement these protocols without specialized software, lack methods of collating the large amounts of data. Problems arise when attempts are made to judge the significance of similar but nonidentical array data. Identifying possible families in these array patterns may be important in understanding the evolution and spread of pathogens such as *M. tuberculosis*. Our method for collapsing array-based data can be used to find and present patterns or signature motifs in these types of data.

As previously noted (1), cluster analysis of large RFLP databases is difficult for a combination of reasons, including software failure and interlaboratory variations. Digital data, such as spoligotyping, mycobacterial interspersed repetitive unit (24), and variable number tandem repeat analyses (25), will probably form the basis for any large DNA fingerprinting projects in the future (26).

The design of large genotyping projects should include multiple methods (2,3,10,23). Strains that cluster by one typing method must be analyzed by other methods to ensure that the groupings represent clusters of true relatedness and not cases of convergent evolution.

Bioinformatic analyses, like logo analysis, may prove useful in obtaining further data from the large *M. tuberculosis* complex DNA typing databases already in existence. Spoligologo analysis is a graphic method of presenting similar spoligotypes that may elicit useful insights into the geographic spread of tuberculosis. Potential families of TB strains could be identified on the basis of their logo; these strains could then be analyzed by additional DNA-typing methods to confirm the relationship, followed examining relevant patient data (e.g., country of birth). A similar analysis was performed for the *M. tuberculosis* W-Beijing family (11) that helped elucidate the

evolution of a multidrug–resistant strain in New York City. Spoligologo analysis could help identify more of these families, determine their global origin, and evaluate their spread.

Determining the sources and spread of tuberculosis is an important tool in preventing further infections. Understanding the geographic origin of an *M. tuberculosis* DNA fingerprint could be useful, especially in understanding the sources and spread of strains in the U.S. foreign-born population, among whom differentiating recently transmitted disease from reactivation of a past exposure can be difficult.

Acknowledgments

We thank Ann Miller and Sharon Sharnprapai for patient data; Max Salfinger, Linda Parsons, Ben Zhao, Adelah Ebrahimzadeh, Paul Elvin, and Alissa Scharf for processing and shipping of cultures; and Christophe Sola for a critical reading of the manuscript.

This research was supported in part by the Centers for Disease Control and Prevention, National Tuberculosis Genotyping and Surveillance Network cooperative agreement, and the Pittsfield Massachusetts Anti-Tuberculosis Society.

Dr. Driscoll works as associate director of the Northeast Regional Tuberculosis DNA Fingerprinting Laboratory at the Wadsworth Center of the New York State Department of Health. His research interests include the evolution and molecular epidemiology of the *Mycobacterium tuberculosis* complex.

References

1. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, et al. *Mycobacterium tuberculosis* isolates from a state with a large rural population. J Infect Dis 1997;175:1446–52.
2. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, et al. Comparison of methods based on different molecular epidemiologic markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
3. Rhee JT, Tanaka MM, Behr MA, Agasino CB, Paz EA, Hopewell PC, et al. Use of multiple markers in population-based molecular epidemiologic studies of tuberculosis. Int J Tuberc Lung Dis 2000;4:1111–9.
4. Salamon H, Behr M, Rhee JT, Small PM. Genetic distances for the study of infectious disease epidemiology. Am J Epidemiol 2000;151:324–34.
5. Sola C, Filliol I, Gutierrez C, Mokrousov I, Vincent V, Rastogi N. Spoligotype database of *Mycobacterium tuberculosis:* biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg Infect Dis 2001;7:390–6.
6. Sola C, Devallois A, Horgen L, Maisetti J, Filliol I, Legrand E, et al. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. Emerg Infect Dis 1999;5:404–14.
7. van Embden JD, van Gorkom T, Kremer K, Jansen R, van der Zeijst BA, Schouls LM. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J Bacteriol 2000;182:2393–401.
8. Niang MN, Niang N, de la Salmoniere YG, Samb A, Hane AA, Cisse MF, et al. Characterization of *M. tuberculosis* strains from West African patients by spoligotyping. Microbes Infect 1999;1:1189–92.
9. Viana-Niero C, Gutierrez C, Sola C, Filliol I, Boulahbal F, Vincent V, et al. Genetic diversity of *Mycobacterium africanum* clinical isolates based on IS*6110*-restrction fragment length polymorphism analysis, spoligotyping, and variable number of tandem DNA repeats. J Clin Microbiol 2001;39:57–65.
10. Niemann S, Richter E, Rusch-Gerdes S. Differentiation among members of the *Mycobacterium tuberculosis* complex by molecular and biochemical features: evidence for two pyrazinamide-susceptible subtypes of *M. bovis*. J Clin Microbiol 2000;38:152–7.
11. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol 2002;10:45–52.
12. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 1995;33:3234–8.
13. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
14. Tanaka MM, Small PM, Salamon H, Feldman MW. The dynamics of repeated elements: applications to the epidemiology of tuberculosis. Proc Natl Acad Sci U S A 2000;97:3532–7.
15. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 1990;18:6097–100.
16. Dale J, Brittain D, Cataldi A, Cousins D, Crawford J, Driscoll J, et al. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardized nomenclature. Int J Tuberc Lung Dis 2001;5:216–9.
17. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
18. Henikoff S, Henikoff JG, Alford WJ, Pietrokovski S. Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene 1995;163:GC17–GC26.
19. Schneider TD. Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. Nucleic Acids Res 2001;29:4881–91.
20. Filliol I, Sola C, Rastogi N. Detection of a previously unamplified spacer within the DR locus of *Mycobacterium tuberculosis:* epidemiological implications. J Clin Microbiol 2000;38:1231–4.
21. Centers for Disease Control and Prevention. Recommendations for prevention and control of tuberculosis among foreign-born persons. Report of the working group on tuberculosis among foreign-born persons. MMWR Morb Mortal Wkly Rep 1998;47(RR-16):1–29.
22. Borgdorff MW, Behr MA, Nagelkerke NJ, Hopewell PC, Small PM. Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. Int J Tuberc Lung Dis 2000;4:285–6.
23. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J Mol Evol 2001;53:680–9.
24. Supply P, Mazars E, Lesjeans S, Vincent V, Gicquel B, Locht C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol 2000;36:762–71.
25. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–90.
26. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 2001;39:3563–71.

Address for correspondence: Jeffrey R. Driscoll, Division of Infectious Diseases, Wadsworth Center, New York State Department of Health, P.O. Box 22002, New Scotland Ave., Albany, NY 12201-2002, USA; fax: 518-473-1326; e-mail: jeffrey.driscoll@wadsworth.org

Molecular Differentiation of *Mycobacterium tuberculosis* Strains without IS*6110* Insertions

Kerry H. Lok,* William H. Benjamin, Jr.,* Michael E. Kimerling,* Virginia Pruitt,† Monica Lathan,‡ Jafar Razeq,‡ Nancy Hooper,‡ Wendy Cronin,‡ and Nancy E. Dunlap*

By using standard restriction fragment length polymorphism, 6 zero-copy IS*6110 Mycobacterium tuberculosis* isolates were identified from 1,180 Maryland isolates as part of the National Tuberculosis Genotyping and Surveillance Network Project. By using various genotyping methods, we demonstrated that this zero band cluster can be differentiated into six genotypes.

IS*6110* restriction fragment length polymorphism (RFLP) has been used to investigate *Mycobacterium tuberculosis* transmission within populations since the early 1990s (1–4). However, some strains do not have any IS*6110* insertions and thus are zero-band strains (5) and are considered a single IS*6110* cluster by exclusion (5,6). These zero-band strains are rare in the United States and usually are found in persons from Southeast Asia. Recently, the spacer oligonucleotide typing (spoligotyping) technique was used to divide some zero-copy IS*6110* strains into subclusters that are associated with distinct geographic origins. As part of the National Tuberculosis Genotyping and Surveillance Network, six zero-copy IS*6110* strains were identified in persons from Maryland during a 5-year period. We conducted further molecular characterization of these strains to ascertain if they were closely related.

Six zero-copy IS*6110* isolates were identified from 1996 to 2000. These isolates, negative for IS*6110* by RFLP analysis, were from 1,180 Maryland cases. The isolates were collected as part of the genotyping network. Of the six patients, three were from Vietnam (two shared a common surname), and the others came from India, Iraq, and Liberia. None were linked epidemiologically to one another.

We identified all isolates as *M. tuberculosis* complex with BACTEC NAP test (BACTEC 460, BD Diagnostic Systems, Sparks, MD) and AccuProbe (Gen-Probe, San Diego, CA) before they were sent to the Alabama Regional Genotyping Laboratory. The isolates were fingerprinted at least twice by IS*6110* RFLP to rule out technical error in the RFLP procedure. After error was ruled out, all zero-copy strains were then tested with three secondary typing methods.

M. tuberculosis isolates were cultured on Lowenstein-Jensen or 7H11 Middlebrook plates for at least 4 weeks before DNA extraction. Chromosomal DNA was extracted from the isolates with chloroform-isoamyl alcohol, and RFLP was performed according to international standards (7). For the zero-copy IS*6110* strains, membranes with negative lanes were subjected to at least twice the normal exposure time to rule out the possibility of missing a faint band.

The use of spoligotyping was based on the presence or absence of 43 variable spacers in the direct repeat (DR) region of *M. tuberculosis*. Spoligotyping membranes were purchased from Isogen Bioscience BV (Isogen, Bilthoven, the Netherlands). We followed the manufacturer's recommendations for hybridizing polymerase chain reaction (PCR) products, as described by Groenen and colleagues (8). The numbering of the spacer regions was done as reported previously (8). Excel (Microsoft Corp., Redmond, WA) was used to analyze the spoligotyping results. For national database reporting, we converted the spoligotyping image into an octal-digital format based on the protocol set by colleagues and the genotyping network (9).

After *Alu*I digestion, the DNA was transferred to a nylon membrane for polymorphic guanine cytosine-rich repetitive sequence (PGRS) Southern blotting. Southern-blotting protocol was followed at the genotyping network except for the use of the plasmid pTBN12 as the probe. The results were compared visually (10).

The variable number of tandem repeats (VNTR) typing method was employed, as previously described, to further investigate these strains (11). The sizes of the fragments were determined by using a DNA ladder and amplicons from strain H37Rv (11).

Four different spoligotype patterns were obtained for the six isolates (Table). Three isolates had the same spoligotype patterns (designated Centers for Disease Control and Prevention [CDC] spoligotype 258): isolates 1–3 with an absence of spacers 19–41. The other three patterns were distinct with deletion of spacers as follows: isolate 4 (deletions at 4–11, 13–36, and 40); isolate 5 (deletions at 12–43); and isolate 6 (deletions at 33–34).

*University of Alabama at Birmingham, Birmingham, Alabama, USA; †Alabama Department of Public Health, Montgomery, Alabama, USA; and ‡Maryland Department of Health and Mental Hygiene, Baltimore, Maryland, USA

Table. Zero-copy IS*6110 Mycobacterium tuberculosis* strains from previously reported studies, other studies from the National Tuberculosis Genotyping Surveillance Network, and strains from current study[a]

No./case	Spoligotype octal description	CDC designation	Country of origin	Yr reported	Source of data
1	N/A	N/A	India	1993	(10)
3	N/A	N/A	Hong Kong	1995	(5)
5	N/A	N/A	India	1995	(5)
3	577767000000011	N/A	Asia	1997	(12)
1	777777777760771	0002	Curacao	1999	(13)
1	477776501013071	N/A	India	1999	(13)
1	777701002001731	N/A	the Netherlands	1999	(13)
1	N/A	N/A	United States	Unpublished	Arkansas
1	777777777760700	0202	China	Unpublished	California
8	777777000000011	0258	Vietnam	Unpublished	California
1	777603000000011	1682	Vietnam	Unpublished	California
1	777777600007771	1128	North Korea	Unpublished	California
1	777777000000011	0258	Vietnam	Unpublished	Michigan
1	777603002000011	0870	Vietnam	Unpublished	Texas
1	N/A	N/A	Vietnam	Unpublished	Texas
1	000000000003771	0034	N/A	2000	(14)
2	777777000000011	0258	Vietnam	2001	(15)
1	777603000000011	1682	Vietnam	2001	(15)
1	741777000000011	n/a	Vietnam	2001	(15)
2	777647000000011	n/a	United States	2001	(15)
3	777777000000011	0258	Vietnam	2002	Study isolates 1–3
1	700100000000731	0968	Iraq	2002	Study isolates 4
1	577600000000000	0742	India	2002	Study isolates 5
1	777777777763771	0169	Liberia	2002	Study isolates 6
43					

[a]CDC, Centers for Disease Control and Prevention; N/A, not applicable.

PGRS also yielded four patterns and divided the strains into groups that were identical to those found with spoligotyping (Figure 1). Three isolates (lanes 1–3) had the same banding pattern, and one isolate (lane 6) had a similar pattern that differed only by the presence of a double rather than single band at 2,760 bp. The two remaining isolates had distinct patterns (lanes 4 and 5). The lack of variability in the PGRS patterns suggests that these strains may be related. The genomes of these strains may be more stable than strains with IS*6110*.

We used VNTR to genotype the three samples (lanes 1–3) (Figure 1) that had identical spoligotype and PGRS results. The profile of these three isolates, which were obtained from three Vietnamese patients, were differentiated only by loci exact tandem repeats (ETR)-A and ETR-D (Figure 2). The remaining three isolates (not shown) had multiple differences.

The first zero-copy IS*6110 M. tuberculosis* strain was reported by van Soolingen and colleagues in 1993 (10). Subsequently, 22 cases have been reported (5,6,12,13,14). Within the genotyping network, another 21 cases were documented between 1996 and 2000. California reported 11 cases, Maryland 6 (this study), Texas 2, Michigan 1, and Arkansas 1. Zero-copy isolates were not found in New Jersey and Massachusetts. These isolates represent only 21 (0.18%) of 11,923 fingerprinted from seven surveillance sites. Therefore, zero-copy IS*6110* strains are extremely rare in the United States. Goguet de la Salmoniere and colleagues reported three zero-copy IS*6110* from 106 cases (2.8%) in three French hospitals during a 1-year study (12). This rate is 10 times higher compared with the genotyping network findings in the United States. The combined data of spoligotype profiles and the patients' countries of origin suggest that most isolates originated in Asia (17). Without additional epidemiologic data, this hypothesis cannot be corroborated.

Using the secondary molecular genotyping techniques, we showed that all six isolates were different. We found that all strains but one had deletions in the DR region that included DR 24 (a common and perhaps original insertion site for IS*6110)*. Only the strain from Liberia has spacer 24 in the DR region. This strain represented an exception in the zero-copy IS*6110* strains. However, the absence of spacer 24 is not an

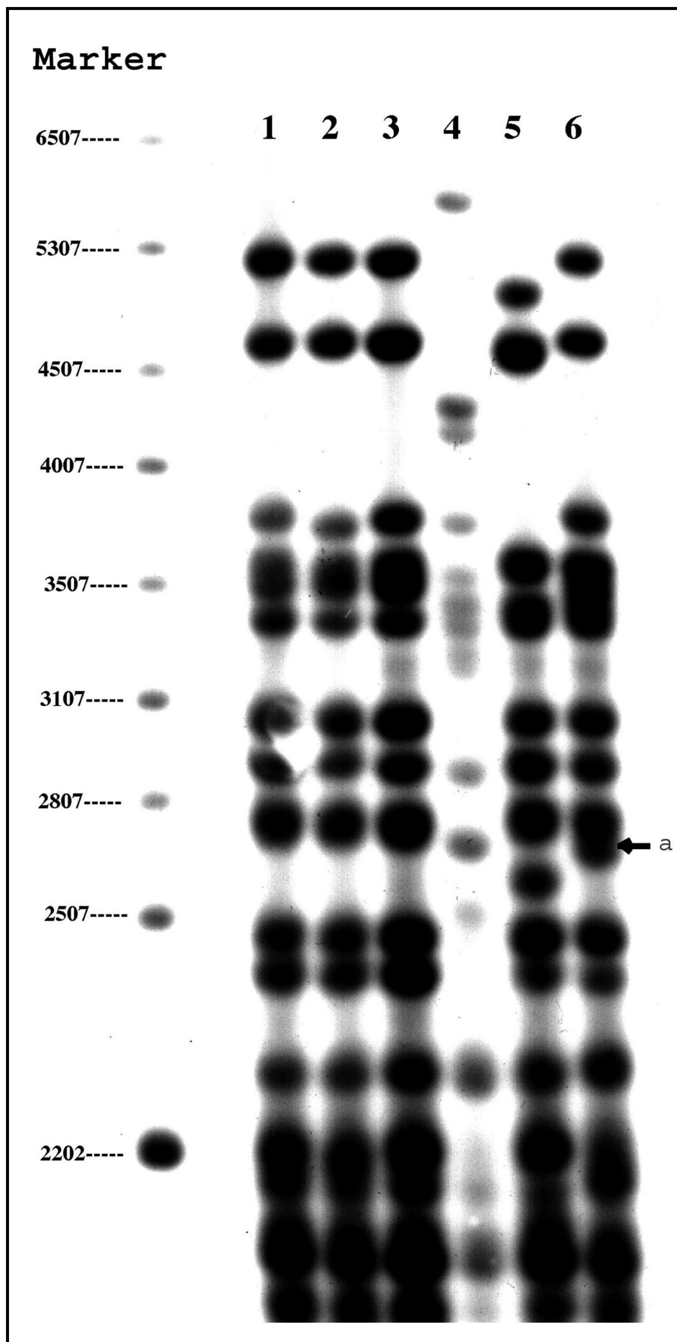

Figure 1. Polymorphic guanine cytosine-rich repetitive sequence restriction fragment length polymorphism results of six zero-copy IS*6110* strains. Lanes 1–6 represent the six cases reported in this study. The arrow indicates an additional band at 2,760 bp in isolate 6 compared to lanes 1–3.

absolute indication of a zero-copy IS*6110* strain. All Beijing strain families have a deletion that includes spacers 1–34, but they also have multiple IS*6110* insertions (15 to 21) and some in the remaining DR region (18).

Among the six zero-copy IS*6110* strains, the three Vietnamese isolates had the same spoligotype and PGRS. By using VNTR genotyping we were able to differentiate these strains; two VNTR loci differentiated them. The importance of this finding is unknown.

Figure 2. The variable number of tandem repeats results of three zero-copy IS*6110* isolates with identical spoligotyping and polymorphic guanine cytosine-rich repetitive patterns. Seven VNTR loci are listed across the top. The second line represents the patient number of the isolates for each locus.

In comparing the six zero-copy IS*6110* cases reported in this study with more recent data from the genotyping network and elsewhere, we identified 43 zero-copy IS*6110* strains (Table). Most of these strains (35 [81%] of 43) originated from southern Asia. Fourteen of 32 spoligotyped strains are similar and have the same CDC designation. Additionally, eight of the strains are similar to CDC spoligotype 258, distinguished by the same absence of spacers 19–41. Most strains spoligotyped (29 [91%] of 32) have a deletion of the spacer 24 region. The other three isolates did have spacer 24, which cannot be explained without sequencing the DR region.

By using multiple genotyping methods, we confirmed the identification of zero-copy *M. tuberculosis* isolates. We further demonstrated that the six unrelated cases were caused by different *M. tuberculosis* strains. However, the close similarity of the Vietnamese genotypes implies an important geographic association. An optimal algorithm for evaluating zero band isolates is yet to be determined and should be based on evolving secondary methods.

Acknowledgments

We thank Alabama State Tuberculosis Laboratory, especially Donna Mulcahy and Nancy Robinson for preparing the strains for the study, and William Bishai for providing pTBN12.

Dr. Lok is a postdoctoral fellow in pulmonary diseases at the University of Alabama at Birmingham. His research interests include the molecular epidemiology of tuberculosis.

References

1. Mazurek GH, Cave MD, Eisenach KD, Wallace RJJ, Bates JH, Crawford JT. Chromosomal DNA fingerprint patterns produced with IS*6110* as strain-specific markers for epidemiologic study of tuberculosis. J Clin Microbiol 1991;29:2030–3.
2. Dunlap NE. The use of RFLP as a tool for tuberculosis control: utility or futility? Int J Tuberc Lung Dis 2000;4:S134–8.
3. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N J Med 1994; 330:1703–9.
4. Jasmer R, Hahn J, Small P, Daley C, Behr M, Moss A, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 1999;130:971–8.

5. Das S, Paramasivan CN, Lowrie DB, Prabhakar R, Narayanan PR. IS*6110* restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, south India. Tuber Lung Dis 1995;76:550–4.
6. Das S, Chan SL, Allen BW, Mitchison DA, Lowrie DB. Application of DNA fingerprinting with IS986 to sequential mycobacterial isolates obtained from pulmonary tuberculosis patients in Hong Kong before, during and after short-course chemotherapy. Tuber Lung Dis 1993;74:47–51.
7. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
8. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol Microbiol 1993;10:1057–65.
9. Dale JW, Brittain D, Cataldi AA, Cousins D, Crawford JT, Driscoll J, et al. Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: recommendations for standardised nomenclature. Int J Tuberc Lung Dis 2001;5:216–9.
10. van Soolingen D, de Haas P, Hermans P, Groenen P, van Embden JD. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 1993;31:1987–95.
11. Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–96.
12. Goguet de la Salmoniere YO, Li HM, Torrea G, Bunschoten A, van Embden J, Gicquel B. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J Clin Microbiol 1997;35:2210–4.
13. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
14. Soini H, Pan X, Amin A, Graviss EA, Siddiqui A, Musser JM. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J Clin Microbiol 2000;38:669–76.
15. Soini H, Musser JM. Molecular diagnosis of mycobacteria. Clin Chem 2001;47:809–14.
16. Soini H, Pan X, Teeter L, Musser JM, Graviss EA. Transmission dynamics and molecular characterization of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J Clin Microbiol 2001;39:217–21.
17. Sola C, Horgen L, Goh KS, Rastogi N. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean island with IS6110 and DRr probes. J Clin Microbiol 1997;35:843–6.
18. Beggs ML, Eisenach KD, Cave MD. Mapping of IS*6110* insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J Clin Microbiol 2000;38:2923–8.

Address for correspondence: Michael E. Kimerling, University of Alabama at Birmingham, RPHB 217, 1530 3rd Ave. South, Birmingham, Alabama 35294-0022, USA; fax: 205-934-1746; e-mail: kimerlin@uab.edu

EMERGING INFECTIOUS DISEASES

Full text free online at www.cdc.gov/eid

The print journal is available at **no charge** to public health professionals

☐ YES, I would like to receive Emerging Infectious Diseases.

Please print your name and business address in the box and return by fax to 404-371-5449 or mail to
EID Editor
CDC/NCID/MS D61
1600 Clifton Road, NE
Atlanta, GA 30333

Moving? Please give us your new address (in the box) and print the number of your old mailing label here______________

EID Online
www.cdc.gov/eid

DNA Fingerprinting of *Mycobacterium tuberculosis*: Lessons Learned and Implications for the Future

Scott J. N. McNabb,* Christopher R. Braden,* and Thomas R. Navin*

DNA fingerprinting of *Mycobacterium tuberculosis*—a relatively new laboratory technique—offers promise as a powerful aid in the prevention and control of tuberculosis (TB). Established in 1996 by the Centers for Disease Control and Prevention (CDC), the National Tuberculosis Genotyping and Surveillance Network was a 5-year prospective, population-based study of DNA fingerprinting conducted from 1996 to 2000. The data from this study suggest multiple molecular epidemiologic and program management uses for DNA fingerprinting in TB public health practice. From these data, we also gain a clearer understanding of the overall diversity of *M. tuberculosis* strains as well as the presence of endemic strains in the United States. We summarize the key findings and the impact that DNA fingerprinting may have on future approaches to TB control. Although challenges and limitations to the use of DNA fingerprinting exist, the widespread implementation of the technique into routine TB prevention and control practices appears scientifically justified.

The capacity to differentiate *Mycobacterium tuberculosis* strain patterns by DNA fingerprinting has shown promise in tuberculosis (TB) control since this tool was first applied to outbreak investigations (1–3) and population-based studies (4,5) in the early 1990s. Evaluating this tool and determining its limitations are important activities in view of the most recent efforts to eliminate TB in the United States after its resurgence in the early 1990s (6). In 1996, the Centers for Disease Control and Prevention (CDC) initiated a 5-year, prospective, population-based study, the National Tuberculosis Genotyping and Surveillance Network. Most findings represented in this issue of Emerging Infectious Diseases come from this study. In this synopsis, we address two important implications for DNA fingerprinting of *M. tuberculosis*: its varied utility as a tool in TB prevention and control and its value in the measurement of the overall diversity of *M. tuberculosis* strain patterns in the United States, including differences by region and population and the prevalence of endemic strains.

Identification of Laboratory Cross-Contamination or Mislabeling

DNA fingerprinting of *M. tuberculosis* has been shown to identify and confirm laboratory cross-contamination or mislabeling. Previous retrospective studies describing *M. tuberculosis* laboratory cross-contamination or mislabeling found rates between 0.9% and 3.5% (7–14). In this issue, Northrup et al. (15) report a rate of 1.5%, which is within the range of published rates. Therefore, of 13,035 culture-positive TB case-patients reported in the United States in 2000, TB may have been misdiagnosed in as many as 117 (0.9%) to 456 (3.5%) persons. Using the previously reported finding that two thirds of case-patients with false-positive cultures are treated (11), we estimate that 78–304 persons may have been misdiagnosed and treated unnecessarily in the United States in 2000.

To measure the direct and indirect financial costs associated with these laboratory cross-contamination or mislabeling events, Northrup et al. (15) discuss data about three persons who were falsely diagnosed as having TB. In 1999 U.S. dollars, the estimated average cost to the health-care system for each person was $32,231 and the estimated average direct cost for each person was $10,744; thus, the estimated average cost for misdiagnosis of TB was $42,975 per person. This study did not include indirect and intangible costs, which can be substantial and are largely paid by the patient. Therefore, these costs are underestimates of the true costs associated with these events. Extrapolating these cost estimates to the national level and assuming that 78–304 persons were misdiagnosed and treated in 2000, we estimate that the preventable costs to the U.S. health-care system were $3.35 million to $13.06 million in 2000.

To examine the potential to predict laboratory cross-contamination or mislabeling, Jasmer et al. (16) established reproducible and predetermined criteria on the basis of DNA fingerprinting. They prospectively reviewed these events in three large, experienced laboratories in California. Laboratory procedures were reviewed at the start of the study; culture-positive results for 6 (2%) of 296 persons were caused by laboratory cross-contamination, which could be identified a priori. In this study, five of the six persons received unnecessary, expensive, and potentially dangerous medical treatment.

*Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Monitoring Interjurisdictional Transmission

The impact of epidemic spread of TB across state or jurisdictional boundaries and the necessity for interjurisdictional public health collaboration are not always fully accepted or appreciated, due in large part to the constitutionally mandated independence of state governments for public health practice. Many reports of localized-only epidemic transmission (17) suggest that most transmission of TB in the United States occurs locally. Indeed, Ellis et al. (18) show that most clusters (66% or 680/1,029) from the National Tuberculosis Genotyping and Surveillance Network were restricted to a single site.

However, a few recent reports illustrate the importance of the wider geographic spread of TB and the necessity for interjurisdictional collaboration (19,20). Ellis et al. also show that 260 (25%) clusters from the genotyping surveillance network were found in two sites, 55 (5%) in three, 19 (2%) in four, 8 (1%) in five, and 7 (1%) in six sites. As expected, clusters that spanned multiple sites included more case-patients. Maximum cluster size and absolute numbers of case-patients with isolates that clustered continued to increase through the end of the study. Though many of the case-patients in clusters that spanned these multiple sites may not be epidemiologically linked, other examples of interjurisdictional transmission exist (19–22).

We know that TB is problematic in certain groups at high risk (i.e., homeless persons) (23–29). With this in mind, an exception to the concept of local-only spread is described by Lathan et al. (30). They present data showing that interjurisdictional collaboration in TB control was necessary to control epidemic spread between adjacent jurisdictions (i.e., Maryland with Washington D.C.). Through combined DNA fingerprinting cluster analyses, these researchers found additional and unsuspected TB transmission that not only crossed state lines but also crossed social lines (i.e., between homeless and nonhomeless persons).

McElroy et al. (31) show the value of a system to compare DNA fingerprinting data across jurisdictions, especially during multistate TB outbreak settings. Facilitating an interjuridictional investigation of TB first recognized in Maryland, they extended the investigation outside of Maryland and discovered an additional 18 case-patients linked to the original 21 previously described.

TB control programs that employ DNA fingerprinting should work with neighboring jurisdictions to devise strategies that promote rapid sharing of results. Comprehensive interjurisdictional monitoring of transmission would require a national registry of DNA fingerprinting data and a system to alert public health officials about interjurisdictional clustering.

Program Evaluation

Understanding the transmission characteristics of multidrug-resistant TB strains is essential for developing successful control strategies. Multidrug-resistant TB is an important problem and represents a life-threatening condition; patients often require prolonged treatment and frequent hospitalization. The potentially serious side effects of second-line TB drugs challenge effective treatment and the resources of the patient and the health-care system (32).

One indicator for the success of a TB control program's performance is a decline in TB incidence rate over time. Incident TB case-patients include those recently infected who progress rapidly to active TB and those with remote infection who progress after years of latency and are later diagnosed with active TB. However, in high transmission circumstances, recent disease transmission likely accounts for most incident TB case-patients. DNA fingerprinting clustering may reflect recent TB transmission (4,5,33,34), albeit with methodological and population characteristic caveats (35–37). Therefore, the decline in the incidence of clustering can indicate the impact of interventions aimed at reducing recent TB transmission (38).

Munsiff et al. (39) describe how DNA fingerprinting, specifically molecular clustering of multidrug-resistant TB strains as a surrogate of recent transmission, was used to monitor an improved TB program in New York City. During a 3-year period (1995–1997), multidrug-resistant TB was diagnosed for 241 case-patients (4.9% of all culture-positive, TB case-patients) in New York City. These 241 case-patients were more likely than culture-positive, non–multidrug-resistant TB case-patients to have been born in the United States, to be infected with HIV, to be health-care workers, and to have positive acid-fast bacilli smear results. During this study period, of 234 multidrug-resistant TB case-patients with DNA fingerprinting results, 153 (65%) were grouped into 19 clusters. Epidemiologic links were identified for 25 (19.8%) case-patients clustered by DNA fingerprinting.

Kong et al. from Denver (40) demonstrate the role of DNA fingerprinting to measure the performance of a tuberculin skin-test program among homeless persons. They showed a decrease in clustering (a surrogate of recent TB transmission) from 49% during the 7-year period before the program was implemented to 14% in the 4-year period after the program. This assessment is a logical extension of the usefulness of DNA fingerprinting technology, since a previous report from Denver by Burman et al. (41) showed that homelessness was a predictor of DNA fingerprint clustering.

Ellis et al. (18) show that, despite a decrease in TB incidence rates at all genotyping surveillance network sites, the proportion of cases in clusters stabilized at a relatively high level (~48%). They suggest that this high proportion of clustering may be due to the inclusion of many low-incidence, stable populations in which persons in chains of transmission from past decades still reside in proximity (42). Alternative explanations include a slightly younger population under study, the presence of old endemic strains that have spread widely, and the limited discrimination of low-copy number IS*6110* restriction fragment length polymorphism (RFLP) patterns, even with the addition of spoligotyping as a secondary test. Although the overall proportion of case-patients in clusters plateaued at approximately 48% over the 5 years of study, the annualized proportion of case-patients in clusters decreased over time with a concomitant decrease in TB incidence. This finding, which

might reflect more effective TB control in limiting ongoing transmission, is provocative and merits further investigation.

Statewide Assessment of Circumstances and Settings for TB Transmission

The statewide use of DNA fingerprinting provides an informed picture of the epidemiologic features of disease transmission (36–38). Cronin et al. (43) describe a specific example in Maryland. During a 5-year period, they used DNA fingerprinting on >99% of all isolates in the state and found that cluster investigations were very effective in identifying additional epidemiologic links, many of which occurred in nontraditional settings. Specifically, isolates from 436 (37%) of 1,172 Maryland case-patients were clustered by DNA fingerprinting. Of those 436 clustered case-patients, these researchers found 155 (36%) to be epidemiologically linked using traditional contact investigation.

Miller et al. (44) provide another example of statewide use of DNA fingerprinting by exploring the impact of DNA fingerprinting used in Massachusetts during investigations of a TB outbreak and a laboratory cross-contamination event. These researchers also describe how DNA fingerprinting affected the identification of *M. tuberculosis* strains and transmission sites and accurate epidemiologic links. Overall, they found that, in addition to 129 epidemiologic relationships found before DNA fingerprinting results were obtained, 12 other epidemiologic relationships involving 20 persons were discovered as a result of cluster investigations. In addition, they determined places of transmission previously unrecognized and used DNA fingerprinting to refute a purported TB outbreak.

Sharnprapai et al. (45) also report on the use of DNA fingerprinting to further understand the epidemiology and transmission patterns of TB in Massachusetts. In this study, 28% of TB case-patients were clustered, and case-patients born in the United States were two times more likely to cluster than case-patients not born in the United States. Furthermore, they point out a very important limitation to interpreting and using DNA fingerprinting data. Despite using a secondary typing method (spoligotyping) with strains that have six or fewer copies of IS*6110*, a limited ability to differentiate these strains exists. In their study, clusters of strains with more than six copies of IS*6110* were more likely to have epidemiologic links found than clusters of strains with six or fewer copies of IS*6110*.

The report by Dillaha et al. (46) describes the subtleties of disease transmission faced by TB programs in low-incidence states (specifically Arkansas), which are in the forefront as the United States moves toward TB elimination. Thirty-five case-patients in a 54-year period with identical or very similar fingerprints were identified. After reviewing the endemic strain, these researchers recognized the lack of success with traditional contact tracing and treatment recommendations for latent TB infection for persons with positive tuberculin skin tests. This critical determination has implications for other low-incidence areas. In addition to the traditional focus on personal contacts, Dillaha et al. recommend case finding and screening on the basis of geographic location. With the advent of geographic interface technology and mapping, this new public health strategy might be feasible. Additionally, these findings support the usefulness of a social network approach to contact investigation.

Outbreak Investigation

The value of DNA fingerprinting has been shown clearly during outbreak investigations (1–3). During every investigation, one overriding question recurs: which case-patients related to the outbreak are part of the chain of transmission and are not unrelated, sporadic cases? DNA fingerprinting can help researchers determine whether patients are related to the outbreak and thus focus the epidemiologic investigation.

An intriguing new benefit of coupling DNA fingerprint information with outbreak investigation lies in the power of this tool to increase understanding of the often difficult to discern transmission patterns of community TB disease (47) and uncover previously unknown outbreaks. Ijaz et al. (48) show this potential use by demonstrating that molecular clusters could show previously unsuspected instances of probable TB transmission, prompting more directed investigations to seek epidemiologic links missed by routine contact investigation. In the study, cluster analysis was based on identical and similar DNA fingerprinting patterns, broadening the group of patients included in the initial analysis.

In their study, secondary typing was accomplished by using a polymorphic GC-rich sequence for identical IS*6110*-based DNA fingerprinting patterns with six or fewer bands or for patterns with more than six bands that were similar but differed by a single band (49). In further investigations of clusters with this “broader net,” among 66% of case-patients, Ijaz et al. uncovered additional epidemiologic links missed during routine contact investigations in Arkansas. During this process, they found an extensive, previously unknown social network that aided public health investigations. Ijaz et al. conclude that patients whose isolates have similar but not identical IS*6110* patterns should be considered potential members of a cluster and be included during epidemiologic investigations.

Oh et al. (50) help establish the value of DNA fingerprinting in an unusual outbreak setting, a zoo. They describe a multispecies epizootic with genotypically identical *M. tuberculosis* strains. Their DNA fingerprint investigation showed that five of six animals had the identical strain and that zoo employees with previous negative tuberculin skin tests were exposed. Skin tests for 55 (18%) of 307 employees were positive, showing evidence of recent infection.

Bennett et al. (51) present data from the genotyping surveillance network with important policy implications. In addition to indicating that contact investigation should be extended to all settings frequented by the source case-patient, they also showed a significant positive association between being a smear-negative source case-patient and having unconfirmed transmission. This finding suggests that the identification of a

smear-negative source case-patient (as an index case-patient) should not preclude the ongoing investigation for other possible sources. They also suggest that transmission from smear-negative case-patients is not negligible.

Sun et al. (52) report data gathered on transmission of tuberculosis to children <5 years of age. Representing a sentinel health event, thorough investigation of the circumstances of childhood tuberculosis remains critical to effective public health practice. They found that routine public health investigations conducted by local health departments, within the National Tuberculosis Genotyping and Surveillance Network, identified suspected source patients for 57 (51%) of 111 culture-confirmed case-patients <5 years of age. For 8 (15%) of these 57 patients, DNA fingerprinting suggested infection with different strains. These children were more likely to be older than other children and source case-patients with identical strains. The findings in this study highlight the requirement of rigorous case and contact investigation efforts, especially in household settings.

DNA Fingerprinting Laboratory Techniques

Other articles in this issue present data that improve our understanding of both laboratory facets of DNA fingerprinting (i.e., IS*6110* fingerprinting) and basic science of *M. tuberculosis.* Braden et al. (53) report the results of an external quality assessment program for the seven network laboratories in which the interlaboratory reproducibility was measured. They found that, overall, an exact match was achieved for 73% of isolates in panels: 90% matched with a one-band difference and 97% matched with a two-band difference. Although they report that final outcomes of pattern analysis and cluster determination in the genotyping surveillance network were probably closer to reality than the results of this quality assurance exercise suggest, they also warn that the variability and nonreproducibility are substantial and should be considered when interpreting the results from the genotyping surveillance network. Crawford et al. (54) demonstrate through the establishment of the genotyping surveillance laboratory, that DNA fingerprinting remains an “art,” and the experience and training of laboratorians are important.

Driscoll et al. (55) describe an evaluation of “logo analyses.” Array-based assays use reverse hybridization. The binary nature of array-based assays allows data to be analyzed usefully with algorithms associated with motif recognition, such as sequence logo analyses. Logo analyses have the potential to aid in visualizing and displaying spoligotyping cluster data and in managing the enormous amount of digital data generated by large-scale DNA fingerprinting projects. This potential is especially relevant now because low-incidence states and countries (including the United States) are considering universal implementation of DNA fingerprinting of *M. tuberculosis*. Using these and other bioinformatic tools, scientists will be able to interpret and understand the data generated by such a project.

Lok et al. (56) demonstrate that secondary typing methods (e.g., spoligotyping) should be used when isolates have no IS*6100* insertions (i.e., zero-band strains). This article describes the differentiation power of secondary typing in these instances. In a second paper, Lok et al. (57) demonstrate the use and power of polymerase chain reaction techniques (e.g., variable number of tandem repeats) to distinguish and characterize the most common *M. tuberculosis* strain pattern in the United States—a two-band IS*6110* RFLP pattern representing 5% of all isolates in the National Tuberculosis Genotyping Surveillance Network.

Diversity of *M. tuberculosis* Strains

The second important implication for DNA fingerprinting of *M. tuberculosis* is its ability to measure the overall diversity of *M. tuberculosis* strain patterns, including differences within the United States, differences by region and population, and prevalence of endemic strains. The genotyping surveillance network database demonstrated this diversity in the United States (57). The 10,883 patients in the study represent approximately 11.6% of all new TB cases in the United States from 1996 through 2000. Through this study, DNA fingerprinting of 10,883 isolates was performed by using the IS*6110* RFLP method, yielding 6,128 distinct patterns.

Cowan et al. (58) report that family analysis of IS*6110* patterns revealed 497 patterns related to the W-Beijing family (19); these patterns represent 946 isolates or 9% of all isolates in the genotyping surveillance network. Six new families of related DNA fingerprint patterns were also proposed for isolates containing 6–15 copies of IS*6110*. These families contain up to 251 patterns and 414 isolates; together, they contain 21% of isolates in this copy-number range and may represent endemic strains distributed across the United States.

The 8,245 isolates with more than six copies of IS*6110* yielded 5,640 fingerprint patterns. Of these, 4,846 (86%) were identified for a single isolate, and 794 patterns grouped 3,399 isolates into fingerprint-defined clusters. Of 457 fingerprint patterns identified among the 2,507 isolates with low-copy numbers (six or fewer copies of IS*6110*), 314 (69%) were reported for a single isolate, and 143 grouped 2,193 isolates into clusters. Clustering was much greater among isolates with low-copy numbers (87%) than among isolates with high-copy numbers (41%).

Limitations, Challenges, and Future Considerations

The “state of the art” in applied DNA fingerprinting technology has scientific and molecular limitations, as well as stumbling blocks to practical use in the field. The lack of reproducibility of the RFLP DNA fingerprinting technique and the difficulty in comparing patterns in an RFLP DNA fingerprint database remain important limitations in developing strategies for universal implementation. The other important factor related to RFLP is the time required for obtaining results. In order for DNA fingerprinting to provide value to

routine contact investigations, these molecular data must be available on a timely basis, so that public health intervention specialists can use them in cluster investigations.

Furthermore, the precision of the interpretation of these data is evolving. Clustering, by itself and in its entirety, is not always equivalent to recent transmission; however, a portion of it is. The limitation of clustering interpretation must be scientifically established, especially if that interpretation is used as a marker for public health practice performance and as an indication of progress toward elimination of TB. We know that the population under study and the length of observation time play important roles in interpreting this measure. In addition, the specificity of the clustering case-definition factors into this equation. Additional investigation into this area is necessary.

How well these research techniques can be expanded to assist TB control programs is not clear. Laboratory programs must be established to provide understandable, real-time results in a manner that influences decisions. The expansion of these techniques to assist TB control programs holds great promise. However, cluster investigation must be incorporated into routine public health practice, including the standardization of protocols. TB control officials require further training to interpret DNA fingerprinting results and act on these results in an effective way. The National Tuberculosis Controllers Association, in collaboration with the CDC, is currently drafting a DNA fingerprinting handbook to help health workers in the field understand and interpret DNA fingerprinting data.

We intend to continue analyses of the National Tuberculosis Genotyping and Surveillance Network data to gain additional insight into the value of cluster investigations. If DNA fingerprinting is to be implemented universally, this approach should be flexible enough to adapt to future laboratory techniques, as they become available. We think that DNA fingerprinting will become an essential tool in investigating TB transmission in difficult populations and unusual circumstances; consequently, DNA fingerprinting will be vital in the effort to eliminate TB.

Acknowledgments

The authors acknowledge the data management skill of Steve Kammerer and the valuable wisdom and counsel provided by Michael Iademarco, John Jereb, and Ken Castro.

This work was supported by the Centers for Disease Control and Prevention and the National Tuberculosis Genotyping and Surveillance Network.

Dr. McNabb is chief of the Epidemiologic Studies Section, Surveillance and Epidemiology Branch, Division of Tuberculosis Elimination, Centers for Disease Control and Prevention; Project Officer for the National Tuberculosis Genotyping and Surveillance Network; and adjunct associate professor at the Rollins School of Public Health, Emory University. His primary research interests are public health surveillance, molecular epidemiology, and the primary prevention of allergic diseases.

References

1. Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams EM, Sordillo KR, et al. An outbreak of multi-drug resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;326:1514–21.
2. Jereb JA, Berwin DR, Dooley SW, Haas,WH. Nosocomial outbreak of tuberculosis in a renal transplant unit: application of a new technique for restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates. J Infect Dis 1993;168:1219–24.
3. Daley CL, Small PM, Schecter GF, Schoolink GK, McAdam RA, Jacobs WR Jr, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus: an analysis using restriction-fragment-length polymorphisms. N Engl J Med 1992;326:231–5.
4. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, 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.
5. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiologic methods. N Engl J Med 1994;330:1710–6.
6. Centers for Disease Control and Prevention. Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Recomm Rep 2000;49:1–51.
7. Maurer JR, Desmond EP, Lesser MD, Jones WD Jr. False-positive cultures of *Mycobacterium tuberculosis*. Chest 1984;86:439–43.
8. Niven B, Kaye K, Munsiff SS. Detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures. Clin Infect Dis 1997;25:943.
9. Bhattacharya M, Dietrich S, Mosher L, Siddiqui F, Reisberg BE, Paul WS, et al. Cross-contamination of specimens with *Mycobacterium tuberculosis*: clinical significance, causes, and prevention. Am J Clin Pathol 1998;109:324–30.
10. Segal-Maurer S, Kreiswirth BN, Burns JM, Lavie S, Lim M, Urban C, et al. *Mycobacterium tuberculosis* specimen contamination revisited: the role of laboratory environmental control in a pseudo-outbreak. Infect Control Hosp Epidemiol 1998;19:101–5.
11. Burman WJ, Stone BL, Reves RR, Wilson ML, Yang Z, El-Hajj H, et al. The incidence of false-positive cultures for *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:321–6.
12. Van Duin JM, Pijnenburg JEM, van Rijswoud CM, de Haas PEW, Hendriks WDH, van Soolingen D. Investigation of cross contamination in a *Mycobacterium tuberculosis* laboratory using IS*6110* DNA fingerprinting. Int J Tuberc Lung Dis 1998;2:425–9.
13. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SM. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996;77:407–13.
14. Braden CR, Templeton GL, Stead WW, Bates JH, Cave D, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. Clin Infect Dis 1997;24:35–40.
15. Northrup JM, Miller AC, Nardell E, Sharnprapai S, Etkind S, Driscoll J, et al. Estimated costs of false laboratory diagnoses of tuberculosis in three patients. Emerg Infect Dis 2002;8;1264-70.
16. Jasmer RM, Roemer M, Hamilton J, Bunter J, Braden CR, Shinnick TM, et al. A prospective, multicenter study of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures. Emerg Infect Dis 2002;8;1260-3.
17. Bradford WZ, Koehler J, El-Hajj H, Hopewell PC, Reingold AL, Agasino CB, et al. Dissemination of *Mycobacterium tuberculosis* across the San Francisco Bay area. J Infect Dis 1998;177:1104–7.
18. Ellis BA, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S, et al. Molecular epidemiology of tuberculosis in a sentinel surveilance population. Emerg Infect Dis 2002;8;1197-1209.

19. Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, et al. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. JAMA 1996;275:452–7.
20. Casper C, Singh SP, Rane S, Daley CL, Schechter GS, Riley LW, et al. The transcontinental transmission of tuberculosis: a molecular epidemiological assessment. Am J Public Health 1996;86:551–3.
21. Agerton T, Valway SE, Blinkhorn RJ, Shilkret KL, Reves R, Schluter WW, et al. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis,* across the United States. Clin Infect Dis 1999;29:85–95.
22. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med 1998;338:633–9.
23. Barclay DM III, Richardson JP, Fredman L. Tuberculosis in the homeless. Arch Fam Med 1995;4:541–6.
24. Barnes PF, El-Hajj H, Preston-Martin S, Cave MD, Jones BE, Otaya M, et al. Transmission of tuberculosis among the urban homeless. JAMA 1996;275:305–7.
25. Concato J, Rom WN. Endemic tuberculosis among homeless men in New York City. Arch Intern Med 1994;154:2069–73.
26. Gelberg L, Panarites CJ, Morgenstern H, Leake B, Anderson RM, Koegel P. Tuberculosis skin testing among homeless adults. J Gen Intern Med 1997;12:25–33.
27. Layton MC, Cantwell MF, Dorsinville GJ, Valway SE, Onorato IM, Frieden TR. Tuberculosis screening among homeless persons with AIDS living in single-room-occupancy hotels. Am J Public Health 1995;85:1556–9.
28. Pilote L, Tulsky JP, Zolopa AR, Hanh JA, Schecter GF, Moss AR. Tuberculosis prophylaxis in the homeless. A trial to improve adherence to referral. Arch Intern Med 1996;156:161–5.
29. Zolopa AR, Hahn JA, Gorter R, et al. HIV and tuberculosis infection in San Francisco's homeless adults. JAMA 1994;272:455–61.
30. Lathan M, Mukasa LN, Hooper N, Golub J, Baruch N, Mulcahy D, et al. Cross-jurisdictional transmission of *Mycobacterium tuberculosis* in Maryland and Washington, D.C., 1996-2000, linked to the homeless. Emerg Infect Dis 2002;8;1249-51.
31. McElroy PD, Sterling TR, Driver CR, Kreiswirth B, Woodley CL, Cronin WA, et al. Use of DNA fingerprinting to investigate a multiyear, multistate tuberculosis outbreak. Emerg Infect Dis 2002;8;1252-6.
32. Centers for Disease Control and Prevention. Outbreak of multidrug-resistant tuberculosis—Texas, California, and Pennsylvania. MMWR Morb Mortal Wkly Rep 1990;39:369–72.
33. Jasmer RM, Daley CL, Hopewell PC. Trends in tuberculosis transmission (letters). Ann Intern Med 2000;132:416.
34. McConkey SJ, Williams M, Weiss D, Adams H, Cave MD, Yang Z, et al. Prospective use of molecular typing of *Mycobacterium tuberculosis* by use of restriction fragment length polymorphism in a public tuberculosis-control program. Clin Infect Dis 2002;34:612–9.
35. Glynn JR, Vynnycky E, Fine PEM. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am J Epidemiol 1999;149:366–71.
36. Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JDA, Fine PEM. The effect of age and study duration on the relationship between 'clustering' of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. Epidemiol Infect 2001;126:43–62.
37. Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. Am J Epidemiol 2002;155:565–71.
38. Geng E, Kreiswirth B, Driver C, Li J, Burzynski J, DellaLatta P, et al. Changes in the transmission of tuberculosis in New York City from 1990 to 1999. N Engl J Med 2002;346:1453–8.
39. Munsiff SS, Bassoff T, Nivin B, Li J, Sharma A, Bifani P, et al. Molecular epidemiology of multidrug-resistant tuberculosis, New York City, 1995-1997. Emerg Infect Dis 2002;8;1230-8.
40. Kong P-M, Tapy J, Calixto P, Burman WJ, Reves RR, Yang Z, et al. Skin-test screening and tuberculosis transmission among the homeless. Emerg Infect Dis 2002;8;1280-4.
41. Burman WJ, Reves RR, Hawkes AP, Reitmeijer CA, Yang Z, El-Hajj H, et al. DNA fingerprinting with two probes decreases clustering of *Mycobacterium tuberculosis*. Am J Respir Crit Care Med 1997;155:1140–6.
42. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro K, et al. Interpretation of DNA restriction fragment polymorphism analysis of *M. tuberculosis* isolates from a state with a large rural population. J Infect Dis 1997;175:1446–52.
43. Cronin WA, Golub JE, Lathan MJ, Mukasa LN, Hooper N, Razeq JH, et al. Molecular epidemiology of tuberculosis in a low- to moderate-incidence state: are contact investigations enough? Emerg Infect Dis 2002;8;1271-9.
44. Miller AC, Sharnprapai S, Suruki R, Corkren E, Nardell EA, Driscoll JR, et al. Impact of genotyping of *Mycobacterium tuberculosis* on public health practice in Massachusetts. Emerg Infect Dis 2002;8;1285-9.
45. Sharnprapai S, Miller AC, Suruki R, Corkren E, Etkind S, Driscoll J, et al. Genotyping analyses of tuberculosis cases in U.S.- and foreign-born Massachusetts residents. Emerg Infect Dis 2002;8;1239-45.
46. Dillaha JA, Yang Z, Ijaz K, Eisenach KD, Cave MD, Wilson FJ, et al. Transmission of *Mycobacterium tuberculosis* in a rural community, Arkansas, 1945-2000. Emerg Infect Dis 2002;8;1246-8.
47. Tabet SR, Goldbaum GM, Hooton TM, Eisenach KD, Cave MD, Nolan CM. Restriction fragment length polymorphism analysis detecting a community-based tuberculosis outbreak among persons infected with human immunodeficiency virus. J Infect Dis 1994;169:189–92.
48. Ijaz , Yang Z, Matthews HS, Bates JH, Cave MD. *Mycobacterium tuberculosis* transmission between cluster members with similar fingerprint patterns. Emerg Infect Dis 2002;8;1257-9.
49. Chaves F, Yang Z, El Hajj H, Alonso M, Burman W, Eisenach KD, et al. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J Clin Microbiol 1996;34:1118–23.
50. Oh P, Granich R, Scott J, Sun B, Joseph M, Stringfield C, et al. Human exposure following *Mycobacterium tuberculosis* infection of multiple animal species in a metropolitan zoo. Emerg Infect Dis 2002;8;1290-3.
51. Bennett DE, Onorato IM, Ellis BA, Crawford JT, Schable B, Byers R, et al. DNA fingerprinting of *Mycobacterium tuberculosis* isolates from epidemiologically linked case pairs. Emerg Infect Dis 2002;8;1224-9.
52. Sun SJ, Bennett DE, Flood J, Loeffler AM, Kammerer S, Ellis BA. Identifying the sources of tuberculosis in young children: a multistate investigation. Emerg Infect Dis 2002;8;1216-23.
53. Braden CR, Crawford JT, Schable BA. Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network. Emerg Infect Dis 2002;8;1210-5.
54. Crawford JT, Braden CR, Schable BA, Onorato IM. National Tuberculosis Genotyping and Surveillance network: design and methods. Emerg Infect Dis 2002;8;1192-6.
55. Driscoll JR, Bifani PJ, Mathema B, McGarry MA, Zickas GM, Kreiswirth BN, et al. Spoligologos: a bioinformatic approach to displaying and analyzing *Mycobacterium tuberculosis* data. Emerg Infect Dis 2002;8;1306-9.
56. Lok KH, Benjamin WH Jr, Kimerling ME, Pruitt V, Lathan M, Razeq J, et al. Molecular differentiation of *Mycobacterium tuberculosis* strains without IS*6110* insertions. Emerg Infect Dis 2002;8;1310-3.
57. Lok KH, Benjamin WH Jr, Kimerling ME, Pruitt V, Mulcahy D, Robinson N, et al. Molecular typing of *Mycobacterium tuberculosis* strains with a common two-band IS*6110* pattern. Emerg Infect Dis 2002;8;1303-5.
58. Cowan LS, Crawford JT. Genotype analysis of *Mycobacterium tuberculosis* isolates from a sentinel surveillance population. Emerg Infect Dis 2002;8;1294-1302.

Address for correspondence: Scott J. N. McNabb, Division of Tuberculosis Elimination National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop E10, Atlanta, GA 30333, USA; fax: 404-639-8959; e-mail: sym3@cdc.gov

Rifampin- and Multidrug-Resistant Tuberculosis in Russian Civilians and Prison Inmates: Dominance of the Beijing Strain Family

Francis Drobniewski,* Yanina Balabanova,* Michael Ruddy,* Laura Weldon,* Katya Jeltkova,† Timothy Brown,* Nadezdna Malomanova,‡ Elvira Elizarova,‡ Alexander Melentyey,§ Ebgeny Mutovkin,§ Svetlana Zhakharova,‡ and Ivan Fedorin‡

Consecutive patient cultures (140) of *Mycobacterium tuberculosis* were collected from five Russian civilian and prison tuberculosis laboratories and analyzed for rifampin (*rpoB*) and isoniazid resistance (*inhA, katG, ahp*C); transmission of Beijing family isolates; and the importance of prison and previous therapy in drug resistance. Rifampin, isoniazid, and multidrug resistance occurred in 58.2%, 51.6%, and 44.7% of cultures, respectively; 80% of prison cultures were rifampin resistant. Spoligotyping and variable number tandem repeat (VNTR) fingerprinting divided the isolates into 43 groups. Spoligotyping demonstrated that a high proportion (68.1%) of patients were infected with Beijing family strains and that most (69.1%) were rifampin resistant; the highest proportion (81.6%) occurred in prison. One VNTR subgroup (42435) comprised 68 (72.3%) of the Beijing isolates with a small number of IS*6110* types; 50 (73.5%) were rifampin resistant. Rifampin-resistant Beijing isolates are dominant within the patient population, especially among prisoners, and threaten treatment programs.

The true extent of drug-resistant tuberculosis (TB) globally is unknown. Few national surveys are representative and validated by external laboratories (1,2). Accurate drug-susceptibility testing is difficult to do well. The quality of results can be further compromised if bacterial cultures originate from a selected population from which key groups, such as prisoners, have been omitted. Only limited susceptibility-testing data are available in those countries with the highest rates of TB such as India, Pakistan, Indonesia, China, and Russia.

The spread of resistant isolates, especially multidrug-resistant strains that are resistant to at least isoniazid and rifampin, compromises both clinical efficacy and public health control measures. Patients with infectious multidrug-resistant cases continue to expectorate smear-positive sputum longer than patients with drug-sensitive cases, increasing the probability of spread of infection. The transmission of highly drug-resistant strains has been documented within health-care institutions and prisons (3–6), and the spread of one highly resistant strain, W, has been documented from New York City to other U.S. states and Puerto Rico (7).

In Russia, TB incidence rates declined from the 1950s to the 1990s; the lowest incidence and mortality rates were recorded in 1991 (34.0/100 000 and 8.1/100 000, respectively). By 1999, these rates had risen to 85.2/100 000 and 20.0/100 000, respectively (8–10). The average age of TB patients has declined, reflecting high levels of recent transmission. Data on drug resistance for the whole of the Russian Federation are scarce; information comparable with international data have been reported by the World Health Organization from only 2 of the 89 oblasts (regions). In Ivanavo and Tomsk Oblasts, the prevalence of multidrug-resistant TB was 9.0% and 6.5%, respectively, in 1998–1999 (1,2).

We initiated a pilot study in Samara, Russia, 1 of 89 oblasts to determine the following: the value of genotypic methods for identifying rifampin resistance and multidrug-resistant TB; the extent of rifampin and multiple-drug resistance within the civilian and prison systems; the extent to which this drug resistance was associated with dominant strains, such as those of the Beijing family, by using molecular DNA fingerprinting techniques; and whether drug resistance was associated with being a prisoner or with previous TB treatment.

DNA fingerprinting that uses techniques such as restriction fragment length polymorphism (RFLP) based on IS*6110* is the international standard for documenting transmission at the molecular level (11–14). Rapid techniques that use polymerase chain reaction (PCR) amplification such as spacer-oligonucleotide typing (spoligotyping) (13–15) are usually less discriminating but may provide greater information on evolutionary origins and can be used when cultures are contaminated or nonviable.

*Guy's, King's and St Thomas' Medical School, King's College Hospital, London, United Kingdom; †Central Tuberculosis Research Institute, Moscow, Russian Federation; ‡Samara Oblast Dispensary, Samara City, Russian Federation; and §Samara Prison TB Service, Samara City, Russian Federation

Spoligotyping is particularly valuable in defining strains belonging to the Beijing family as well as subspeciating strains of the *Mycobacterium tuberculosis* complex and does not require viable organisms. The Beijing family of isolates was first reported at a high rate in the Beijing area in 1995 (16) and less frequently in other parts of Asia such as Vietnam (17). These isolates may be associated with a higher occurrence of drug resistance; for instance, the highly drug-resistant W strain, first identified in New York, is a member of the Beijing family (7,18). These isolates have previously been frequently seen in convenience samples taken in the Russian Federation (18,19). In our study, isolates were collected from both the prison and civilian sectors at the same time in one region of the Russian Federation. These isolates were analyzed with molecular epidemiologic techniques (spoligotyping, variable number tandem repeat [VNTR], rapid epidemiological typing [RAPET], IS*6110*) to characterize the role of Beijing isolates and molecular genotypic methods to determine rifampin, isoniazid, and multidrug resistance.

Methods

Population

All cultures of *M. tuberculosis* growing from February 1 to March 31, 2001, and May 14 to August 30, 2001, from five principal civilian and prison laboratories were analyzed: the Oblast Regional TB Laboratory (oblast laboratory), the Samara City Dispensary Laboratory no 1 (city dispensary), Samara City Hospital no. 1 (city hospital), Novokuibyshevsk Novokuibyshevsk Town Laboratory (Novokuibyshevsk laboratory), and Prison Colony no. 19 (prison colony). A single culture from each patient was analyzed. These five laboratories cover 40% of the population of this region with TB and are representative of the TB caseload in Samara Oblast. The oblast laboratory acts as the oblast reference/referral center; it has direct responsibility for rural areas, data collection for the oblast, and methodologic and organizational issues. The city dispensary receives cultures for drug-susceptibility testing from all five laboratories covering the outpatient caseload from the capital, Samara City; city hospital is the main inpatient facility. Novokuibyshevsk laboratory is the central TB laboratory; it serves a medium-sized industrial town. All patients with TB in the prison sector receive their inpatient treatment within the prison colony; the prison laboratory analyzes all positive cultures from prisoners. Gender, date of birth or age, date of culture, phenotypic drug-susceptibility results, prison status, and history of previous treatment were determined by direct questioning of each person and examining the medical and laboratory notes. The study was approved and supervised by the local drug-susceptibility testing and ethics steering committee.

Drug Resistance

All sputum specimens were cultured on Lowenstein-Jensen media; specimens were coded so that those performing the laboratory analysis were unaware of the epidemiologic data. Rifampin, isoniazid, and multidrug resistance were determined genotypically. DNA was chloroform-extracted (20), and rifampin resistance was determined by using a nested PCR-amplification approach. The PCR product was reverse hybridized to probes immobilized on a membrane to detect mutations within the *rpo*B gene (21–23) (INNO-LiPA Rif.Tb; Innogenetics, Ghent, Belgium).

With the use of biotinylated primers, an inhouse macroarray was developed and used to identify mutations associated with isoniazid resistance in the genes *inhA, katG,* and *ahpC* (24). Briefly, crude bacterial lysates were prepared from each isolate as described above (20). With the use of published *M. tuberculosis* sequences, digoxigenin-labeled PCR products were generated in a multiplex PCR amplifying four genomic sequences associated with resistance to isoniazid: 251- and 232-bp regions of *katG,* including codon 315 and codon 463, a 241-bp region of *aphC-oxyR;* and a 265-bp region of the *inhA* locus, including the regulatory region. The sequences used to design the PCR primers were MTU06270 for *katG*, MTU16243 for *aphC,* and MTU66801 for *inhA*. Primer sequences were: tomkp 1 = GGCCCCGAACCCGAGGCTGC; tomkp2 = AACGGGTCCGGGATGGTGCCG; tomkp3 = GCCGACGAG TTCGCCAAGGCC; tomkp4 = ACGACGCCGCCGCCCAT-GCG; tomap1 = CCGCCGATGAGAGCGGTGAGC; tomap2 = CCACTGCTTTGCCGCCACCGC; tomip1 = CACCCG-CAGCCAGGGCCTCG; and tomip2 = CGATCCCCCGGTTT CCTCCGG.

PCR was conducted in a 25-µL reaction containing 2.5 µL 10X reaction buffer (Qiagen, Crawley, UK), 0.5 µL 2 mM dNTP (deoxynucleoside triphosphate; Pharmacia, Little Chalfont, UK), 20 µM each of the eight primers, 0.05 nmol dUTP (digoxigenin-labeled deoxyuridine triphosphate; Roche, Lewes, UK), 0.5 U of HotStartaq (Qiagen), and 1 µL DNA extract. Cycling was conducted with a PE2400 thermal cycler (Applied Biosystems, Foster City, CA) programmed to hold 95°C for 15 min then 30 cycles of 15 sec at 95°C, 30 sec at 60°C, and 60 sec at 72°C.

Eleven probes for the macroarray were diluted in water to 20 µM and applied to printed cells on a numbered nylon membrane (Osmonics, Minnetonka, MN). The oligonucleotides were UV-crosslinked for 1 min in an Amplirad UV box (GRI, Braintree, UK). Membranes were washed twice for 5 min in 0.5XSSC (1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and air-dried. The individual arrays were then separated and placed in a 2.5-mL minifuge tube. These were stored in darkness at room temperature until use. A reverse hybridization procedure was used to interrogate the macroarray, and hybridization was visualized by detecting the digoxigenin label colorimetrically according to the manufacturer's instructions (Roche). Extensive precautions (including the use of a three-room PCR suite, dedicated equipment, and multiple positive and negative controls) were taken to avoid cross-contamination.

Molecular Epidemiology

All cultures were analyzed by spoligotyping (15) and VNTR analysis (25) by using standard methods. For VNTR analysis, PCR-amplification was performed to identify the number of exact tandem repeats at five loci, A to E. PCR products were electrophoresed on a sieving agarose gel to determine the size of the product and compared against a standard to determine the number of exact tandem repeats at each locus. The result can be read as a numerical code that can be compared across large groups of isolates. RAPET was performed as described by Yates et al. (2). Fifty-five viable cultures were available for RFLP typing, targeting the IS*6110* by using the internationally standardized protocol described by van Embden et al. (11).

Results

Drug Resistance

One hundred forty cultures were collected from the five sites, and genotypic analysis was performed in London by British and Russian scientists to determine rifampin and isoniazid resistance (Table 1). One hundred thirty-four cultures were successfully amplified (four cultures gave no amplification, and two cultures were identified as *M. gordonae*). Overall, 78 (58.2%) had mutations consistent with rifampin resistance. The Innolipa assay may underestimate true rifampin resistance by 5% to 7% (21,23). Detection of mutations, however, is indicative of true rifampin resistance in nearly all cases. Table 1 indicates the overall rate of rifampin resistance, the rate for each laboratory center, and the rate of molecular rifampin resistance as a fraction of each laboratory culture population.

A macroarray was used to identify isoniazid resistance in 128 cultures that could be evaluated (amplification products were absent in 6 cultures successfully amplified for *rpoB*). When this method was used, 66 (51.6%) isolates were isoniazid resistant, and 62 were sensitive. This technique only detects approximately 75% to 90% true resistance so the rate of resistance is underestimated. Table 1 indicates the proportions of isoniazid-resistant cultures from each center; most came from the prison TB laboratory. Patient age and being a prisoner were significant risk factors for rifampin but not isoniazid resistance; previous treatment was a risk factor for resistance to both drugs (Table 2).

Beijing Family and Molecular Epidemiology

Of the 140 original samples, 138 produced a spoligotype profile. The samples were divided into 27 groups with 16 individual isolates and 10 clustered groups: 5 clusters containing 2 isolates, 2 with 3 isolates, 1 with 4 isolates, and 1 with 8 isolates. The 10th cluster accounted for 94 (68.1%) of the 138 cultures and comprised the characteristic Beijing strain (i.e., binding occurred only to the final nine spoligotype probes). When VNTR typing was used, 136 isolates (4 did not amplify) were divided into 25 groups: 13 were individual profiles with 4 clusters containing 2 isolates, 2 clusters of 3 isolates, 1 cluster of 4 isolates, 2 clusters of 5 isolates, 1 cluster of 9 isolates, 1 cluster of 13 isolates, and a dominant cluster of 73 strains. Combining spoligotyping and VNTR typing divided the 138 strains into 44 types in total.

The proportion of Beijing isolates compared to the total number of cultures submitted by each site was calculated (Table 3). Both the total number and highest proportion of Beijing family cultures came from the prison and the oblast TB dispensary, which would be inclined to have complicated cases, including ones in former prisoners.

Beijing Strains and Drug Resistance

Overall, 65 (69.1%) of 94 Beijing family cultures were rifampin resistant (Table 4); the highest proportion of rifampin-resistant Beijing isolates (31/38 or 81.6%) occurred in the prison. Proportions at the other centers varied from 60.0% to 77.8%.

High rates of rifampin resistance were seen in Beijing isolates from all centers. Overall, the proportion of Beijing isolates with rifampin resistance was over twice that of non-Beijing isolates (69.1% vs. 29.5%). The Beijing cluster of 94 isolates was subtyped by using VNTR (Table 5); 92 isolates were successfully analyzed, and no amplification occurred in 2 isolates. Overall, the 91 Beijing isolates were subdivided into

Table 1. Molecular resistance to rifampin and isoniazid in *Mycobacterium tuberculosis* isolates with proportional analysis by center/laboratory, Samara Oblast, Russian Federation

Laboratory location	No. of isolates examined for rifampin resistance (n=130)[a]	No. of rifampin-resistant isolates (%)	No. of isoniazid-resistant isolates (%)[b]
All	134	78 (58.2)	66 (–)
Prison TB colony	45	36 (80.0)	27 (60.0)
City dispensary 1	11	4 (36.4)	4 (36.4)
City hospital 1	34	18 (52.9)	14 (41.2)
Novokuibyshevsk town laboratory	21	9 (42.9)	11 (52.4)
Oblast regional TB laboratory	19	11 (57.9)	10 (52.6)

[a]Of 140 cultures, 134 results were produced, but only 130 were available for results by center; 2 were *M. gordonae*, amplification failed for 4 cultures, and center could not be definitively identified for 4.
[b]Of 140 cultures, amplification products that could be evaluated were obtained from 128 cultures.

Table 2. Patient factors associated with rifampin- and isoniazid-resistant isolates[a]

Factor	Rifampin resistance (n=78)	Isoniazid resistance (n=66)
Age	$p<0.05$[b]	$p>0.05$
Known prisoner	RR 1.68,[b] CI 1.29 to 2.18	RR 1.36, CI 0.97 to 1.89
Known previous treatment	RR 1.71,[b] CI 1.07 to 2.74	RR=2.04,[b] CI 1.01 to 4.10

[a]RR, risk ratio; CI, confidence interval.
[b]Statistically significant

13 groups with 8 individual VNTR types and 2 different clusters of 2 isolates, 1 cluster each of 3 isolates and 1 of 9 isolates. The 13th cluster contained 68 isolates (VNTR 42435) or 72.3% of the total number of Beijing isolates seen. Most isolates within this VNTR type were rifampin resistant: 50 (73.5%) of 68 were resistant, 16 (23.5%) were sensitive, and 2 had equivocal results. Within this dominant VNTR type, the proportion of the type that was rifampin resistant by institution was calculated (Table 6). Within the prison and oblast dispensary 90% to 91% of the strains of the VNTR 42435 type were rifampin resistant.

Of the 55 viable isolates that were RFLP-IS*6110* fingerprinted, 23 were members of the main VNTR 42435 group (Figure). A small number of similar but distinct isolates were seen across the sites, including the W148 strain seen in Siberian prisons and elsewhere in the former Soviet Union (18). This observation was confirmed by RAPET typing (data not shown).

Discussion

This analysis is part of an ongoing program to develop the capacity to accurately determine drug resistance and monitor the epidemiology of drug-resistant TB in both civilian and prison sectors in Russia. For multidrug-resistant TB, cure rates are lowered and infectious patients remain sputum smear-positive longer, increasing the probability that others will be infected.

As part of this program, we initiated a collaborative study in Samara, Russia, 1 of 89 regions or oblasts within the Russian Federation. In 2000, the new case rate in Samara was 87.6/100,000 (2,890 cases) in a population of 3,308,000 (including nonresidents and prisoners) with a death rate of 13.6 /100,000 (443 cases) (9,10). TB appears to be a problem in all Russian prisons including those in Samara, where the total number of TB patients at the time of the study was 1,800 cases (~34% of all registered TB patients in the region) (9,10). Because of an increasing number of intravenous drug users in the region, the problem of coinfection with HIV is likely to become increasingly important as the number of HIV-positive persons increases (>12,000 HIV-infected persons were registered in Samara Region in May 2001).

When molecular genotypic resistance analysis was used, 58.2% of cultures were rifampin resistant, and at least 51.6% were shown to be isoniazid resistant. Within this population, there were 123 molecular results for both rifampin and isoniazid, and 55 cases were definitely multidrug-resistant TB (55/123 or 44.7%); and 24 were possible cases (i.e., resistant to rifampin but sensitive to isoniazid with genotypic methods). These rates are high for both rifampin resistance and multidrug-resistant TB. Nevertheless, care must be taken in interpreting these results: they are likely to be overestimates since cultures were derived from new and chronic cases. Very high rates of rifampin resistance were seen in cultures drawn from the prison TB colony (80%), and 35 (79.5%) of 44 patients with known rifampin-resistant cases had been in prison before. Being a prisoner and patient age were risk factors for rifampin

Table 3. Proportion of Beijing family in isolates of *Mycobacterium tuberculosis* with distribution by center/laboratory, Samara Oblast, Russian Federation

Location	No. of isolates available for typing (n= 138)[a]	No. of Beijing family isolates (%)
All	138	94[b] (68.1)
Prison TB colony	49	40 (81.6)
City dispensary 1	11	6 (54.5)
City hospital 1	34	18 (52.9)
Novokuibyshevsk town laboratory	21	12 (57.1)
Oblast regional TB laboratory	19	14 (73.7)

[a]Of 140 cultures, 138 were available for spoligotyping (2 were not multiresistant TB); however, 4 cultures could not be attributed to a center.
[b]138 cultures were evaluated, but 4 could not be attributed to a specific center, i.e., 134 attributable cultures.

Table 4. Number and proportion of rifampin resistant isolates within Beijing family at each center

Location	No. of Beijing isolates	No. of Beijing isolates resistant to rifampin (%)	No. of isolates resistant to rifampin	% of rifampin-resistant isolates that belong to Beijing family
All	94	65 (69.1)	78	83.3
Prison TB colony	40	31 (77.5)	36[a]	86.1
City hospital 1	18	14 (77.8)	18	77.8
Novokuibyshevsk town laboratory	12	7 (58.3)	9	77.8
Oblast regional TB laboratory	14	10 (71.4)	11	90.9

[a]Four gave no amplification; numbers too small to evaluate for city dispensary l.

Table 5. Variable number tandem repeat (VNTR) analysis of Beijing isolates (n=91) and comparison with rifampin sensitivity

VNTR	Total in each VNTR type	No. of rifampin resistant in each type	No. of rifampin sensitive in each type
12435	3	3	0
12534	9	8	1
12535	1	1	0
22232	2	0	2
22435	1	0	1
32413	1	-	-[a]
32433	1	0	1
32435	1	0	1
42234	1	0	1
42434	2	0	2
42435	68	50	16 (2 mixed=nonreadable)[a]
42436	1	1	0
42532	1	1	0

[a] In three, no amplification of *rpoB* gene or mixed reaction.

resistance (but not for isoniazid resistance). Previous treatment was a risk factor for both rifampin and isoniazid resistance.

It remains unclear whether rifampin resistance is being introduced into the oblast prison from the pretrial centers or developing within prison because of interrupted therapy caused by poor adherence to treatment or release before treatment is completed. Previous treatment has been shown to be a risk factor for drug resistance (1,2).

Molecular epidemiologic techniques can help monitor the spread of TB isolates. These techniques will be of particular value where drug resistance is common and where strains such as the Beijing family appear more commonly associated with the type of drug resistance likely to lead to therapeutic failure and prolonged infectivity. In this study, rapid PCR-based techniques such as spoligotyping were used; although less discriminating than IS*6110*-RFLP analysis, PCR techniques permitted safe, rapid analysis of a specific family of strains. Spoligotyping's ability to discriminate was greater than that of VNTR; in combination, the techniques divided the isolates into 43 groups, indicating the benefits of combining these systems. Spoligotyping demonstrated a high proportion (94 [68.1%] of the 138 cultures) of the characteristic Beijing family of strains, which has been previously associated with drug resistance.

Some researchers have argued that the Beijing genotypes may have a selective advantage over other genotypes; BCG-induced immunologic protection may not protect against this strain, which would be an "escape mutant" (17). In that study in Vietnam, although the Beijing isolates were occurring more frequently among those vaccinated with BCG, this difference was not statistically significant (17). In that study, more Beijing isolates occurred in younger age groups, suggesting possible recent transmission; in our analysis, patients with Beijing isolates were younger than those with non-Beijing isolates (data not shown), which might support this conjecture. Other researchers have argued against any selective advantage, pointing out that Beijing isolates have spread widely in the United States, where BCG vaccination is not used (18). Nevertheless, BCG vaccination may accelerate the dominance of this family in regions once the strains have been introduced. The Russian Federation TB Service has a comprehensive program of BCG vaccination, which might create a selection pressure. Analysis using VNTR also demonstrated a dominant VNTR type containing 68 (72.3%) isolates (VNTR 42435) of the total number of Beijing isolates. A small number of similar but distinct isolates were seen on RFLP-IS*6110* fingerprinting, including the W148 strain (17 band) seen in Siberian prisoners and elsewhere in the former Soviet Union (18). In 1998–2000, this strain was seen in 190 prisoners with multidrug-resistant TB in a prison in Tomsk, Western Siberia (18,19). None of the Samaran prisoners were from Siberia.

Another probable reason for the successful expansion of the Beijing family is its association with multidrug-resistant TB; Beijing types such as W4 and "210" appear widespread in China but are fully susceptible to drugs (18). Low cure rates of smear-positive prisoners would lead to prolonged infectivity. The oblast prison, in particular, has a high proportion of Beijing isolates and a high rate of rifampin resistance, suggesting that the Beijing family is dominant there and accounts for a large proportion of resistance. Nevertheless, drug resistance per se cannot be the only explanation for the success with these isolates.

The widespread population movements during the first half of the 20th century may have helped to distribute these isolates into new communities, which were then subsequently

Table 6. Proportion of dominant variable number of tandem repeat (VNTR) isolates of type 42435 known to be resistant or sensitive, by institution

Location	Proportion (n=68)	No. of VNTR type resistant to rifampin (%)[a]	Proportion of VNTR type sensitive to rifampin (%)
Prison TB colony	30	27 (90)	3 (10)
City dispensary	5	3 (60)	2 (40)
City hospital 1	7	3 (42.9)	4 (57.1)
Novokuibyshevsk town laboratory	12	7 (58.3)	5 (41.7)
Oblast regional TB laboratory	11	10 (90.9)	1 (9.1)

[a]No *rpoB* amplification in one culture in this group.

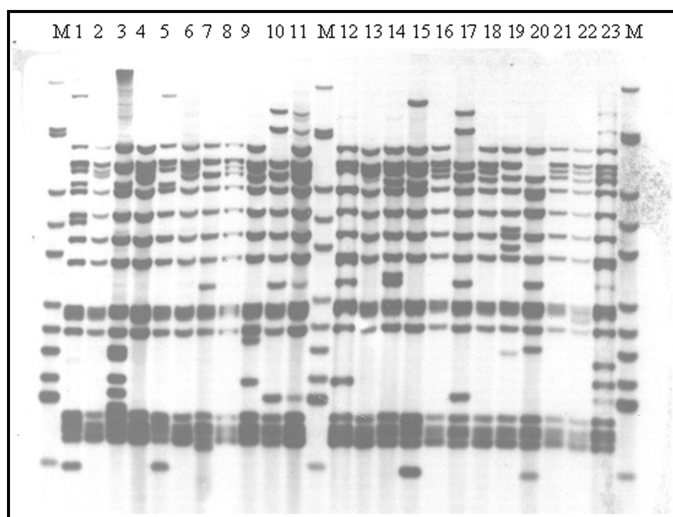

Figure. IS*6110* restriction fragment length polymorphism analysis of isolates from a dominant variable number tandem repeat group of Beijing family isolates at all sites, Samara, Russia. M indicates *Mycobacterium tuberculosis* strain MT14323. Isolates were from all five sites including the prison (tracks 13, 16,18, and 20).

selected for by the introduction of a comprehensive BCG vaccination policy and by the later development of drug resistance. Concluding that Beijing family isolates have undiscovered advantages is also reasonable. In this study a large proportion of the Beijing isolates were of the same VNTR type (42435), but little is known of any differences in biologic function that might be related to different types. A recent study has demonstrated that the coding sequence Rv3710 (*leu*A) encodes the production of active alpha-isopropylmalate synthase. Within the sequence lies the locus of VNTR 4155, and this may have a modifying role on the function of the enzyme (26).

High-quality national or regional drug resistance surveys are needed in other parts of the Russian Federation. Such surveys would lead to a clearer understanding of the true level of drug resistance and in turn facilitate clinical management and permit better empirical treatment strategies. Genotypic techniques are of value in determining rifampin resistance and likely multidrug-resistant TB. Although expensive compared to drug-susceptibility testing on solid media, genotyping techniques may be justified in populations with high levels of multidrug-resistant TB and high rates of concurrent HIV. Further analysis is required to confirm the spread of the Beijing family and determine whether it is imported into the prison and spread within the prison community.

Acknowledgments

We thank Malcolm Yates and Heather Jebbari for microbiological and statistical advice, respectively. We also thank all the microbiologists and clinical experts in Russia who contributed to the program.

The United Kingdom Department of International Development funded this program (CNTR 00 0634).

Mr. Drobniewski is director and consultant medical microbiologist of the U.K. Public Health Laboratory Service Mycobacterium Reference Unit, consultant physician in clinical tuberculosis at King's College Hospital, and professor of tuberculosis and mycobacterial diseases, University of London. His interests include all aspects of the diagnosis, epidemiology, and management of tuberculosis, particularly in the context of coinfection with HIV.

References

1. World Health Organization/International Union Against Tuberculosis. Antituberculosis drug resistance. Geneva: The Organization;1997.
2. World Health Organization/International Union Against Tuberculosis. Antituberculosis drug resistance. Geneva: The Organization; 2000.
3. Coronado VG, Beck-Sague CM, Hutton MD, Davis BJ, Nicholas P, Villareal C, et al. Transmission of drug-resistant *Mycobacterium tuberculosis* among persons with human immunodeficiency virus infection in urban hospital: epidemiologic and restriction fragment length polymorphism analysis. J Infect Dis 1993;168:1052–5.
4. Dooley SW, Jarvis WR, Marlon WJ, Snider DE. Multi-drug resistant tuberculosis. Ann Intern Med 1992;117:257–9.
5. Edlin BR, Tokars JL, Grieco MH, Crawford JT, Williams J, Sordillo E, et al. An outbreak of multi-drug resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N Engl J Med 1992;26:1514–21.
6. Breathnach AS, De Ruiter A, Holdsworth GMC, Bateman NT, O'Sullivan DGM, Rees PJ, et al. An outbreak of multi-drug resistant tuberculosis in a London teaching hospital. J Hosp Infect 1998;39:111–7.
7. Agerton T, Valway SE, Blinkhorn RJ, Shilkret KL, Reves R, Schluter WW, et al. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. Clin Infect Dis 1999;29:85–92.
8. Shilova MV, Dye C. The resurgence of TB in Russia. Philos Trans R Soc Lond B Biol Sci 2001;356:1069–75
9. Drobniewski FA, Balabanova Y. The diagnosis and management of multiple drug resistant tuberculosis at the beginning of the new millennium. Int J Infect Dis 2002;6:S21–31.
10. Balabanova Y, Fedorin I, Melentyeiv A, Kuznetzov S, Drobniewski F. Tuberculosis in Samara, Central Russia 1999–2000 [Abstract P42]. In: Abstracts of the 29th European Society for Mycobacteriology, Berlin, Germany; 2001 July 2–6. Berlin: European Society for Mycobacteriology; 2001.
11. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
12. van Soolingen D, de Haas PE, Hermans PW, van Embden JD. DNA fingerprinting of *Mycobacterium tuberculosis*. Methods Enzymol 1994;235:196–205.
13. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Palittapongarnpim P, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strain: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
14. Diaz R, Kremer K, de Haas PE, Gomez RI, Marrero A, Valdivia JA, et al. Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994–June 1995: utility of spoligotyping versus IS*6110* restriction fragment length polymorphism. Int J Tuberc Lung Dis 1998;2:743–50.
15. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
16. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portael F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of East Asia. J Clin Microbiol 1995;33:3234–8.

17. Ahn D, Borgdorff MW, Van LN, Lan NT, van Gorkom T, Kremer K, et al. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. Emerg Infect Dis 2000;6:302–5.
18. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol 2002;10:45–52.
19. Portaels F, Rigouts L, Bastian I. Addressing multidrug resistant tuberculosis in penitentiary hospitals and in the general population of the former Soviet Union. Int J Tuberc Lung Dis 1999;3:582–8.
20. Yates MD, Drobniewski FA, Wilson SM. Evaluation of a rapid PCR-based epidemiological typing method for routine studies of *Mycobacterium tuberculosis*. J Clin Microbiol 2002;40:712–4.
21. De Beenhouwer H, Lhiang Z, Jannes W, Mijs L, Machtelinckx L, Rossau R, et al. Rapid detection of rifampin resistance in sputum and biopsy specimens from tuberculosis patients by PCR and line probe assay. Tuber Lung Dis 1995;76:425–30.
22. Watterson SA, Wilson SM, Yates MD, Drobniewski FA. Comparison of three molecular assays for rapid detection of rifampin resistance in *Mycobacterium tuberculosis*. J Clin Microbiol 1998;36:1969–73.
23. Drobniewski FA, Watterson SA, Wilson SM, Harris GS. A clinical, microbiological and economic analysis of a national UK service for the rapid molecular diagnosis of tuberculosis and rifampin resistance in *Mycobacterium tuberculosis*. J Med Microbiol 2000;49:271–8.
24. Brown T, Maxwell O, Baker L, Drobniewski F, Anthony R, French G. Screening for mutations associated with INH resistance in clinical *Mycobacterium tuberculosis* isolates [Abstract P60]. In: Abstracts of the UK Public Health Laboratory Service Annual Scientific Conference, 2001, Nov 2–7. London: Public Health Laboratory Service; 2001.
25. Frothingham R, Meeker-O'Connell WE. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology 1998;144:1189–96.
26. Chanchaem W, Palittapongarnpim PA. A variable number of tandem repeats result in polymorphic alpha-isopropylmalate synthase in *Mycobacterium tuberculosis*. Tuberculosis 2002;82:1–6.

Address for correspondence: F. Drobniewski, Public Health Laboratory Service, Mycobacterium Reference Unit, Department of Microbiology and Infection, Guy's, King's and St Thomas' Medical School, King's College Hospital (Dulwich), East Dulwich Grove, London SE22 8QF, UK; fax: 02073466477; e-mail: francis.drobniewski@kcl.ac.uk

EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends Vol.8, No.5, May 2002

Tuberculosis-Related Deaths within a Well-Functioning DOTS Control Program

Maria de Lourdes García-García,* Alfredo Ponce-de-León,† Maria Cecilia García-Sancho,‡ Leticia Ferreyra-Reyes,* Manuel Palacios-Martínez,* Javier Fuentes,§ Midori Kato-Maeda,† Miriam Bobadilla,† Peter Small,¶ and José Sifuentes-Osornio,†

To describe the molecular epidemiology of tuberculosis (TB)-related deaths in a well-managed program in a low-HIV area, we analyzed data from a cohort of 454 pulmonary TB patients recruited between March 1995 and October 2000 in southern Mexico. Patients who were sputum acid-fast bacillus smear positive underwent clinical and mycobacteriologic evaluation (isolation, identification, drug-susceptibility testing, and IS*6110*-based genotyping and spoligotyping) and received treatment from the local directly observed treatment strategy (DOTS) program. After an average of 2.3 years of follow-up, death was higher for clustered cases (28.6 vs. 7%, p=0.01). Cox analysis revealed that TB-related mortality hazard ratios included treatment default (8.9), multidrug resistance (5.7), recently transmitted TB (4.1), weight loss (3.9), and having less than 6 years of formal education (2). In this community, TB is associated with high mortality rates.

In both humanistic and economic terms, the cost of deaths due to tuberculosis (TB) is staggering. In 1990 alone, approximately 2.5 million people died of TB, accounting for >25% of avoidable adult deaths in the developing world (1,2). Directly observed treatment strategy (DOTS), a comprehensive approach to TB control, is one of the most cost-effective health interventions available (3,4). In the context of a well functioning DOTS program, cure rates in excess of 80% can be expected. While these outcomes are assumed to decrease mortality rates, the detailed epidemiology of deaths in a well-functioning DOTS program by using modern molecular techniques has not been described.

Since 1995, we have conducted a population-based molecular epidemiologic study of TB in a health district in southern Mexico. Previous reports have documented the TB control program approaches World Health Organization benchmarks (5) and drug resistance is considerable and has an important negative impact on treatment outcomes (6). We now report the short- and long- term mortality rates due to TB in this cohort of TB patients. These data suggest that, as has been described with other diseases, excess mortality may persist for months or years after treatment completion, default, or failure (7,8).

Methods

The study site, described previously (5,6,9), is located in a predominantly urban region in the Orizaba Health Jurisdiction in the state of Veracruz, which encompasses 134 square km and has a population of 284,728 (10). The incidence rate of TB during the year 2000 for the state was higher than that for the nation (28.0 vs. 15.9 per 100,000 inhabitants) (11).

Community-based screening of chronic coughers (>2 weeks) was performed from March 1995 to October 2000. Patients with positive AFB sputum smears underwent epidemiologic, clinical (standardized questionnaire, physical exam, chest x-ray, and HIV test), and mycobacteriologic evaluation. Treatment was provided in accordance with official norms (12,13). Treatment outcomes were classified as previously described (6). Annual follow-up was performed for treatment outcome and vital status. Deaths were confirmed with death certificates. A close caregiver was interviewed to elicit signs and symptoms of the terminal illness and "probable cause of death" was assigned by two of the authors (JF, LF). Informed consent was obtained from participants. The study was approved by the appropriate institutional review boards.

Microbiologic Evaluation

Mycobacterial culture, identification, and susceptibility testing were performed on sputa from each enrolled patient. In brief, unconcentrated sputum was spread onto Lowenstein-Jensen media (DIFCO, Detroit, MI) at the local laboratory, and the remaining sputum was frozen at –70°C. The tubes were examined on a weekly basis until growth was detected. Cultures were reported as negative if no growth occurred after 8 weeks. Cultures with visible growth were forwarded to the department of Mycobacteriology at the Instituto de Diagnostico y Referencia Epidemiológicos (March 1995 to December 1997) or to the Clinical Microbiology Laboratory of the Instituto Nacional de Ciencias Médicas y Nutrición (INCMNSZ) (January 1998 to October 2000) for definitive biochemical

*Instituto Nacional de Salud Pública, Cuernavaca, México; †Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán," México City, México; ‡Instituto Nacional de Enfermedades Respiratorias, México City, Mexico; §Secretaría de Salud del Estado de Veracruz, Xalapa, México; and ¶Stanford University, Palo Alto, California, USA

identification at the species level (14,15). The frozen sputum sample was processed if the first sample was contaminated or had no growth. Identification and drug susceptibility tests were carried out using conventional methods and the BACTEC system (Becton Dickinson, Cockeysville, MD).

Mycobacterium tuberculosis Fingerprinting

Mycobacteria isolated from study patients were genotyped at Stanford University from March 1995 to February 1997 and at INCMNSZ from March 1997 to October 2000 with the internationally standardized IS*6110*-based restriction fragment length polymorphism (RFLP) technique and compared by using a computer-assisted visual approach (Bioimage AQ-1 analyzer and Molecular Fingerprinting Analyzer, version 2.0) (16,17). Isolates with identical IS*6110* fingerprints that contained five or fewer hybridizing bands underwent spoligotyping at INCMNSZ as described (18,19). To assess transmission of *M tuberculosis* that rapidly progressed to disease, we established a 1-year period for defining clustering as described (20). Cases were considered "clustered" if two or more isolates from different patients were identified within a year that had 1) six or more IS*6110* bands in an identical pattern, or 2) five or fewer bands with identical IS*6110* fingerprints and matching spoligotypes. The first patient diagnosed in each cluster and those with unique fingerprints were classified as having reactivated disease.

Analysis

Deaths were attributed to TB based on the death certificate (TB listed as the cause of death), interview of close caregiver (TB identified as probable cause of death), and active TB at the time of death (positive AFB smear after last treatment or during treatment if the patient did not complete treatment performed ≤6 months before death). A patient's death was attributed to TB if two or more of these conditions were met. Bivariate and multivariate analyses were performed to describe association between sociodemographic (age, sex, household characteristics, occupation, ethnicity, years of formal education, place of residence, and social security), behavioral (usage of drugs and alcohol, and previous incarceration), clinical (previous diagnosis of TB; other associated diseases, previous hospitalizations; HIV infection, duration of symptoms previous to diagnosis; time elapsed between diagnosis and initiation of treatment and between initiation of treatment and smear conversion; and symptoms such as cough, hemoptysis, fever, night sweats, weight loss, and general malaise), bacteriologic (number of bacilli per microscopic field, drug resistance, RFLP, or spoligotype pattern), therapeutic (compliance, treatment outcome, and retreatment) variables, and death due to TB and other causes. Survival analyses included Kaplan-Meier curves and Cox proportional hazards model for TB and non-TB–related deaths using as reference time the period elapsed from diagnosis. Variables were entered into the models according to their statistical significance in univariate analysis and their biological relevance. The percentage of population attributable risk was calculated for variables that were included in the final model (21). DBASE IV and STATA 5.0 programs were used for data analysis (22).

Results

During the study period, 454 patients were diagnosed with pulmonary TB. Most patients were men (270 [59.5%] of 454) and the median age was 42 years (range 12–97 years). Most came from lower socioeconomic status as indicated by household characteristics, formal years of education, and occupation. Of 438 patients, 74 (17%) lived in households with earthen floor, 214 (49.3%) of 434 had no access to potable water within the household, 293 (65.4%) of 448 had <6 years of formal education, and 104 (23.4%) of 445 were manual workers. The prevalence of HIV was 2.1% (9 of 429), combined resistance to isoniazid and rifampin was seen in 26 (6.7%) of 388 patients, and other resistance patterns in an additional 61 (15.7%) of 388. One hundred fifty-one (37%) of 413 patients had chest radiographs with evidence of pulmonary cavitation, and 72 (17.4%) of 413 patients had interstitial pulmonary infiltrates. Median time from initiation of symptoms to treatment was 101 days (range 3–2,307 days), median time from diagnosis to initiation of treatment was 5 days (range 0–594 days), and median time for sputum conversion was 42 days (range 15–1,228 days).

Treatment and Follow-Up

Of 11 patients who refused treatment, 4 died. Of 443 initiating treatment, 75.4% initiated treatment <10 days after diagnosis; 96.5% received supervised treatment. Outcomes for patients were as follows: 357 (80.6%) were cured of whom 314 (70.9%) were bacteriologically confirmed, 41 (9.3%) defaulted, and 20 (4.5%) failed treatment; 16 (3.6%) died during treatment, and 9 (2%) transferred out of study area. Patients were tracked for a median of 839 days (range 3–2,402). Sixty-one additional patients died during follow-up after treatment. Death was due to TB in 34 (41.9%) of 81 instances; 2 deaths were in patients who did not receive treatment, 11 in patients receiving treatment, and 21 after treatment. Tuberculosis mortality rates were higher during treatment versus after treatment (1.3/10,000 days vs. 0.7/10,000 days, $p<0.01$). Crude comparison of sociodemographic, bacteriologic, and clinical characteristics of patients who died from TB or from other causes and surviving patients are shown in Table 1. Patients who died from TB had a higher probability of having been treated for a previous TB episode, showed more severe clinical symptoms, and had drug-resistant isolates. Less frequently they had other coexisting chronic conditions, such as HIV infection or hepatic cirrhosis. The patients who died from TB also had longer delays before diagnosis, treatment, and sputum conversion. These patients had higher probabilities of default, failure, and having subsequent TB episodes. The most common non-TB causes of death included diabetes (12), cirrhosis and other liver diseases (9), AIDS (6), cancer (2), and cardiovascular diseases (3). Kaplan

Table 1. Sociodemographic, clinical, bacteriologic, and therapeutic characteristics of smear-positive pulmonary tuberculosis (TB) patients according to cause of death, Orizaba, Veracruz, 1995–2000

Variables	Died from TB (n=34) (%)	Died from other causes (n=47) (%)	Survived (n=373) (%)	p value[a]
Sociodemographic				
Median age (range)	30 (24–73)	47 (22–70)	40.5 (12–82)	0.05
Men	52.9	76.6	57.9	0.04
Indigenous origin	20.6	4.3	16.1	0.07
<6 years formal education	73.5	76.6	62.2	0.02
Rural and industrial workers	11.8	25.5	23.6	0.2
Previous imprisonment	14.7	34.0	27.9	0.2
Previous TB treatment	47.1	34.0	14.2	<0.0001
Previous hospitalization	58.8	51.1	45.6	0.2
Residence in shelters	5.9	10.6	5.1	0.3
Alcohol use	38.2	66.0	43.7	0.005
Household crowding	47.1	27.7	37.5	0.2
Household with earthen floor	20.6	10.6	16.6	0.4
Clinical				
HIV infection	8.8	10.6	.3	<0.0001
Hepatic cirrhosis	0	6.4	1.3	0.06
Body mass index (≤18)	47.1	27.7	20.4	<0.0001
Hemoptysis	26.5	38.3	37.8	0.5
Fever	44.1	61.7	44	0.04
Night sweats	58.8	59.6	57.6	0.2
Weight loss (>15 %)	47.1	46.8	29.2	0.002
Radiologic nodes	5.9	6.4	7.0	0.9
Cavities	44.1	25.5	33.2	0.2
Median time interval between initiation of symptoms and treatment (range in days)	17 (1–212)	8 (0–158)	5 (0–322)	0.004
Median time interval between diagnosis and treatment (range in days)	141.5 (78–991)	126 (4–439)	99.5 (4–1,723)	0.01
Bacteriologic				
Resistance to isoniazid and rifampin	29.4	17.0	2.1	<0.0001
Other resistance	11.8	19.1	12.9	0.2
<10 bacilli per 100 fields	79.4	87.2	86.1	0.5
Median time interval between treatment and sputum conversion (range in days)	95 (35–530)	44 (19–182)	42.5 (15–348)	0.03
Treatment outcome				
Cure	5.9	61.7	87.4	<0.0001
Failure	20.6	10.6	2.1	<0.0001
Default	32.4	10.6	6.7	<0.0001
Retreatment	17.6	23.4	5.4	<0.0001

[a]Chi square test, analysis of variance test.

Meier survival probabilities from TB deaths were 97.3% after the first 6 months, 95.8% after 1 year, 93.7% after 2 years, and 91% after 3 years.

RFLP and Spoligotyping Results

M. tuberculosis culture, drug test, and IS*6110* RFLP and spoligotyping were available for 326 (72%) isolates. Comparison of patients whose isolates were available for genotyping to those whose isolates were unavailable indicated that patients for whom fingerprint analysis was not performed had a higher probability of being of native origin in Mexico (30 [24.2%] of 124 vs. 39 [12%] of 324, p= 0.001) and of living in households with earthen floor (31 [25.8%] of 120 vs. 43 [13.5%] of 318, p=0.002). Forty (12.3%) of the 326 evaluated cases were

in clusters. The frequency of being members of clusters of recently transmitted disease was higher among patients dying from TB than among those dying from other causes or surviving (8 [28.5%] of 28 patients who died from TB vs. 32 [10.7%] of 298 patients who died from other causes and survivors, p=0.01).

Factors Associated with Mortality Rates

Predictors of death due to TB by Cox regression analysis included treatment default, resistance to isoniazid and rifampin, and recently transmitted TB controlling for time of occurrence of death, weight loss >15%, and years of formal education (Table 2). The effect was not modified when gender, age, HIV infection, crowding in the household, household characteristics, occupation, ethnicity, previous treatment, delay in seeking treatment, specific symptoms, type of radiologic lesions, and other kinds of diseases were introduced into the model.

After controlling for age, predictors of non-TB death included HIV-infection, hepatic cirrhosis, and weight loss. Recently transmitted TB was not associated with other causes of death (Table 2).

The proportion of death due to TB and to other causes attributable to the different categories of risk factors is shown in Table 2. In the study population, 60% of deaths due to TB were attributable to drug resistance and treatment default.

Patients with recently transmitted disease had a lower probability of survival compared with patients with reactivated disease (p=0.007) (Figure). When causes of death were analyzed according to genotype, we found that TB was the cause of death in 8 (20%) of 40 patients with recently transmitted disease and in 20 (7%) of 286 patients with reactivated disease (p=0.01).

Sensitivity Analysis

Analysis of the distribution of the time between diagnosis dates of successive matching fingerprints indicated that 49.4% (95% confidence interval [CI] 44% to 54%) of all isolates with matching fingerprints patterns were identified within 1 year. When the interval was modified to 6, 18, 24, 30, and 36 months, we found that the proportion of clustered cases increased (8.3%, 13.8%, 13.8%, 14.4%, and 14.7%, respectively), (χ^2 trend, p=0.02). The association between clustered cases and death due to TB continued to be positive for each of the other definitions of interval for clustering: 3.6 (1.1 to 10.5), p=0.01; 2.8 (1.0 to 7.3), p=0.03; 2.8 (1.0 to 7.3), p=0.03; 2.7 (0.9 to 6.8), p=0.04; and 2.6 (0.9 to 6.6) p=0.04.

Discussion

This study describes high mortality rates from TB in a cohort of pulmonary TB patients who resided in an area with a low rate of HIV infection and were treated in the context of a well-functioning DOTS program. Patients were followed for an average of 2.3 years after diagnosis. Tuberculosis was associated with high rates of deaths both during treatment and after treatment completion, default, or failure. The main independent risk factors for death due to TB were treatment default and being infected with multidrug-resistant *M. tuberculosis*. Additionally, data indicate that cases due to ongoing transmission of TB may have higher mortality rates than cases due to reactivation of latent disease. These results suggest that current techniques underestimate death associated with TB and provide further impetus not only to treat but also to prevent TB.

Case completion rates are the standard by which the effectiveness of DOTS-based treatment programs are judged. A large body of evidence collected under diverse settings demonstrates completion rates of 81% and cure rates of 73% (23). The epidemiology of death in these programs is less well studied. Studies performed in HIV-endemic settings constitute an exception as the high rate of deaths occurring while patients are still on antituberculous therapy are cited as an impediment to achieving desired cure rates (24).

Table 2. Population attributable-risk percent and hazard ratios for death among smear-positive tuberculosis (TB) patients, Orizaba, Veracruz, 1995–2000[a]

Variables	Population attributable risk (%)	Adjusted hazard ratio	95% CI	p value[b]
Death due to TB[c]				
Treatment default	28.7	8.9	3.3 to 24.4	<0.0001
Resistance to isoniazid and rifampin	25.9	5.7	2.0 to 16.3	<0.001
Clustered	18.8	4.1	1.6 to 10.0	0.002
Weight loss (>15%)	-	3.9	1.5 to 10.9	0.007
Formal education <6 yr	-	1.8	0.6 to 5.2	0.3
Death due to other causes				
HIV/AIDS	11.1	33.1	11.4 to 95.4	<0.0001
Hepatic cirrhosis	6.6	5.7	1.6 to 19.7	0.006
Weight loss (>15%)	-	3.3	1.6 to 6.7	0.001
Age (yrs)	-	1.02	0.99 to 1.04	0.07

[a]CI, confidence interval; -, not applicable.
[b]Cox proportional hazards model.
[c]Controlling for death before or after treatment completion, default, or failure.

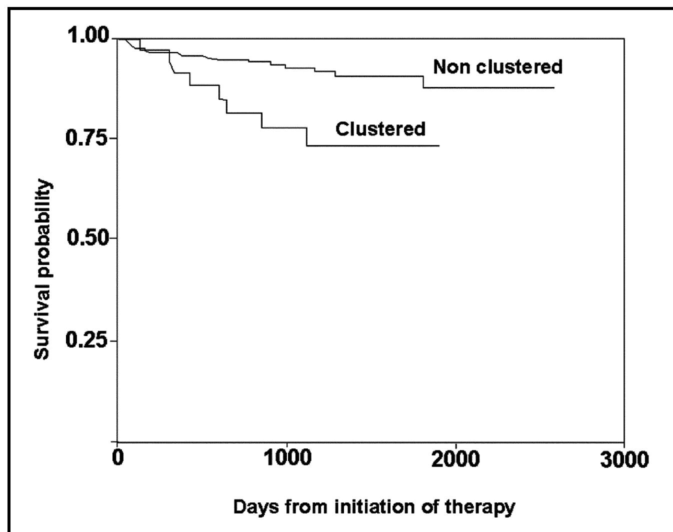

Figure. Estimated survival of smear-positive pulmonary tuberculosis patients according to clustered or unique fingerprint pattern in a low HIV-prevalence community (p=0.01).

Our study reports TB mortality rates both during and after treatment completion, default, or failure. Of the 443 patients who were started on therapy, 16 died before completing therapy. This frequency (3.6%) compares favorably to that reported in other studies (4% in the Gambia [25], 15% in Bolivia [26], 25% in Malawi [27], and 28% in India [28]). However, in contrast to previous reports, we had a median of 833 days of follow-up after the completion of therapy, during which 21 additional patients (4.6%) died of TB, comparable with reports from the 1960s and 1970s (29). This mortality rate exceeds that expected according to age-adjusted state mortality statistics (30).

Although not novel, the risk factors for death due to TB identified in this cohort are noteworthy; these findings suggest that, even in a high-quality TB control program, additional efforts could yield important benefits. Treatment default has been previously described as associated with higher mortality rates in Mexico (31) and elsewhere (32,33). Concerted efforts to further reduce default may disproportionately decrease deaths. We also found that drug resistance, and particularly multidrug resistance, are associated with death (6,14,34). Although the most appropriate strategy for managing drug-resistant TB is debated, these data provide additional impetus for evaluating novel approaches, such as those recently introduced in Mexico, for managing drug-resistant TB (13,35). We estimate that 60% of deaths due to TB in this setting were attributable to treatment default and multidrug-resistance. Additional studies are needed to quantitate the presumably greater mortality rates that result from default and drug resistance in other settings.

Although relatively uncommon in this setting, HIV was associated with a high overall mortality rate. Most of these deaths were due to non-TB–related death. HIV/AIDS was the main predictor for non-TB deaths in this cohort. In our study, eight of the nine HIV-seropositive persons died, three of them a year or more after completing antituberculous therapy. Several studies have demonstrated excess mortality rates after successful TB treatment in HIV-infected patients (24,36,37); the high rates have been attributed to HIV-related disease (38). These data emphasize the need to improve integration of quality treatment for both TB- and HIV-infected persons in populations that suffer from both diseases.

The most striking finding of the molecular epidemiologic component of this study is the association between clustering, which we interpret as indicative of recently transmitted disease, and TB-related death. Death due to TB was significantly more common among those who had recently transmitted disease than those with reactivated disease (28.6% vs. 7%, p=0.01). This association was independent of treatment default, multidrug resistance, time of occurrence of death, weight loss, and years of formal education. Furthermore, the effect was not modified when other variables indicative of sociodemographic level (such as occupation, characteristics of the household, or ethnicity) or clinical variables indicating other diseases, including HIV infection were introduced in the model. Inferring causality from such associations is difficult, it seems biologically implausible that being more likely to die made people more likely to acquire recently transmitted TB. A more plausible explanation is that rapidly progressing to disease contributes to deterioration of the health of these patients and thus increases their likelihood of death. The phenomenon is well described for other diseases such as measles. However, why this would be more pronounced for recently transmitted disease is unclear.

The association between clustering and death could be spurious because of limitations in molecular or conventional epidemiology. The validity of clustering as a proxy for recent transmission might be challenged by the fact that we did not perform fingerprint analysis on all isolates (39). Comparison between patients whose isolates were genotyped and those whose isolates were not genotyped showed that results of this study may not be generalizable to indigenous or lower socioeconomic groups. Our selection of the time interval allowed us to consider TB that was transmitted very recently and progressed to disease. The validity of using a 1-year interval was confirmed with the sensitivity analysis using different time periods (6, 18, 24, 30, and 36 months) as association between clustered cases and death persisted, despite modification of the time interval and by the fact that almost 50% of isolates with matching DNA fingerprint patterns occurred within 1 year of identification of the previous case. Conventional epidemiologic approaches to studying the cause of death are difficult. Death certificate data are notoriously unreliable, and whether patients died of TB or other causes is not certain (40). Therefore, we added other criteria that included the interview of a close caregiver and activity of TB at the time of death to validate our definition of death due to TB. The cause-of-death profile derived from interview of a close caregiver has been demonstrated to be useful for planning purposes (41). We consider that this definition adequately identified TB-related death

as characteristics of patients dying from TB differed from patients dying from other causes in several important aspects (drug resistance, coexisting chronic conditions, and clinical severity) and allowed the identification of different risk factors for TB-related death and for death due to other causes.

If confirmed in other settings, the conclusions of this study have important implications for control programs. Most importantly, given that current surveillance data are collected at the conclusion of therapy, this method probably underestimates the true impact of TB on a population's death. Given the increasing role of cost-efficacy modeling in setting health-care priorities, this oversight has important consequences. In addition, if recently acquired TB exhorts a greater mortality rate than that due to reactivated infection, the importance of interrupting TB transmission is further elevated.

Acknowledgments

We thank Luis Juarez and Bulmaro Cano for data processing; Ed Desmond, Travis Jobe, and Areli Martínez-Gamboa for training and technical support with spoligotyping; Carmen Soler for support for HIV tests; Manuel Tielve and Rubén Acevedo for support in interpreting chest x-rays; and the physicians, nurses, chemists, health promoters, recruiters, and interviewers in Orizaba who supported the diagnosis, treatment, and follow-up of patients.

This study was supported by the National Institutes of Health of the United States project no. A135969, by the Wellcome Trust, by the Howard Hughes Medical Institute (ID 55000632) and by the Mexican Council of Science and Technology, project nos. G26264M and 30987-M.

Dr. García-García is senior researcher and head of the Tuberculosis Unit at the National Institute of Public Health in Cuernavaca, Mexico, and International Research Scholar of the Howard Hughes Medical Institute. Her research interests lie in the epidemiology of TB.

References

1. Murray CJL, Styblo K, Rouillon A. Tuberculosis in developing countries: burden, intervention and cost. Bulletin of the International Union against Tuberculosis and Lung Disease 1990;65:6–24.
2. Visschedijk J, Simeant S. Targets for health for all in the 21st century. World Health Stat Q 1998;51:56–67.
3. Floyd K, Wilkinson D, Gilks C. Comparison of cost effectiveness of directly observed treatment (DOT) and conventionally delivered treatment for tuberculosis: experience from rural South Africa. BMJ 1997;315:1407–11.
4. Migliori GB, Ambrosetti M, Besozzi G, Farris B, Nutini S, Saini, et al. Cost-comparison of different management policies for tuberculosis patients in Italy. AIPO TB Study Group. Bull World Health Organ 1999;77:467–76.
5. García-García ML, Small PM, Garcia-Sancho C, Mayar ME, Ferreyra RL, Palacios M, et.al. Tuberculosis epidemiology and control in Veracruz, Mexico. Int J Epidemiol 1999;28:135–40.
6. García-García ML, Ponce de Leon A, Jimenez Corona MA, Jimenez Corona AJ, Palacios Martinez M, Balandrano Campos S, et al. Clinical consequences and transmissibility of drug resistant tuberculosis in Southern Mexico. Arch Intern Med 2000;160:630–6.
7. Koster FT, Curling GC, Aziz KM, Haque A. Synergistic impact of measles and diarrhoea on nutrition and mortality in Bangladesh. Bull World Health Organ 1981;59:90:1–8.
8. Aaby P, Andersen M, Knudsen K. Excess mortality after early exposure to measles. Int J Epidemiol 1993;22:156–62.
9. García-García ML, Palacios Martínez M, Ponce de Leon A, Jimenez Corona ME, Jimenez Corona A, Balandrano-Campos S, et al. The role of core groups in transmitting tuberculosis in a high prevalence community in Southern Mexico. Int J Tuberc Lung Dis 2000;4:12–7.
10. Instituto Nacional de Estadística, Geografía e informática. Estados Unidos Mexicanos. XII censo general de población y vivienda 2000. Resultados preliminares. Aguascalientes; Ags; Instituto Nacional de Estadística, Geografía e Informática; 2000. p. 10.
11. Dirección General de Epidemiología. Sistema Único de Información. Secretaría de Salud 2001;18:4–7.
12. Secretaría de Salud. Norma Oficial Mexicana NOM-006-SSA2 1993, para la prevención y control de la tuberculosis en la atención primaria a la salud. Diario Oficial de la Federación, 1995;26 de enero:20–29.
13. Modificación a la Norma Oficial Mexicana NOM-006-SSA2-1993. Para la prevención y control de la tuberculosis en la atención primaria a la salud. Diario Oficial de la Federación, 2000;octubre 31.
14. Nolte FS, Mechock B. Mycobacterium. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH, editors. Manual of clinical microbiology, 6th ed. Washington: ASM Press; 1995. p. 400–37.
15. Instituto Nacional de Diagnóstico y Referencia Epidemiológicas. Manual de Técnicas y Procedimientos de Laboratorio en Tuberculosis. México: Secretaría de Salud, 1992. p. 9–20.
16. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
17. Woelffer GB, Bradford WZ, Paz A, Small PM. A computer-assisted molecular epidemiologic approach to confronting the reemergence of tuberculosis. Am J Med Sci 1996;311:17–22.
18. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
19. Van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 1995;33:3234–8.
20. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann Intern Med 1999;130:971–8.
21. Hennekens CH, Buring J. Epidemiology in medicine. Boston: Little, Brown and Company; 1986. p. 87–90.
22. Computing Resource Center. Stata reference manual: release 3. 5th ed. Santa Monica (CA): The Center; 1992.
23. World Health Organization. Global tuberculosis control (WHO/CDS/TB/2001.287). Geneva, Switzerland: The Organization; 2001.
24. Connolly C, Davies GR, Wilkinson D. Impact of the human immunodeficiency virus epidemic on mortality among adults with tuberculosis in rural South Africa, 1991–1998. Int J Tuberc Lung Dis 1998;2:919–25.
25. Lienhardt C, Manneh K, Bouchier V, Lahai G, Milligan PJ , McAdam KP. Factors determining the outcome of treatment of adult smear-positive tuberculosis cases in The Gambia. Int J Tuberc Lung Dis 1998;2:712–8.
26. Olle-Goig JE. Patients with tuberculosis in Bolivia: why do they die? Rev Panam Salud Publica 2000;8:151–5.
27. Harries AD, Nyirenda TE, Banerjee A, Boeree MJ, Salaniponi FM. Treatment outcome of patients with smear-negative and smear-positive pulmonary tuberculosis in the National Tuberculosis Control Programme, Malawi. Trans R Soc Trop Med Hyg 1999;93:443–6.
28. Datta M, Radhamani MP, Selvaraj R, Paramasivan CN, Gopalan BN, Sudeendra CR, et al. Critical assessment of smear-positive pulmonary tuberculosis patients after chemotherapy under the district tuberculosis programme. Tuber Lung Dis 1993;74:180–6.

29. Grzybowski S, Enarson D. [Results in pulmonary tuberculosis patients under various treatment program conditions]. Bulletin of the International Union against Tuberculosis. 1978;53:70–5.
30. Mexican Ministry of Health. Vital Statistics. [cited 2001 July 4] Available from: URL: http://www.ssa.gob.mx/unidades/dgied/sns/vitales/cuadro3.htm
31. Bustamante-Montes LP, Escobar-Mesa A, Borja-Aburto VH, Gomez-Muñoz A, Becerra-Posada F. Predictors of death from pulmonary tuberculosis: the case of Veracruz, Mexico. Int J Tuberc Lung Dis 2000;4:208–15.
32. Lawn SD, Acheampong JW. Pulmonary tuberculosis in adults: factors associated with mortality at a Ghanaian teaching hospital. West Afr J Med 1999;18:270–4.
33. Burman WJ, Cohn DL, Rietmeijer CA, Judson FN, Sbarbaro JA, Reves RR. Noncompliance with directly observed therapy for tuberculosis. Epidemiology and effect on the outcome of treatment. Chest 1997;111:1168–73.
34. Espinal MA, Kim SJ, Suarez PG, Kam KM, Khomenko AG, Migliori GB, et al. Standard short-course chemotherapy for drug-resistant tuberculosis: treatment outcomes in 6 countries. JAMA 2000;283:2537–45.
35. Espinal MA, Dye C, Raviglione M, Kochi A. Rational 'DOTS plus' for the control of MDR-TB. Int J Tuberc Lung Dis 1999;3:561–3.
36. Connolly C, Reid A, Davies G, Sturm W, McAdam KP, Wilkinson D. Relapse and mortality among HIV-infected and uninfected patients with tuberculosis successfully treated with twice weekly directly observed therapy in rural South Africa. AIDS 1999;13:1543–7.
37. Perriens JH, St. Louis ME, Mukadi YB, Brown C, Prignot J, Pouthier F, et al. Pulmonary tuberculosis in HIV-infected patients in Zaire. A controlled trial of treatment for either 6 or 12 months. N Engl J Med 1995;332:779–84.
38. Lucas SB, Hounnou A, Peacock C, Beaumel D, Djomand G, N´gbichi JM, et al. The mortality and pathology of HIV infection in a West African city. AIDS 1993;7:1569–79.
39. Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am J Epidemiol 1999;149:366–71.
40. Heldal E, Naalsund A, Kongerud J, Tverdal A, Boe J. Deaths from active tuberculosis: can we rely on notification and mortality figures? Tuber Lung Dis 1996;77:215–21.
41. Kahn K, Tollman SM, Garenne M, Gear JS. Who dies from what? Determining cause of death in South Africa's rural north-east. Trop Med Int Health 1999;4:433–41.

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the Centers for Disease Control and Prevention or the institutions with which the authors are affiliated.

Address for correspondence: Ma. de Lourdes García García, Unidad de tuberculosis, Instituto Nacional de Salud Pública Ave., Universidad No. 655, Col. Sta. María Ahuacatitlán, Cuernavaca, México, 62508; fax: 52777 317 55 29; e-mail: garcigar@correo.insp.mx

Antique poster (detail) from the collection of Joel and Jahan Montague. Joel Montague is Board Chairman of Partners for Development, a nonprofit organization. Jahan Montague is associate professor of nephrology, University of Massachusetts Medical School in Worcester.

RESEARCH

Molecular Analysis of Sarcoidosis Tissues for *Mycobacterium* Species DNA

Wonder Puryear Drake,* Zhiheng Pei,† David T. Pride,† Robert D. Collins,* Timothy L. Cover,* and Martin J. Blaser†

We performed polymerase chain reaction analysis, for *Mycobacterium* species 16S rRNA, *rpoB*, and IS*6110* sequences, on 25 tissue specimens from patients with sarcoidosis and on 25 control tissue specimens consisting of mediastinal or cervical lymph nodes and lung biopsies. *Mycobacterium* species 16S rRNA sequences were amplified from 12 (48%) and *rpoB* sequences from 6 (24%) of the sarcoidosis specimens. In total, 16S rRNA or *rpoB* sequences were amplified from 15 sarcoidosis specimens (60%) but were not detected in any of the control tissues (p=0.00002, chi square). In three specimens, the sequences resembled *Mycobacterium* species other than *M. tuberculosis*. All specimens with sequences consistent with *M. tuberculosis* were negative for IS*6110*. We provide evidence that one of a variety of *Mycobacterium* species*,* especially organisms resembling *M. tuberculosis,* is found in most patients with sarcoidosis.

Sarcoidosis is a multisystem inflammatory disease that mainly affects lymph nodes and pulmonary tissues and is characterized by noncaseating granulomata in affected organs (1). Although the cause of sarcoidosis remains unknown, several microorganisms have been proposed as possible etiologic agents, including bacteria (*Borrelia burgdorferi, Proprionibacterium acnes,* and *Mycobacterium* species) and viruses (*Human herpesvirus* 8, Epstein-Barr virus, *Cytomegalovirus*, and Coxsackie B) (2). Metals (beryllium and zirconium), minerals (talc and clay), and organic substances (pine tree pollen) have also been proposed as etiologic agents (2). Efforts to identify an infectious agent for sarcoidosis using methods such as histologic staining and routine microbial culture have been unsuccessful.

Polymerase chain reaction (PCR) analysis for microbial DNA serves as an alternative method for identifying infectious agents. PCR was used to identify the etiologic agents of bacillary angiomatosis (*Bartonella henselae)* (3) and Whipple's disease (*Tropheryma whippelii)* (4). Because of the substantial pathologic (5), immunologic (6), epidemiologic (7), and clinical similarities (8,9) between sarcoidosis and infections caused by *Mycobacterium* species (particularly tuberculosis), we analyzed tissue specimens from patients with sarcoidosis for evidence of mycobacterial genes. The results of previous studies have been inconclusive; some investigators were unable to demonstrate mycobacterial DNA in sarcoid lesions (10,11), whereas others have amplified mycobacterial DNA of different species (12,13). We examined sarcoidosis and control paraffin-embedded pulmonary, mediastinal, and cervical tissue specimens for *Mycobacterium* species 16S rRNA, *rpoB*, and IS*6110* sequences.

*Vanderbilt University School of Medicine and Veterans Affairs Medical Center, Nashville, Tennessee, USA; and †New York University School of Medicine and Veterans Affairs Medical Center, New York, New York, USA

Materials and Methods

Patients and Samples

For this study, we selected paraffin-embedded tissue specimens from patients who had had mediastinal or cervical lymph node resection from 1991 to 2001. Specimens from 44 patients with sarcoidosis and 57 controls were included. Patients were included for further study if they met the pathologic and clinical features described and if the specimens, after processing and DNA extraction, were positive for human β-actin with PCR analysis. We were unable to retrieve purified protein derivative status on a systematic basis**.** Based on these criteria, 25 control and 25 sarcoidosis specimens were further analyzed.

For inclusion in this study, the following criteria was used for patients with sarcoidosis: 1) clinical features had to be consistent with sarcoidosis (i.e., acute respiratory illness accompanied by erythema nodosum, hilar adenopathy and arthritis [Lofgren's syndrome], or indolent progressive pulmonary decompensation associated with radiographic findings, such as hilar adenopathy, reticulonodular infiltrates, or pulmonary fibrosis); 2) histopathologic features had to be consistent with sarcoidosis (i.e., specimens from each patient had to have confluent noncaseating granulomas, well circumscribed within the surrounding tissue with a variable amount of peripheral lymphocytic infiltration [5]); 3) known microbial causes for granulomata had to be excluded (i.e., specimens were negative for microorganisms by hematoxylin and eosin (H&E), fungal, acid fast bacilli (AFB), and auramine-O stains and on routine bacterial, fungal, and AFB cultures). In each case, histopathologic specimens were independently reviewed by two pathologists.

Control lymph node specimens were selected from patients who had undergone mediastinoscopy or cervical node biopsy during the same period. In each case, a definitive diagnosis other than sarcoidosis was made. Control patients were selected from patients for whom the final diagnoses were fungal infection, lymphoma, and primary or metastatic lung malignancies (Table 1).

DNA Extraction

For each patient enrolled in the study, the original paraffin-embedded tissue block was retrieved from the archives, and eleven 10-µm sections were cut from each. One section was stained with H&E for microscopic examination, four sections were used for extraction of DNA, and the remaining six sections were stored for future analysis. The specimens were randomly processed for slide preparation, and the microtome blade was changed between each tissue block. For each section from patients with sarcoidosis and for the control specimens with the granulomata, granulomata were microdissected and extracted by using disposable surgical blades. For those control specimens without granulomata, all tissue from the four sections was used for DNA extraction. For all specimens, DNA was extracted with the Qiagen DNAeasy extraction kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, except that 60 µL of proteinase K was used at a concentration of 20 mg/mL. Tissue dissection and DNA preparation were performed in a dedicated clean room, which was separate from the rooms used for PCR analysis and sequencing. The extracted DNA was stored at –20°C. Groups of tissue specimens from patients with sarcoidosis and controls were processed in parallel during all steps of the procedure, including extraction of the DNA, amplification and detection of mycobacterial DNA, and sequence analysis.

PCR Analysis for 16S rRNA, *rpoB*, and IS*6110*

Before PCR amplification, to ensure that the extracted DNA was of proper quality, we used PCR to verify that DNA sequences encoding human β-actin could be amplified. The primers used were 5´ATCATGTTTGAGACCTTCAAC3´ (forward primer) and 5´CAGGAAGGAAGGCTGGAAGAG3´ (reverse primer). The PCR conditions were 35 cycles of amplification carried out in a DNA Thermal Cycler 480 (Perkin-Elmer, Wellesley, MA); each cycle consisted of 1 min of denaturing at 94°C, 1 min of annealing at 54°C, and 1 min of extension at 72°C (14). As required, all 50 tissue specimens yielded human β-actin amplicons and were tested further for the presence of bacterial DNA.

For amplification of 16S rRNA sequences, a nested PCR analysis was performed. The primers FO16S, 5´GATAA GCCTGGGAAACTGGGTC3´ and RO16S, 5´TTCTCCAC-CTACCG TCAATCCG3´ were selected to amplify a 344-bp fragment of the 5´ region (nt 134–477) of mycobacterial 16S rRNA. Primers FI16S (5´CATGTCTTGTGGTGGAAAG CG3´) and RI16S (5´TACCGTC AATCCGAGAGAACCC3´) were selected as nested primers to amplify a 288-bp fragment (nt 181–468). The PCR conditions for both sets of primers were as follows: 5 min of denaturing at 94°C, followed by 35 cycles of amplification, consisting of 1 min of denaturing at 94°C, 1 min of annealing at 58°C, and 1 min of extension at 72°C. At the end of the 35 cycles, a final extension cycle of 7 minutes at 72°C was performed.

For amplification of *rpoB* sequences, a nested PCR also was performed. The primers FOrpoB (5´GCAGACGC TGTTGGAAAACTTG3´) and ROrpoB, (5´TGTTCTGGTCC ATGAATTGGCTC3´) were selected to amplify a 455-bp fragment of the β subunit (nt 1,940–2,394) of the *M. tuberculosis* RNA polymerase gene. The inner primers were designed as previously described and used in a nested fashion (nt 1,965–2,324) to amplify a 360-bp product (15). The PCR conditions were as described previously for 16S rRNA.

For amplification of IS*6110* sequences, PCR analysis included the use of primers IS*1* (5´CCTGCGAGCGTAGGCG TCGG3´) and IS*2* (5´CTCGTCCAGCGCCGCTTCGG3´), designed to amplify a 123-bp fragment (nt 1,510–1,632) of the *M. tuberculosis* IS*6110* element (16). The assay used the same conditions as described previously, with the exception that the PCR analysis included 30 rather than 25 cycles.

Negative and positive controls were run in parallel with each PCR assay. We used genomic DNA extracted from *M. tuberculosis* strain H37rv as positive controls, and DNA extracted from a paraffin-embedded tissue biopsy from an AIDS patient with ileocecal tuberculosis as positive controls. We included the following as negative controls in each PCR reaction: DNA extracted from nonsarcoid paraffin-embedded tissue, PCR master mix inoculated with 5 µL of sterile water, and PCR master mix alone. The DNA extraction was performed in the same manner as described for the sarcoid and control specimens.

Determination of DNA Sequence of Amplified Products

The PCR products were purified with the Qiagen QIAquick PCR purification kit (Qiagen, Valencia, CA) and sequenced directly on both strands in the Vanderbilt Cancer Center Core Sequencing Laboratory. In cases in which the signal was ambiguous, PCR products were cloned into the plasmid vector system, pGEM T-Easy (Promega, Madison, WI), and the nucleotide sequences were then determined.

Alignments of the 16S rRNA, *rpoB*, and IS*6110* sequences were performed with the NCBI BLAST program. Statistical evaluation of significance was determined by using chi-square analysis or Fisher's exact test, depending upon anticipated cell size. Sequences were aligned with ClustalW and subjected to phylogenetic analysis with HKY85 distance matrices with Paup 4.0b8 (Sinauer Associates, Sunderland, MA).

Results

Patient and Specimen Characteristics

Of the 25 patients with sarcoidosis, 12% were African American and 88% Caucasian; 36% were men, and 64% were

Table 1. Demographic and pathologic information for sarcoidosis and control tissue specimens[a]

Patient	Age (yrs)/sex	Race	Source of specimen	Pathologic diagnosis	Presence of granuloma
1	61/M	C	Lung	Sarcoidosis	Yes
2	60/F	C	Cervical node	Sarcoidosis	Yes
3	43/F	C	Mediastinum	Sarcoidosis	Yes
4	55/F	C	Mediastinum	Sarcoidosis	Yes
5	42/M	AA	Mediastinum	Sarcoidosis	Yes
6	54/F	C	Mediastinum	Sarcoidosis	Yes
7	48/F	C	Mediastinum	Sarcoidosis	Yes
8	34/M	C	Mediastinum	Sarcoidosis	Yes
9	72/F	AA	Mediastinum	Sarcoidosis	Yes
10	42/F	C	Mediastinum	Sarcoidosis	Yes
11	42/F	C	Mediastinum	Sarcoidosis	Yes
12	68/F	C	Mediastinum	Sarcoidosis	Yes
13	37/F	C	Mediastinum	Sarcoidosis	Yes
14	45/F	C	Mediastinum	Sarcoidosis	Yes
15	46/F	C	Mediastinum	Sarcoidosis	Yes
16	38/M	C	Mediastinum	Sarcoidosis	Yes
17	33/F	C	Mediastinum	Sarcoidosis	Yes
18	26/M	C	Mediastinum	Sarcoidosis	Yes
19	55/F	AA	Mediastinum	Sarcoidosis	Yes
20	31/M	C	Mediastinum	Sarcoidosis	Yes
21	42/M	C	Lung	Sarcoidosis	Yes
22	42/F	C	Mediastinum	Sarcoidosis	Yes
23	38/M	C	Mediastinum	Sarcoidosis	Yes
24	54/F	C	Mediastinum	Sarcoidosis	Yes
25	78/M	C	Mediastinum	Sarcoidosis	Yes
26	33/M	AA	Lung	Hodgkin’s disease	No
27	70/F	C	Lung	Histoplasmosis	Yes
28	73/M	C	Mediastinum	Mesothelioma	No
29	56/F	C	Mediastinum	Adenocarcinoma	No
30	24/M	C	Lung	Cryptococcus	Yes
31	75/M	C	Lung	Renal cell cancer	No
32	32/M	AA	Mediastinum	Coccidiomycosis	Yes
33	41/F	C	Mediastinum	Breast cancer	No
34	74/F	C	Mediastinum	Adenocarcinoma	No
35	72/F	C	Mediastinum	Large-cell cancer	No
36	77/M	C	Mediastinum	Large-cell cancer	No
37	72/M	AA	Mediastinum	Large-cell cancer	Yes
38	78/F	C	Mediastinum	Adenocarcinoma	No
39	72/M	C	Mediastinum	Squamous cell cancer	No
40	52/F	AA	Mediastinum	Adenocarcinoma	Yes
41	52/F	C	Mediastinum	Breast cancer	Yes
42	73/M	C	Mediastinum	Adenocarcinoma	Yes

Table 1, continued. Demographic and pathologic information for sarcoidosis and control tissue specimens[a]

Patient	Age (yrs)/sex	Race	Source of specimen	Pathologic diagnosis	Presence of granuloma
43	18/M	AA	Mediastinum	Histoplasmosis	No
44	47/M	C	Mediastinum	Hodgkin's lymphoma	No
45	74/M	C	Mediastinum	Large-cell lymphoma	No
46	76/M	C	Mediastinum	Adenocarcinoma	No
47	75/M	C	Mediastinum	Small-cell cancer	No
48	60/M	C	Mediastinum	Adenocarcinoma	No
49	73/M	C	Mediastinum	Adenocarcinoma	No
50	85/M	C	Mediastinum	Lymphoma	No
51	40/M	AA	Ileum	Tuberculosis	No

[a]F, female, M, male; C, Caucasian; AA, African-American.

<50 years of age. No specimens were obtained from persons <18 years of age (Table 1). The control population was 20% African American and 80% Caucasian; 68% were men. Most (76%) of the control patients were >50 years of age; the age and sex of the control patients reflect the patient population undergoing mediastinoscopy to obtain a tissue diagnosis for probable malignancy. The control population consisted of patients with lung cancer (72%), chronic fungal infections (16%), or lymphoma (12%) (Table 1). Mediastinal lymph nodes were the source of specimens from 88% and 84% of the sarcoid and control patients, respectively. The remaining specimens from each group were obtained from either pulmonary or cervical lymph node biopsies. Granulomas were present in all of the sarcoidosis tissue specimens and in 7 of the 25 control specimens.

PCR Assay Sensitivities

The sensitivity of the PCR assay for each gene was determined by PCR analysis of serially diluted genomic DNA from *M. tuberculosis* strain H37rv, ranging from 5 ng to 0.05 fg per µL. One *M. tuberculosis* genome is estimated to have a mass of 5 fg (17). For PCR analyses of 16S rRNA, *rpoB,* and IS*6110*, we consistently achieved a sensitivity of 1–2 gene copies in each assay.

16S rRNA PCR of Tissue Specimens

In the PCR for 16S rRNA sequences (Table 2), 12 (48%) of the 25 sarcoidosis specimens tested positive compared to none of 25 of the control specimens (p=0.0003, chi square). Sequence analysis of the PCR products from the sarcoidosis specimens showed that 8 of the 12 had 100% positional identity with *M. tuberculosis*, and 1 possessed 99% positional identity with *M. tuberculosis* (patient 15). Sequencing of the 16S PCR product of patient 15 showed a C→T substitution at position 289 and an A→G substitution at position 355 (based on the *M. tuberculosis* 16S rRNA sequence, GenBank accession nos. Z83862.1, AJ131120.1, X52917.1, and X58890.1). Three other sequences were found (in patients 7, 19, and 24) that most closely resembled other *Mycobacterium* species*.* The amplicon sequence from patient 7 possessed an A→G substitution at position 299 and a C→A substitution at position 380, yielding 99% positional identity with *M. kansasii.* Notably the sequences of *M. kansasii, M. avium, M. visibilis,* and *M. paratuberculosis* are identical within this region; therefore, distinguishing between these species is not possible (18). The amplicon sequence from patient 19 contained a T→C substitution at position 434, yielding 99% positional identity with *M. gordonae.* The amplicon sequence from patient 24 contained 100% positional identity with *M. gordonae* and *M. bohemicum* (19). The phylogenetic relationships of the mycobacterial sequences are shown in the Figure and are deposited in GenBank (accession nos. AF468214, AF468215, and AF468216).

PCR of Tissue Specimens with Other Mycobacterial Primers

With the *rpoB* PCR, 6 (24%) of the 25 specimens from sarcoidosis patients yielded a product of 360 bp, which by sequence analysis in each case was most consistent with *M. tuberculosis.* Five sequences amplified from sarcoidosis specimens had 98%–100% positional identity with *M. tuberculosis* (patients 1, 3, 8, 14, and 24), whereas one had 95% positional identity (patient 16) (Table 2). The difference in the rate of finding *M. tuberculosis rpoB* sequences (24% in the sarcoidosis specimens and none in the control specimens) was also significant (p=0.02, Fisher's exact test). In total, 15 (60%) of the sarcoidosis specimens had either *Mycobacterium* 16S rRNA or *rpoB* sequences compared with none of the control specimens (p=0.00002, chi square). Mycobacterial 16S rRNA and *rpoB* fragments were amplified from 3 (12%) of the 25 sarcoidosis specimens (patients 1, 14, and 24). Patients 1 and 14 possessed sequences that had 100% positional identity with *M. tuberculosis* 16S rRNA and 98%–99% positional identity with *M. tuberculosis rpoB* DNA. The products amplified from patient 24 possessed 100% positional identity with *M. gordonae* 16S rRNA and 99% positional identity with *M. tuberculosis rpoB* DNA. In the region amplified with our 16S primers, a difference of 14 nt existed between the 16S rRNA of *M. tuberculosis* and *M. gordonae.* In the region amplified by using *rpoB* primers, a difference of 39 nt existed between *rpoB* of *M. tuberculosis* and *M. gordonae.*

Table 2. Analysis of sarcoidosis and control tissue specimens for *Mycobacterium* 16SrRNA, *rpoB*, and *IS6110*[a]

Sarcoidosis patient	16S rRNA	rpoB	IS*6110*	Control patient	16S rRNA	rpoB	IS*6110*
1	*M. tuberculosis*	*M. tuberculosis*	--	26	-	-	-
2	*M. tuberculosis*	-	-	27	-	-	-
3	-	*M. tuberculosis*	-	28	-	-	-
4	*M. tuberculosis*	-	-	29	-	-	-
5	*M. tuberculosis*	-	-	30	-	-	-
6	-	-	-	31	-	-	-
7	*M. kansasii*	-	-	32	-	-	-
8	-	*M. tuberculosis*	-	33	-	-	-
9	*M. tuberculosis*	-	-	34	-	-	-
10	-	-	-	35	-	-	-
11	*M. tuberculosis*	-	-	36	-	-	-
12	-	-	-	37	-	-	-
13	-	-	-	38	-	-	-
14	*M. tuberculosis*	*M. tuberculosis*	-	39	-	-	-
15	NM	-	-	40	-	-	-
16	-	*M. tuberculosis*	-	41	-	-	-
17	-	-	-	42	-	-	-
18	*M. tuberculosis*	-	-	43	-	-	-
19	*M. gordonae*	-	-	44	-	-	-
20	-	-	-	45	-	-	-
21	-	-	-	46	-	-	-
22	-	-	-	47	-	-	-
23	-	-	-	48	-	-	-
24	*M. gordonae*	*M. tuberculosis*	-	49	-	-	-
25	-	-	-	50	-	-	-
				51	*M. tuberculosis*	*M. tuberculosis*	*M. tuberculosis*
				H37rv	*M. tuberculosis*	*M. tuberculosis*	*M. tuberculosis*

[a]NM, novel mycobacterium, resembling *M. tuberculosis*; H37rv, *M. tuberculosis*; -, indicates a negative result.

None of the sarcoidosis or control patient tissue specimens yielded IS*6110* amplicons (Table 2). In contrast, IS*6110* products were consistently found in the positive controls, genomic DNA from *M. tuberculosis* H37rv, and DNA extracted from a paraffin-embedded tissue biopsy of an AIDS patient with ileocecal tuberculosis. Both positive controls were positive in the 16S rRNA, *rpoB*, and IS*6110* PCR assays, and sequence analysis of the products indicated 100% homology with *M. tuberculosis.* The negative controls were consistently negative.

Discussion

For this study, we chose patients whose cases were consistent with sarcoidosis or in whom an alternative diagnosis was made conclusively. We chose this stringent design so that no borderline tissues were examined. Many cases of disease eventually diagnosed as sarcoidosis have atypical findings. If our present observations are confirmed, such cases will be important for future analyses.

We found evidence of mycobacterial DNA in the granulomas of 24% of sarcoidosis specimens when assessing for *rpoB,* 48% in the same population when assessing for 16S rRNA; in total, 60% were positive for either. We acknowledge the limitations of studying archival tissue and the possibility of contamination; however, control tissues did not demonstrate positive results, making contamination less likely. Earlier studies have identified the presence of mycobacterial DNA in sarcoidosis tissue specimens with 30%–50% prevalence (12,13,20). Instead of a single organism being present, we provide evidence for a heterogenous population of *Mycobacterium* species in the sarcoidosis tissue specimens studied. Although we found evidence of organisms resembling *M. tuberculosis, M. gordonae*, and *M. kansasii,* other studies also have identified *M. avium* sequences (12,13).

We also provide DNA sequence evidence for novel mycobacteria in patient 15. Although most DNA sequences from the study patients most closely resemble *M. tuberculosis,*

sequences resembling other mycobacterial species also were identified (Table 2, Figure). In several previous studies, non-tuberculosis mycobacteria also have been reported (13). One novel sequence is most closely related to 16S rDNA from *M. tuberculosis,* a known pulmonary pathogen, rather than to sequences from other mycobacterial species of lesser virulence. The consistent presence of two single polymorphisms in the same location in the novel sequence suggests a true polymorphism rather than an error introduced by Taq PCR. Moreover, the novel sequence was consistently absent from water, non-sarcoid paraffin-embedded tissue, and *M. tuberculosis* DNA controls. Synonymous substitutions in the *M. tuberculosis* genome are relatively rare, although genomic variations have been found in genes associated with antibiotic resistance (21). The DNA with the polymorphism suggests that a variant of *M. tuberculosis*, or a closely related novel mycobacterium, may be present in the sarcoidosis specimen.

The presence of *M. tuberculosis* DNA in 48% of sarcoidosis specimens is notable because clear clinical connections between sarcoidosis and tuberculosis have been made. On occasion, patients with documented tuberculosis develop sarcoidosis while on antituberculous treatment or vice versa (22–24). Mycobacterial DNA in sarcoidosis specimens may explain the clinical correlation between sarcoidosis and tuberculosis. That patients have developed sarcoidosis while on antituberculous therapy suggests that in those patients *M. tuberculosis* was not the etiologic agent of sarcoidosis. That 60% of the specimens we examined showed mycobacterial DNA agrees with certain previous studies (12,13), but other studies were negative for mycobacterial DNA (10,11). One possible explanation for these discordant results is that sarcoidosis represents one group of host responses to infectious agents of which mycobacteria represent the largest associated group. Alternatively, *Mycobacterium* species are present in many of the lesions but at extremely low levels, on either side of the threshold of detection. Such a hypothesis of small numbers of organisms provoking an intense inflammatory response, analogous to tuberculoid leprosy (25), could explain why organisms cannot be detected except by ultrasensitive methods. Yet another alternative explanation was our observation of degradation of the mycobacterial signal in the total DNA extract. We observed that mycobacterial DNA could be amplified from the positive specimens consistently over a 6–8 month period if the original DNA extract was maintained at –20°C. After this time period, fresh DNA extractions were necessary to demonstrate the presence of mycobacterial DNA. The original specimens, in which the mycobacterial DNA could no longer be amplified, remained positive for human β-actin by PCR analysis, although the band was weaker, suggesting either that the eukaryotic DNA degraded more slowly than prokaryotic DNA or that more signal was originally present. This degradation occurred despite minimizing freeze-thaws of extracted DNA and maintaining the DNA at –20°C. Our observation suggests that isolation of mycobacterial DNA from sarcoidosis specimens is best achieved by performing PCR analysis on fresh DNA extractions, which may help explain why other investigators had negative findings.

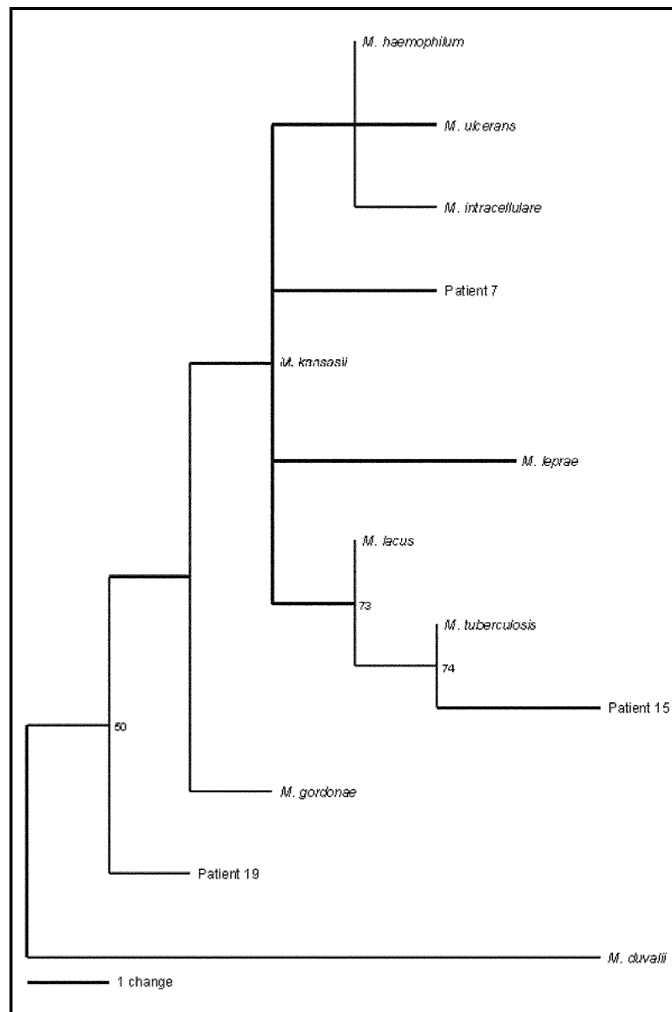

Figure. Analysis of 16S rRNA sequences from nine *Mycobacterium* species and from 12 patients with sarcoidosis. Phylograms based on nucleotide alignments were generated with HKY85 distances matrices using Paup 4.0b10 (Sinauer Associates, Sunderland MA). Bootstrap values ≥50 (based on 500 replicates) are represented at each node, and the branch length index is represented below the phylogram. For eight patients, the 16S rRNA sequence was identical to *M. tuberculosis*, and for one patient (patient 24) was identical to *M. gordonae*. For patient 15, the sequence was closely related to *M. tuberculosis*; for patient 19, the sequence was closely related to *M. duvalii* and *M. gordonae*; for patient 7, the sequence was less closely related to *M. leprae*.

Based on these observations, we examined whether *M. tuberculosis* DNA was present in the sarcoid granuloma by testing for the presence of IS*6110*. PCR analysis for IS*6110* is useful, since IS*6110* is typically present in 1–25 copies in members of the *M. tuberculosis* complex. *M. bovis* BCG has only a single of copy of IS*6110*, whereas the higher copy numbers are typically found in *M. tuberculosis* isolates (26). We found no evidence of IS*6110* DNA in our sarcoidosis or control tissue specimens.

Several possible explanations exist for the presence of mycobacterial 16S rRNA and *rpoB,* and the absence of IS6110 in the sarcoid specimens, although these three amplicons were

consistently present in our positive controls. First, our assay for IS*6110* may not have been sufficiently sensitive to detect the very low numbers of *M. tuberculosis* genomes in the sarcoidosis tissue specimens. In serial dilution studies, the assay was sensitive enough to detect one bacterial genome, comparable to results for the nested PCRs for 16S and *rpoB*. However, correlating the sensitivity of DNA extracted from bacterial culture to DNA extracted from formalin-fixed, paraffin-embedded tissue is not possible. Other laboratories that reported an assay sensitivity of 1–2 genome copies for IS*6110* in sarcoidosis tissue extract were also unable to detect any IS*6110* (11,27–29), which was consistent with our results. Studies assessing for IS*6110* reflect a substantial portion of the literature that does not support the presence of mycobacterial DNA in the sarcoidosis tissue specimens (11,27,28,30).

A second possibility is that *M. tuberculosis* is present but the strains do not contain IS*6110*, since strains that possess one copy or no copies of IS*6110* have been reported (31,32). In the United States, all of approximately 14,000 strains of *M. tuberculosis* tested have been shown to possess IS*6110*; some in low-copy number (33). Therefore, this scenario seems unlikely.

A third explanation is that while the agent we found associated with sarcoidosis has a close genetic relationship with *M. tuberculosis,* it is not *M. tuberculosis.* The genes for 16S and 23S are particularly suitable as targets for identifying microorganisms, since they are both well conserved and show variation indicative of their evolution and interrelationship with other organisms (34). This genetic variation is the basis for identifying the species of microorganisms in a particular genus, as this genetic variation is a constant property. Other members of the *M. tuberculosis*-complex (*M. tuberculosis* BCG, *M. bovis, M. microti*, and *M. africanum*) have 100% 16S and *rpoB* homology with *M. tuberculosis* but belong to different species; these strains are usually differentiated from *M. tuberculosis* by biochemical and clinical features. Although we could not attempt isolation of microorganisms from the formalin-fixed, paraffin-embedded specimens, future studies targeted to mycobacteria would be especially useful in confirming our observations and in characterizing any association with sarcoidosis.

We have found evidence for mycobacterial 16S rRNA and *rpoB* sequences in sarcoidosis tissue specimens but not in control tissue specimens. Upon sequence analysis, the products were most consistent with *M. tuberculosis*, but IS*6110* could not be detected from these species. We also provide evidence of the presence of a heterogeneous mycobacterial population, including organisms highly related to *M. tuberculosis, M. gordonae,* and *M kansasii*. This heterogeneous population was found in individual sarcoidosis samples and, in one case, in the same sample (patient 24). These findings suggest that while *M. tuberculosis* and other *Mycobacterium* species may not be the sole microbial agents present in sarcoidosis tissues, they are commonly present and may play important roles. Further investigation into their presence and any putative etiologic agent is warranted.

Acknowledgments

We thank David Relman for guidance throughout the project, Colorado State University (NIAID NO1 AI-75320) for supplying our laboratory with the *Mycobacterium tuberculosis* strain H37rv, and Edward McDonald for providing paraffin-embedded lymph node biopsy specimens.

Dr. Drake is supported in part by the Robert Wood Johnson Minority Medical Faculty Development Program, the Medical Research Service of the Department of Veterans Affairs, and by RO1 GM 63270 from the National Institutes of Health.

Dr. Drake is assistant professor of medicine at Vanderbilt University School of Medicine in the Department of Infectious Diseases. Her interests include the investigation of possible infectious etiologic agents for chronic inflammatory states, such as sarcoidosis, by using molecular assays.

References

1. Crystal RG, Bitterman PB, Rennard SI, Hance AJ, Keogh BA. Interstitial lung diseases of unknown cause: disorders characterized by chronic inflammation of the lower respiratory tract. N Engl J Med 1984;310:154–66.
2. American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of sarcoidosis and Other Granulomatous Disorders (WASOG). Statement on sarcoidosis. Am J Respir Crit Care Med 1999;160:736–55.
3. Relman DA, Loutit JS, Schmidt TM, Falkow S, Tompkins LS. The agent of bacillary angiomatosis. An approach to the identification of uncultured pathogens. N Engl J Med 1990;323:1573–80.
4. Relman DA, Schmidt TM, MacDermott RP, Falkow S. Identification of the uncultured bacillus of Whipple's disease. N Engl J Med 1992;327:293–301.
5. Scadding JG, Mitchell DN. Sarcoidosis. 2nd ed. London: Chapman and Hall; 1985. p 36–41.
6. Foley N, Lambert C, McNicol M, Johnson N, Rook G. An inhibitor of tumour necrosis factor in the serum of patients with sarcoidosis, tuberculosis and Crohn’s disease. Clin Exp Immunol 1990;80:395–9.
7. Brett G. Epidemiological trends in tuberculosis and sarcoidosis in a district of London between 1958 and 1963. Tubercle 1965;46:412–6.
8. Mitchell DN, Rees RJW. Some diseases of possible mycobacterial aetiology. In: Ratledge C, Stanford J, editors. The biology of mycobacteria. London: Academic Press; 1982. p. 525–36.
9. Rook GA, Stanford JL. Slow bacterial infections or autoimmunity? Immunol Today 1992;13:160–4.
10. Richter E, Greinert U, Kirsten D. Assessment of mycobacterial DNA in cells and tissues of mycobacterial and sarcoid lesions. Am J Respir Crit Care Med 1996;153:375–80.
11. Bocart D, Lecossier D, de Lassence A, Valeyre D, Battesti JP, Hance A J. A search for mycobacterial DNA in granulomatous tissues from patients with sarcoidosis using the polymerase chain reaction. Am Rev Respir Dis 1992;145:1142–8.
12. Popper HH, Klemen H, Hoefler G, Winter E. Presence of mycobacterial DNA in sarcoidosis. Hum Pathol 1997;28:796–800.
13. Li N, Bajoghi A, Kubba A, Bhawan J. Identification of mycobacterial DNA in cutaneous lesions of sarcoidosis. J Cutan Pathol 1999;26:271–8.
14. Hall BL, Finn OJ. PCR-based analysis of the T-cell receptor V beta multigene family: experimental parameters affecting its validity. Biotechniques 1992;13:248–57.
15. Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, Chae GT, et al. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). J Clin Microbiol 1999;37:1714–20.
16. Eisenach KD, Cave MD, Bates JH, Crawford JT. Polymerase chain reaction amplification of a repetitive DNA sequence specific for *Mycobacterium tuberculosis*. J Infect Dis 1990;161:977–81.

17. Del Portillo P, Munillo LA, Patarroyo ME. Amplification of a species-specific DNA fragment of *Mycobacterium tuberculosis* and its possible use in diagnosis. J Clin Microbiol 1991;29:2163–8.
18. Edwards U, Rogall T, Blocker H, Emde M, Bottger EC. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. Nucleic Acid Res 1989;17:7843–53.
19. Rogall T, Wolters J, Flohr T, Bottger EC. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int J Syst Bacteriol 1990;40:323–30.
20. El-Zaatari FA, Naser SA, Markesich DC, Kalter DC, Engstand L, Graham DY. Identification *of Mycobacterium avium* complex in sarcoidosis. J Clin Microbiol 1996;34:2240–5.
21. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissimenation. Proc Natl Acad Sci U S A 1997;94:9869–74.
22. Wong CF, Yew WW, Wong PC, Lee J. A case of concomitant tuberculosis and sarcoidosis with mycobacterial DNA present in the sarcoid lesion. Chest 1998;114:626–9.
23. Emerson PA, Young FH. Sarcoidosis following tuberculosis. Tubercle 1956;4:116–9.
24. Kent DC, Houk VN, Elliott RC, Sokolowski JW, Baker JH, Sorensen K. The definitive evaluation of sarcoidosis. American Review of Respiratory Diseases 1970;101:721–7.
25. Sieling RA, Modlin RL. Cytokine patterns at the site of mycobacterial infection. Immunobiology 1994;191:378–87.
26. van Soolingen D, Hermans PWM, De Haas PEW, Soll DR, van Embden JDA. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 1991;29:2578–86.
27. Wilsher ML, Menzies RE, Croxson MC. *Mycobacterium tuberculosis* DNA in tissues affected by sarcoidosis. Thorax 1998;53:871–4.
28. Vokurka M, Lecossier D, du Bois RM, Wallaert B, Kambouchner M, Tazi A, et al. Absence of DNA from mycobacteria of the *M. tuberculosis* complex in sarcoidosis. Am J Respir Crit Care Med 1997;156:1000–3.
29. Klemen H, Husain AN, Cagle PT, Garrity ER, Popper HH. Mycobacterial DNA in recurrent sarcoidosis in the transplanted lung—a PCR-based study on four cases. Virchows Arch 2000;436:365–9.
30. Chao SC, Yan JJ, Lee JY. Cutaneous sarcoidosis among Taiwanese. J Formosan Med Assoc 2000;99:317–23.
31. Se Thoe SY, Tay L, Sng EH. Evaluation of amplicon- and IS6110-PCR for direct detection of *Mycobacterium tuberculosis* complex in Singapore. Trop Med Int Health 1997;2:1095–101.
32. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol 1993;31:1987–95.
33. Bifani PJ, Shopsin B, Alcabes P, Mathema B, Kreiswirth BN, Liu Z, et al. Molecular epidemiology and tuberculosis control. JAMA 2000;284:305–7.
34. Pace NR. A molecular view of microbial diversity and the biosphere. Science 1997;276:734–40.

Address for correspondence: Wonder Drake, Division of Infectious Diseases, Vanderbilt University Medical Center, A-3310 MCN, Nashville, TN 37232, USA; fax: 615-343-6160; e-mail: wonder.drake@mcmail.vanderbilt.edu

SEARCH
EID
ONLINE
www.cdc.gov/eid

Genomewide Pattern of Synonymous Nucleotide Substitution in Two Complete Genomes of *Mycobacterium tuberculosis*

Austin L. Hughes,* Robert Friedman,* and Megan Murray†

Comparison of the pattern of synonymous nucleotide substitution between two complete genomes of *Mycobacterium tuberculosis* at 3,298 putatively orthologous loci showed a mean percent difference per synonymous site of 0.000328 ± 0.000022. Although 80.5% of loci showed no synonymous or nonsynonymous nucleotide differences, the level of polymorphism observed at other loci was greater than suggested by previous studies of a small number of loci. This level of nucleotide difference leads to the conservative estimate that the common ancestor of these two genotypes occurred approximately 35,000 ago, which is twice as high as some recent estimates of the time of origin of this species. Our results suggest that a large number of loci should be examined for an accurate assessment of the level of nucleotide diversity in natural populations of pathogenic microorganisms.

Surveys of genetic diversity in the pathogenic bacterium *Mycobacterium tuberculosis* have revealed a contradictory picture. In spite of known polymorphism at the phenotypic level and abundant polymorphism associated with repetitive elements (1), surveys of single nucleotide polymorphism in protein-coding genes have shown surprisingly low levels of polymorphism in comparison with other eubacterial species (2). The apparent low level of nucleotide polymorphism has led to the hypothesis that the ancestor of this species occurred quite recently, perhaps 15,000–20,000 years ago (2,3). However, if the number of substitutions per site is low, the error of estimation of this number would be expected to be substantial unless a very large number of sites are surveyed. We addressed the question of polymorphism in *M. tuberculosis* by comparing protein-coding genes in two completely sequenced genotypes, H37Rv and CDC1551 (4).

*University of South Carolina, Columbia, South Carolina, USA; and †Harvard School of Public Health, Boston, Massachusetts, USA

Methods

We applied the BLASTP program (5) to identify, for each predicted protein sequence in the H37Rv genome (GenBank accession no. AL123456), the closest homolog in the CDC1551 genome (GenBank accession no. AE000516). Following GenBank annotations, we compared 3,972 predicted proteins in H37Rv with 4,187 predicted proteins in CDC1551. We used a strict search criterion ($E = 10^{-50}$) to identify truly orthologous gene pairs. We aligned (6) the putative orthologous pairs of amino acid sequences (n=3,428), then imposed this alignment on the DNA sequences.

Visual inspection of amino acid alignments showed that certain alignments, usually near the N-terminus or C-terminus, had regions of very low sequence identity. Examination of the DNA sequences of the corresponding genes showed that these regions of low identity were typically caused by a frameshift in one of the two genomes relative to the other. Whether these frameshifts are biologically real or result from sequencing error was uncertain; therefore, we eliminated 119 such gene pairs from our data set. For the remaining gene pairs (n=3,309), we computed the proportion of synonymous substitutions per synonymous site (p_S) and the proportion of nonsynonymous substitutions per site (p_N) by using Nei and Gojobori's method (7). Because values of p_S and p_N were very low in most cases, we did not correct for multiple hits.

Because p_S values appeared to fall into two groups (see Results), we used a simple probabilistic model to separate these two sets of gene pairs. We assumed that the probability of synonymous substitution followed two separate binomial distributions, designated models A and B, with probabilities of "success" (i.e., of a synonymous difference) designated p_A and p_B, respectively. Using the Bayes equation, for each gene pair with a given p_S value, we computed the probability that model A applies, given the observed p_S: $P(A|p_S) = (p_{SA}) f_A / [(p_{SA}) f_A + (p_{SB}) f_B]$, where f_A is the frequency of cases to which model A applies, f_B the frequency of cases to which model B applies, p_{SA} is the binomial probability of obtaining the observed p_S, given the number of synonymous sites in the gene and a probability of a synonymous difference equal to p_A; and p_{SB} is the binomial probability of obtaining the observed p_S, given the number of synonymous sites in the gene and a probability of a synonymous difference equal to p_B. The probability that model B applies, given p_S, is $P(B|p_S) = (p_{SB}) f_B / [(p_{SA}) f_A + (p_{SB}) f_B]$

By comparing these two probabilities, we assigned each gene pair to one of two groups assumed to evolve according to the two models, respectively (Groups A and B). We reassigned group membership in iterative fashion, computing p_A and p_B from the mean p_S values for each group. We started the process with $f_A = 0.995$ and continued until group memberships were stable.

Results

Of 3,309 pairs of putatively homologous protein-coding genes in the H37Rv and CDC1551 genomes of *M. tuberculosis*, 2,662 (80.5%) showed no synonymous or nonsynonymous

nucleotide differences between the two genomes, and 3,010 (91.0%) showed no synonymous differences between the two genomes. However, in a small number of gene pairs, the proportion of synonymous differences per synonymous site (p_S) was surprisingly high. In 13 (0.4%) gene pairs, p_S was >0.01, and in 3 gene pairs p_S exceeded 4%. These extreme p_S values seen in a small number of gene pairs are much higher than generally observed between alleles at neutrally evolving loci in eukaryotes (8). Thus, the comparison of protein-coding genes between the two *M. tuberculosis* genomes suggested the existence of two distinct groups of gene pairs: a large group having few or no synonymous differences and a much smaller group with a substantial degree of synonymous divergence.

We used a simple probabilistic model (see Methods) to separate these two sets of gene pairs, designated Group A and Group B, respectively (Figure). The application of this method showed 11 loci with unusually high p_S values and probabilities of assignment to group A of <50% (Figure).

We assumed that Group A members are truly orthologous gene pairs that diverged at the time of the common ancestor of the H37Rv and CDC1551 genomes. Group A included 3,298 pairs, with mean p_S for all genes of 0.000328 ± 0.000022 standard error. When p_S was estimated for the 3,298 genes concatenated together (a total of 934,413 synonymous sites), an estimate of p_S = 0.000348 ± 0.00019 was obtained. The range of p_S values in Group A was between zero and 0.012; a total of 288 loci in Group A had p_S values other than zero. These results show a substantial level of nucleotide diversity, approximately half the level of nucleotide diversity in humans (9).

Rates of nucleotide substitution per unit time are difficult to estimate in bacteria given the lack of calibration from the fossil record (10). To obtain an estimate of the rate of synonymous nucleotide substitution, we used published data on comparisons of *Escherichia coli* and *Salmonella typhimurium* (11,12), which are believed to have diverged approximately 100 million years ago (13,14) (Table 1). This procedure yielded estimates for the last common ancestor of H37Rv and CDC1551 in the range of 34,000–38,000 years (Table 1). These estimates are approximately twice previous estimates of the age of the common ancestor of worldwide *M. tuberculosis* (2,3). To obtain the observed mean p_S value between H37Rv and CDC1551 within 15,000–20,000 years would require a rate of synonymous substitution approximately twice that observed in *Enterobacteria*.

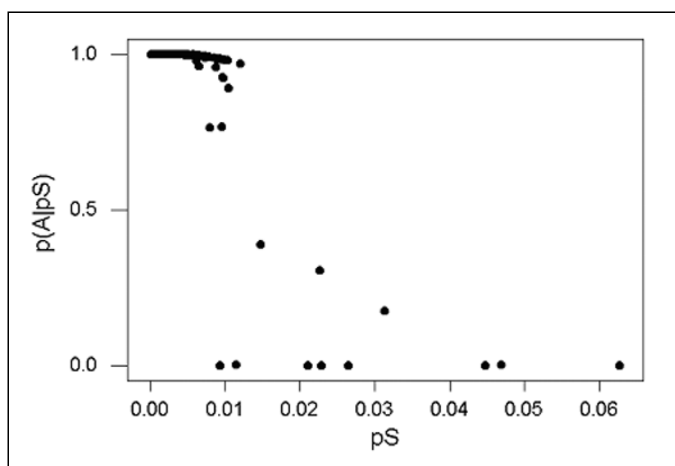

Figure. A plot of p(A|p_S), the probability of assignment of a locus to group A given the observed p_S value, as a function of p_S at 3,309 loci compared between the H37Rv and CDC1551 genotypes of *Mycobacterium tuberculosis*. The plot shows the bimodal nature of the distribution of p_S values, with overall higher values of p_S at the 11 loci having p(A|p_S) >50%.

Group B consisted of 11 gene pairs with mean p_S of 0.0286 ± 0.0050 (Table 2).

In *Enterobacteria*, a negative correlation exists between observed proportions of synonymous difference and codon bias (11). In the case of *Mycobacterium*, codon bias results mainly from the very high third position G+C content of most genes (15). In our data, however, we observed no correlation between p_S and proportion G+C at third codon positions (r = -0.010; not significant).

Table 1. Estimates[a] of the divergence time of the H37Rv and CDC1551 genotypes of *Mycobacterium tuberculosis*

Reference	No. loci	Synonymous substitutions/site/yr	Divergence time (H37Rv and CDC1551)
11	67	$4.7 \pm 0.2 \times 10^{-9}$	34,900 ± 2,300 [b] (33,500–36,400) [c]
12	128	4.4×10^{-9}	37,300 ± 2,500 [b]

[a]Based on synonymous substitutions between *Escherichia coli* and *Salmonella typhimurium*, assumed to have diverged 100 million years ago (13,14).
[b]Estimates are shown ± standard error, based on standard error of mean p_S.
[c]Range based on standard error of rate estimate.

Discussion

A number of additional possibilities may explain the occurrence of gene pairs with higher than expected p_S values: 1) Balanced polymorphism. Selectively maintained polymorphisms are expected to be much older than neutral polymorphisms and may even predate speciation events (16). In the case of haploid organisms such as bacteria, balancing selection would take the form of frequency-dependent selection rather than overdominant selection. 2) Differential deletion. In a multi-gene family, if one member of an orthologous pair of genes were deleted in one genotype, the gene pairs would involve paralogous, not orthologous comparisons. 3) Horizontal gene transfer. A gene obtained by one of the two genotypes from another bacterial species would be expected to be more divergent than other genes in that genotype.

One indication of a balanced polymorphism is a higher rate of nonsynonymous than synonymous substitution (8). There was no strong evidence of such selection in the present case; p_S was greater than p_N at 10 of the 11 loci, and p_N exceeded p_S only slightly at one locus (Table 2). In addition, we compared p_S and p_N in sliding windows of 30 codons along the length of these genes. No regions were observed in which p_N was greater than p_S (data not shown). Thus, there was no evidence of positive selection acting on specific regions of these genes. On the other hand, differential deletion can probably explain some

Table 2. Proteins for which the nearest homologous comparison between the H37Rv and CDC1551 genotypes of *Mycobacterium tuberculosis* has a high p_S value (Group B)

	Accession nos.	Protein function	p_S	p_N
Probable differential deletion				
	NP_216309, NP_335079	unknown	0.0470	0.0000
	NP_215713, NP_335504	unknown	0.0628	0.0043
	NP_216319, NP_336310	PE repeat family	0.0115	0.0094
	NP_215965, NP_335949	PE repeat family	0.0448	0.0185
Possible horizontal gene transfer				
	NP_214910, NP_334815	unknown	0.0226	0.0105
	NP_216104, NP_336077	unknown	0.0265	0.0084
	NP_216281, NP_336535	unknown	0.0210	0.0161
	NP_215835, NP_335809	adenylate cyclase	0.0229	0.0068
	NP_216564, NP_336573	polyketide synthase	0.0093	0.0036
	NP_217029, NP_337080	unknown	0.0148	0.0156
	NP_216862, NP_335679	unknown	0.0313	0.0000
Mean ± S.E.			0.0286 ± 0.0050	0.0085 ±, 0.0019[a]

[a]Paired sample t-test of the hypothesis that $p_S = p_N$, p<0.01. The quantities p_S and p_N are the proportion of nucleotide difference per synonymous site and per nonsynonymous site, respectively.

cases, most notably members of the PE multi-gene family (11) (Table 2). The remaining gene pairs are possibly cases of horizontal gene transfer (Table 2), for which there is some recent evidence in *M. tuberculosis* (17). Presumably a related species of *Mycobacterium* was the source of such gene transfers.

Our results did not support the hypothesis that the common ancestor of *M. tuberculosis* was relatively recent (2). Rather, the pattern of nucleotide substitution at synonymous sites suggested a divergence time for the two available genotypes of this species approximately 35,000 years ago. Since H37Rv and CDC1551 represent two genotypes sampled from within the species, they are probably not the most divergent genotypes possible. Thus, the last common ancestor of *M. tuberculosis* likely occurred considerably earlier than 35,000 years ago.

While the difference between an estimate of 15,000–20,000 years and one of 35,000 years is not large on an evolutionary time scale, such a difference is substantial on the scale of human history. For example, the existence of two genotypes in the current population of *M. tuberculosis* with a common ancestor at 35,000 years is evidence against the hypothesis that *M. tuberculosis* arose, presumably from *M. bovis*, at the time of human domestication of cattle (18). Our result is thus consistent with phylogenetic analyses based on insertion-deletion events, which suggest that the *M. tuberculosis* lineage was a human pathogen well before the origin of *M. bovis* (19). Thus, along with recent evidence of an ancient origin and extensive polymorphism in the malaria parasite *Plasmodium falciparum* (20,21), our study provides evidence against the long-held view that virulent pathogens are invariably evolutionarily recent (22).

Our estimate is conservative because the rate of synonymous substitution may actually be lower in *Mycobacterium* than in *Enterobacteriaceae*, given the highly skewed G+C content in the former. Furthermore, our estimate of the mean proportion of synonymous difference was conservative because we excluded 119 loci with potential frameshifts between the two genotypes as well as a set of 12 loci with unusually high p_S values. In addition to the 12 loci assigned to our Group B, certain other loci might also have originated from horizontal gene transfer. However, even if horizontal gene transfer has occurred at other loci besides those in Group B, eliminating further loci with relatively high p_S values from Group A will not affect the results greatly. For example, if we eliminate the 10 loci with highest p_S values from Group A, mean p_S will be reduced only to 0.000299, and the estimated age of the common ancestor will be barely affected.

The degree of polymorphism observed in this study is unlikely to have been substantially influenced by sequencing errors. The error rate for finished sequences from the Institute for Genomic Research (where CDC1551 was sequenced) has been independently estimated at <1 in 88,000 bases (23). Assuming a similar error rate for both 4.4 mega-bp *M. tuberculosis* genomes, we would expect to see approximately 100 differences between them due to sequencing errors. Approximately 21 such differences would be expected in the 938,778 synonymous sites in Group A and Group B genes. In fact, 411 synonymous differences were observed at these sites; thus, even if present, sequencing errors are likely to have made up only a small fraction (approximately 5%) of the total synonymous polymorphism. At such a rate, sequencing errors would have little effect on our estimates of nucleotide diversity at synonymous sites or the age of the common ancestor of the two genomes.

In addition, the hypothesis that the single nucleotide polymorphisms (SNPs) observed between these genotypes are real received strong support from a recent study that observed a

number of the same SNPs in clinical isolates (24). Moreover, since sequencing errors are expected to occur at random with respect to the reading frame of coding sequences, the fact that mean p_S exceeded mean p_N in both Group A and Group B was strong evidence against the hypothesis that a substantial proportion of the observed polymorphism was due to sequencing error.

Simple considerations of probability can explain why earlier studies produced relatively low estimates of this species' age. If we assume that the per-site probability of a synonymous difference between two *M. tuberculosis* genomes is equal to the mean p_S observed between H37Rv and CDC1551 (0.000328), then the probability is approximately 95% that no synonymous differences will be seen in a gene with 150 synonymous sites. The probability that no synonymous differences will be seen in 10 such loci chosen at random is approximately 60%, and the probability that no synonymous differences will be observed at 20 such loci is approximately 37%. On the other hand, the probability that no synonymous differences will be seen at 100 such loci is <1%.

These calculations emphasize the need to examine a very large number of nucleotide sites to obtain a reliable estimate of nucleotide diversity and thus of the age of the most recent common ancestor in cases where the frequency of substitution is less than one in a thousand. Even when the frequency of substitution is between one in a thousand and one in a hundred, substantial stochastic error is possible if the number of loci examined is small. Thus, any study that estimates population parameters from nucleotide sequence data needs to survey a substantial number of loci. These considerations are particularly important in the case of pathogenic microorganisms, where a number of factors (including both natural selection and horizontal gene transfer) may lead to substantial differences among loci with respect to the level of nucleotide diversity.

Comparison of two complete genomes of *M. tuberculosis* showed a greater extent of sequence polymorphism than would be expected on the basis of previous studies, in turn suggesting that analysis of additional genomes will likely show further polymorphism. Polymorphism in any species of pathogen may complicate therapeutic strategies because it implies the existence of variation on which selection can act, including selection imposed by human vaccines and pharmacologic agents (20). On the other hand, known polymorphisms may prove useful to investigators in reconstructing the evolutionary relationships among clinical isolates and in providing markers for understanding the genetic basis of complex phenotypic traits.

This research was supported by National Institutes of Health grants GM34940 and GM66710 to A.L.H. and AI01430 and AI1046669 to M.M.

Dr. Hughes is director of the Biotechnology Institute and professor in the Department of Biological Sciences at the University of South Carolina. His research uses computational analysis of molecular sequence data to understand host-parasite co-evolution, genome evolution, and the population biology of human pathogens.

References

1. Kato-Maeda M, Bifani PJ, Kreiswirth BN, Small PN. The nature and consequence of genetic diversity within *Mycobacterium tuberculosis*. J Clin Invest 2001;107:533–7.
2. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent dissemination. Proc Natl Acad Sci U S A 1997;94:9869–74.
3. Kapur V, Whittam TS, Musser JM. Is *Mycobacterium tuberculosis* 15,000 years old? J Infect Dis 1994;170:1348–9.
4. Betts JC, Dodson P, Quan S, Lewis AP, Thomas PJ, Duncan K, et al. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. Microbiology 2000;146:3205–16.
5. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
6. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–80.
7. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986;3:418–26.
8. Hughes AL. Adaptive evolution of genes and genomes. New York: Oxford University Press; 1999.
9. The International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–33.
10. Ochman H, Elwyn S, Moran NA. Calibrating bacterial evolution. Proc Natl Acad Sci U S A 1999;96:12638–43.
11. Sharp PM. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J Mol Evol 1991;33:23–33.
12. Smith NGC, Eyre-Walker A. Nucleotide substitution rate estimation in enterobacteria: approximate and maximum-likelihood methods lead to similar conclusions. Mol Biol Evol 2001;18:2124–6.
13. Ochman H, Wilson AC. Evolution in bacteria: evidence for a universal substitution rate in cellular organisms. J Mol Evol 1987;26:74–86.
14. Doolittle RF, Feng D-F, Tsang S, Cho G, Little E. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 1996;271:470–7.
15. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537–44.
16. Takahata N, Nei M. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 1990;124:967–78.
17. Le Dantec C, Winter N, Gicquel B, Vincent V, Picardeau P. Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. J Bacteriol 2001;183:2157–64.
18. Stead WW. The origin and erratic global spread of tuberculosis. How the past explains the present and is the key to the future. Clin Chest Med 1997;18:65–77.
19. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A 2002;99:3684–9.
20. Hughes AL, Verra F. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. Proc R Soc Lond B Biol Sci 2001;268:1855–60.

21. Mu J, Duan J, Makova KD, Joy DA, Huynh CQ, Branch OH, et al. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. Nature 2002;418:323–6.
22. Burnet M. Natural history of infectious disease. Cambridge: Cambridge University Press; 1940.
23. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, et al. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. Science 2002;296:2028–33.
24. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Sequencing of the *M. tuberculosis* genome: comparison of a recent clinical isolate with the laboratory strain. J Bacteriol 2002. In press.

Address for correspondence: Austin L. Hughes, Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia, SC 29208, USA; fax: 803-777-4002; e-mail: austin@biol.sc.edu

EMERGING INFECTIOUS DISEASES
Tracking trends and analyzing new and reemerging infectious disease issues around the world
A peer-reviewed journal published by the National Center for Infectious Diseases Vol.6, No.2, Mar–Apr 2000

Lyme Disease Tuberculosis Vaccines

Global Distribution of *Mycobacterium tuberculosis* Spoligotypes

Ingrid Filliol,* Jeffrey R. Driscoll,† Dick van Soolingen,‡ Barry N. Kreiswirth,§ Kristin Kremer,‡ Georges Valétudie,* Dang Duc Anh,¶ Rachael Barlow,# Dilip Banerjee, Pablo J. Bifani§, Karine Brudey,* Angel Cataldi,†† Robert C.Cooksey,‡‡ Debby V. Cousins,§§ Jeremy W. Dale,¶¶ Odir A. Dellagostin,## Francis Drobniewski,*** Guido Engelmann,††† Séverine Ferdinand,* Deborah Gascoyne-Binzi,# Max Gordon,* M. Cristina Gutierrez,‡‡‡ Walter H. Haas,§§§ Herre Heersma,‡ Gunilla Källenius,¶¶¶ Eric Kassa-Kelembho,### Tuija Koivula,¶¶¶ Ho Minh Ly,¶ Athanasios Makristathis,**** Caterina Mammina,†††† Gerald Martin,‡‡‡‡ Peter Moström,* Igor Mokrousov,§§§§ Valérie Narbonne,¶¶¶¶ Olga Narvskaya,§§§§ Antonino Nastasi,#### Sara Ngo Niobe-Eyangoh,‡‡‡ Jean W Pape,*****††††† Voahangy Rasolofo-Razanamparany,‡‡‡‡‡ Malin Ridell,§§§§§ M. Lucia Rossetti,¶¶¶¶¶ Fritz Stauffer,##### Philip N. Suffys,****** Howard Takiff,†††††† Jeanne Texier-Maugein,‡‡‡‡‡‡ Véronique Vincent,‡‡‡ Jacobus H. de Waard,§§§§§§ Christophe Sola,* and Nalin Rastogi***

We present a short summary of recent observations on the global distribution of the major clades of the *Mycobacterium tuberculosis* complex, the causative agent of tuberculosis. This global distribution was defined by data-mining of an international spoligotyping database, SpolDB3. This database contains 11,708 patterns from as many clinical isolates originating from more than 90 countries. The 11,708 spoligotypes were clustered into 813 shared types. A total of 1,300 orphan patterns (clinical isolates showing a unique spoligotype) were also detected.

Since the publication of the second version of our spoligotypes database on *Mycobacterium tuberculosis* (1), the causative agent of tuberculosis (TB), the proportion of clustered isolates (shared types [STs]) increased from 84% (2,779/3,319) to 90% (11,708/13,008). Fifty percent of the clustered isolates were found in only 20 STs. Three of these isolates are *M. bovis,* including *M. bovis* BCG (ST 481, 482, and 683). The addition of the next 30 most frequent STs increased the total proportion of clustered isolates (65% instead of 50% initially).

A total of 36 potential subfamilies or subclades of *M. tuberculosis* complex have been tentatively identified, leading to the definition of major and minor visual recognition rules (Table). The ancestral East-African Indian family (EAI) is made up of at least five main subclades, whereas at least three major spoligotyping patterns are found within the Haarlem family (2). Two families found in central and Middle Eastern Asia (CAS1 and CAS2) are newly defined. The X family (3) is also currently split into at least three well-defined subclades. However, the subdivision of family T (T1–T4, likely to represent relatively old genotypes), which differs from the classic ST 53 (all spacers present except 33–36), remains poorly defined. Similarly, the Latino-American and Mediterranean family (LAM) is tentatively split into subclades LAM1–LAM10 (4). Spoligotyping used alone is not well suited for studying the phylogeny of these two clades (T and LAM). Such study will require results from other genotyping methods such as IS*6110*-restriction fragment length polymorphism (5) or mycobacterial interspersed repetitive units–variable number of DNA tandem repeats (6). Among well-characterized major clades of tubercle bacilli, four families represent 35% of 11,708 clustered isolates (Beijing 11%, LAM 9.3%, Haarlem 7.5%, and the X clade 7%).

The global distribution of the most frequently observed spoligotypes by continent in SpolDB3 is as follows. Among the patterns originating in North America (n= 4,276, 32% of the total number of isolates in the database), 16% of the strains are of the Beijing type, 14% belong to ST 137 or ST 119 (X family), and 8% are unique (results not shown). In Central America (n=587, 4.5%), 8% of the strains belong to the ubiquitous ST 53, 7% are ST 50, and 6% are ST 2; the last two STs are part of the Haarlem family. In South America (n=861, 6.6%), the distribution of ST 53 and ST 50 accounts for 10% and 9%, respectively, of the spoligotypes, whereas ST 42 accounts for as much as 9% of the total isolates. The origin of ST 42 remains to be established. In Africa (n=1,432, 11%), ST 59 and ST 53 account for 9% of all isolates studied thus far; however, the values obtained for ST 59 are biased because strains from Zimbabwe are overrepresented. We also observed that *M. africanum* ST 181 accounts for as much as 6% of all spoligotypes from Africa in our sample.

In Europe (n=4,360, 33.5%), ST 53 represents as much as 9% of the spoligotypes, ST 50 and 47 (Haarlem family) repre-

*Institut Pasteur, Pointe-à-Pitre, Guadeloupe, French West Indies; †Wadsworth Center, Albany, New York, USA; ‡National Institute of Public Health and the Environment, Bilthoven, the Netherlands; §Public Health Research Institute, New York, New York, USA; ¶National Institute of Hygiene and Epidemiology, Hanoi, Vietnam; # General Infirmary, Leeds, U.K.; **St. Georges' Hospital Medical School, London, U.K.; ††Instituto de Biotecnologia, Castelar, Argentina; ‡‡Centers for Disease Control and Prevention, Atlanta, Georgia, USA; §§Australian Reference Laboratory for Bovine Tuberculosis, Department of Agriculture, South Perth, Australia; ¶¶University of Surrey, Guildford, Surrey, U.K.; ##Universidade Federal, Pelotas, Brazil; ***Public Health Laboratory Service, Dulwich Hospital, London, U.K.; †††University Children's Hospital, Heidelberg, Germany; ‡‡‡Institut Pasteur, Paris, France; §§§Robert Koch Institute, Berlin, Germany; ¶¶¶Swedish Institute for Infectious Disease Control, Solna, Sweden; ###Institut Pasteur, Bangui, Central African Republic; ****Hygiene-Institut der Universität, Wien, Austria; ††††University of Palermo, Palermo, Italy; ‡‡‡‡Bundesinstitut für gesundheitlichenVerbraucherschutz und Veterinärmedizin, Jena, Germany; §§§§Pasteur Institute of Saint Petersburg, Saint Petersburg, Russia; ¶¶¶¶Centre Hospitalier Universitaire, Brest, France; ####University of Firenze, Firenze, Italy; *****Les Centres Gheskio, Institut National de Laboratoire et de Recherche, Port-au- Prince, Haïti; †††††Cornell University, Ithaca, New York, USA; ‡‡‡‡‡Institut Pasteur, Tananarive, Madagascar; §§§§§Göteborg University, Göteborg, Sweden; ¶¶¶¶¶Universidade Federal do Rio Grande do Sul, Brazil; #####Bundesstaatliche bakteriologisch-serologische Untersuchungsanstalt Wien, Austria; ******Oswaldo Cruz Institute, Rio de Janeiro, Brazil; ††††††Caracas, Venezuela; ‡‡‡‡‡‡Centre Hospitalier Universitaire, Bordeaux, France; and §§§§§§Instituto de Investigacionas Cientificas, Caracas, Venezuela

Table. Excerpt from SpolDB3 database showing prototype spoligotypes, visual recognition rules, and binary and octal description[a]

Rk	ST	Class[b]	Total (n)[c]	Rules[d]	Binary description	Octal
1	1	Beijing	1282	Δ1–34	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■■■■■■	000000000003771
2	53	T1	864	F	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	777777777760771
11	52	T2	163	Δ40 and F	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■□■■■	777777777760731
30	37	T3	71	Δ13 and F	■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	777737777760771
64	40	T4	26	Δ19 and F	■■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■□□□□■■■■■■■	777777377760771
7	47	Haarlem1	246	Δ26–30 and E	■■■■■■■■■■■■■■■■■■■■■■■■■□□□□□□■□□□□■■■■■■■	777777774020771
20	2	Haarlem2	104	Δ1–24, Δ26–30 and E	□□□□□□□□□□□□□□□□□□□□□□□□■□□□□□□■□□□□■■■■■■■	000000004020771
3	50	Haarlem3	519	E	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	777777777720771
6	119	X1	310	C	■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■□□□□■■■■■■■	777776777760771
4	137	X2	427	C and Δ39–42	■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■□□□□■■□□□□■	777776777760601
31	92	X3	70	Δ4–12 and C	■■■□□□□□□□□□□■■■■□■■■■■■■■■■■■■■□□□□■■■□■■■	700036777760731
15	48	EAI1	118	A and Δ40	■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■□■■■■■□■■■	777777777413731
13	19	EAI2	130	Δ3, Δ20–21 and A	■■□■■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■■■■■■	677777477413771
16	11	EAI3	121	Δ2-3, A and Δ37–39	■□□■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■□■■□□□■■■■	477777777413071
8	139	EAI4	234	Δ26–27 and A	■■■■■■■■■■■■■■■■■■■■■■■■■□□■□□□□■□■■■■■■■■■	777777774413771
46	236	EAI5	41	A	■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■□■■■■■■■■■	777777777413771
24	181	Afri1	91	Δ7–9 and Δ39	■■■■■■□□□■■■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■	770777777777671
ND	331	Afri2	9	Δ8–12, Δ21–24 and Δ37–39	■■■■■■■□□□□□■■■■■■■■□□□□■■■■■■■■■■■■□□□■■■■	774077607777071
ND	438	Afri3	3	Δ8–12 and Δ37–39	■■■■■■■□□□□□■■■■■■■■■■■■■■■■■■■■■■■■□□□■■■■	774077777777071
17	482	*M. bovis*-BCG	26	Δ3, Δ9, Δ16 and D	■■□■■■■■□■■■■■■□■■■■■■■■■■■■■■■■■■■■■■□□□□□	676773777777600
ND	641	*M. microti*	8	4-7, 23–24, 37–38	□□□■■■■□□□□□□□□□□□□□□□■■□□□□□□□□□□□□■■□□□□□	074000030000600
ND	592	*M. canetti*	6	30 and 36	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■□□□□□■□□□□□□□	000000000101000
21	26	CAS1	102	Δ4–7, Δ23–34	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□■■■■■■■■■	703777740003771
ND	288	CAS2	6	Δ4–10, Δ23–34	■■■□□□□□□□■■■■■■■■■■■■□□□□□□□□□□□□■■■■■■■■■	700377740003771
12	20	LAM1	152	Δ3 and B	■■□■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	677777607760771
22	17	LAM2	92	Δ3, Δ13 and B	■■□■■■■■■■■■□■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	677737607760771
19	33	LAM3	108	Δ9–11 and B	■■■■■■■■□□□■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	776177607760771
49	60	LAM4	37	Δ40 and B	■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■□■■■	777777607760731
42	93	LAM5	44	Δ13 and B	■■■■■■■■■■■■□■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	777737607760771
37	64	LAM6	47	Δ29 and B	■■■■■■■■■■■■■■■■■■■■□□□□■■■■□■■■□□□□■■■■■■■	777777607560771
36	41	LAM7	48	Δ20, Δ26-27 and B	■■■■■■■■■■■■■■■■■■■□□□□□■□□■■■■■□□□□■■■■■■■	777777404760771
NA	290	LAM8	9	Δ27 and B	■■■■■■■■■■■■■■■■■■■■□□□□■■□■■■■■□□□□■■■■■■■	777777606760771
5	42	LAM9[e]	344	B	■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■■□□□□■■■■■■■	777777607760771
9	61	LAM10	202	Δ23–25 and F	■■■■■■■■■■■■■■■■■■■■■■□□□■■■■■■■□□□□■■■■■■■	777777743760771
26	34	S[f]	82	Δ9–10 and F	■■■■■■■■□□■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	776377777760771
28	451	H37Rv	78	Δ20–21 and F	■■■■■■■■■■■■■■■■■■■□□■■■■■■■■■■■□□□□■■■■■■■	777777477760771

[a]Rk, ranking no.; ND, not done; ST, arbitrary designation; *M., Mycobacterium*.
[b]Class: family definition. See text for the definition of the family acronyms.
[c]Total (n), size of the class; binary and octal, description.
[d]Rule A, absence of spacers 29–32, presence of spacer 33 and absence of spacer 34; rule B, absence of spacers 21–24 and spacers 33–36; rule C : absence of spacer 18 and spacers 33–36; rule D, absence of spacers 39–43; rule E, absence of spacer 31 and spacers 33–36; rule F, absence of spacers 33–36. Clades defined with low sample size, such as Afri2, Afri3, CAS2, and LAM8 are subject to change.
[e]Formerly LAM1.
[f]Formerly LAM2

sent 8% of the cases, and the Beijing family accounts for 4% of the spoligotypes. In the Middle Eastern and central Asian region, where the number of samples obtained is still very low (n=351, 2.7%), a high diversity of strains within the EAI and CAS families has been observed, and no single pattern currently exceeds 5%. Further studies of isolates from these regions are needed, e.g., in India, where our sampling is still anecdotal (n=44 isolates). Notwithstanding the scarcity of available data from this region, the observed diversity suggests that this region might be of great interest for further study of the genetic variation of tubercle bacilli. Contrary to what we observed for the Middle East and central Asia, the Far East Asian region (n=801, 6.1%) is characterized by the prevalence of a single genotype, the Beijing type family, a family linked to emerging multiresistance (7). One out of two strains in the Far East is a Beijing type. In Oceania (n=340, 2.6%), ST 19 and Beijing account for 15% and 13%, respectively, of clustered isolates. Thus, this preliminary analysis of the spoligotype distribution of SpolDB3 clearly shows major differences in the population structure of tubercle bacilli within the eight subcontinents studied (Africa; Europe; North America; Central America; South America; Middle East and Central Asia; Far East Asia; and Oceania).

At present, SpolDB 3 is an experimental tool that has yet to prove its usefulness in tracking epidemics. Nevertheless, the facility with which matches between spoligotypes can be detected suggests that this tool may be a good screening mechanism for population-based studies on recent TB transmission. Indeed, the detection of a rarely found ST in SpolDB3 may be a catalyst that signals researchers to look for the clonality of the isolates and to study their epidemiologic relatedness.

Data-exchange protocols through inter-networking will also be implemented in the near future. Working groups such as the European Network for Exchange of Molecular Typing Information (available from: URL: www.rivm.nl/enemti) are coordinating such initiatives. The expanded use of the Bionumerics software (third upgrade; Applied Maths, St. Martens-Latem, Belgium) may also foster this research field. SpolDB3 will also be instrumental in facilitating better understanding of the driving forces that shape tubercle bacilli evolution. Further research should now emphasize the use of data-mining methods, in combination with experts' knowledge, to tackle the complex dynamics of the population's genetics of tubercle bacilli and TB transmission (3). Our sample represents the compilation of many national studies and, as such, should be considered as an ongoing population-based project aimed at studying global TB genetic diversity. Nevertheless, obtaining a more precise and representative snapshot of the genetic variability of *M. tuberculosis* complex will require a larger sampling. Although only partially representative of worldwide spoligotypes of *M. tuberculosis* complex, Spo1DB3 contains a reservoir of genetic information that has already proved useful for defining the phylogenetic links that exist within the TB genomes and for constructing theoretical models of genome evolution. Much remains to be done to evaluate the potential of global genetic databases to better characterize casual contacts (that could lead to identification of sporadic cases) in TB epidemiology. An improved version of our database, which will focus on areas with a high prevalence of TB, is currently in development; as of August 26, 2002, it had 20,000 isolates and 3,000 alleles. Ongoing population-based genotyping projects will likely help shed light on contemporary and ancient tubercle bacilli's evolutionary history.

This paper was written as part of the EU Concerted Action project QLK2-CT-2000-00630 and partly supported by the Réseau International des Instituts Pasteur et Instituts Associés, Institut Pasteur and Fondation Française Raoul Follereau, France. An electronic, simplified, version of SpolDB3 is available from the corresponding authors upon request.

Dr. Filliol performed this work as part of her doctoral thesis. She has been working at the Institut Pasteur de Guadeloupe for the last 4 years. Her research focuses on molecular epidemiology and phylogeny of tubercle bacilli.

References

1. Sola C, Filliol I, Guttierez CM, Mokrousov I, Vincent V, Rastogi N. Spoligotype database of *Mycobacterium tuberculosis*: biogeographical distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg Infect Dis 2001;7:390–6.
2. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PWM, Martin C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 1999;37:2607–18.
3. Sebban M, Mokrousov I, Rastogi N, Sola C. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis*. Bioinformatics 2002;18:235–43.
4. Sola C, Filliol I, Legrand E, Mokrousov I, Rastogi N. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS*1081*, IS*6110*, VNTR and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J Mol Evol 2001;53:680–89.
5. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993; 31:406–9.
6. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C. Automated high-throughput genotyping for the study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 2001;39:3563–71.
7. Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg Infect Dis 2002;8:843–9.

Address for correspondence: Christophe Sola and Nalin Rastogi, Unité de la Tuberculose et des Mycobactéries, Institut Pasteur de Guadeloupe, BP 484, 97165 Pointe-à-Pitre Cedex, Guadeloupe, French West Indies; fax : 590(0)590- 893880; e-mails: csola@pasteur.gp; rastogi@pasteur.gp

Two Cases of Pulmonary Tuberculosis Caused by *Mycobacterium tuberculosis* subsp. *canetti*

Jean Miltgen,* Marc Morillon,* Jean-Louis Koeck,†
Anne Varnerot,‡ Jean-François Briant,*
Gilbert Nguyen,* Denis Verrot,* Daniel Bonnet,*
and Véronique Vincent‡

We identified an unusual strain of mycobacteria from two patients with pulmonary tuberculosis by its smooth, glossy morphotype and, primarily, its genotypic characteristics. Spoligotyping and restriction fragment length polymorphism typing were carried out with the insertion sequence IS*6110* patterns. All known cases of tuberculosis caused by *Mycobacterium canetti* have been contracted in the Horn of Africa.

The *Mycobacterium tuberculosis* complex includes the following mycobacteria, which are characterized by a slow growing rate: *M. tuberculosis, M. africanum, M. bovis,* and *M. microti* (1). In recently published reports of two cases of lymphatic node tuberculosis (TB), the strains were recognized as belonging to a new taxon of *M. tuberculosis* (2,3). These isolates were characterized by a highly particular growing pattern, and the colonies appeared smooth and glossy. A complete genetic study of these strains led to their integration into the *M. tuberculosis* complex. This strain, identified as *M. tuberculosis* subsp. *canetti* or, more simply, *M. canetti,* was first isolated in 1969 by Georges Canetti from a French farmer. The strain was preserved at the Pasteur Institute where its antigenic pattern was studied extensively. We report two cases of pulmonary TB caused by this strain. The two patients had also lived in East Africa.

The Study

Patient 1

In September 1998, a 36-year-old male soldier in the French Foreign Legion with hemoptysis was sent back to France from Djibouti. He expectorated bloody sputum after running and on a few other occasions. His medical history was not unusual. When the patient was hospitalized, 2 weeks after the initial symptoms, he began to experience progressive fatigue. He did not experience fever, weight loss, night sweats, anorexia, cough, dyspnea, or chest pain, and did not produce sputum.

Results of the clinical examination were normal. The Mantoux test, performed with 10 IU of purified tuberculin (Aventis-Pasteur-MSD, Lyon, France), yielded a maximum transverse diameter of induration of 15 mm. Laboratory values were normal (Table). The chest X-ray showed a triangular consolidation of the left upper lobe with blurred limits and small cavitary lesions. No other contiguous mediastinohilar anomalies were visible. A computed tomographic scan confirmed the cavitary syndrome: three excavated nodular images showed radiating spicules within a micronodular infiltrate. Bronchoscopy showed a moderate inflammation of airway mucosa, especially in the left upper lobe. Biopsy specimens exhibited nonspecific inflammation.

A bronchial washing smear from the left upper lobe was positive for acid-fast bacilli. Serologic tests for HIV-1 and HIV-2 were negative. No evidence of disease was found elsewhere; the patient did not experience bone pain. Results of neurologic and ophthalmologic examinations were normal; no lymphadenopathy or hepatosplenomegaly were found and the genitalia were normal. Auscultation revealed no pericardial fremitus; no ascitic fluid was detected. The urinary sediment contained <1,000 red blood cells/L and <5,000 leukocytes/L. Antituberculosis chemotherapy was begun with four drugs: rifampicin, isoniazid, ethambutol, and pyrazinamide. Cultures revealed a strain identified as *M. tuberculosis* subsp. *canetti* that was susceptible to all primary antituberculous drugs. Therefore, rifampicin and isoniazid were continued for 3 more months for a total treatment period of 6 months. The patient's response to treatment was favorable, and he remained asymptomatic.

Patient 2

A 55-year-old male soldier in the French Foreign Legion, who returned from Djibouti, was hospitalized in September 1999 after his chest x-ray showed abnormal findings. He was a nurse and had been occasionally in charge at the Djibouti Hospital for 2 years. His medical history was unremarkable. Eight months before he returned to France, he experienced asthenia, anorexia, and a weight loss of 3 kg. The symptoms resolved spontaneously after 2 months, and he had been asymptomatic since then. He had no history of cough, sputum production, hemoptysis, dyspnea, fever, or night sweats.

Results of a clinical examination and of laboratory studies were normal (Table), except for hypereosinophilia. Serologic tests for schistosomiasis, hydatidosis, distomiasis, amebiasis, toxocariasis, and trichinosis were negative, and parasites were not found in stool samples. Thoracic radiographs performed when he came back from Djibouti showed parenchymal consolidation of the right upper lobe with small cavities. Sputum was not produced. A gastric aspirate smear was negative for acid-fast bacilli, and a bronchial aspiration smear was positive

*Hôpital d'instruction des armées Laveran, Marseille, France; †Hôpital d'instruction des armées Val de Grâce, Paris, France; ‡Institut Pasteur, Paris, France

Table. Laboratory values for both patients infected with *Mycobacterium tuberculosis* subsp. *canetti*

Laboratory test	Patient 1	Patient 2
Sedimentation rate (mm)	3	3
C-reactive protein (mg/L)	7.3	4.28
Fibrinogen (g/L)	3.64	6.3
Blood count		
Hemoglobin (g/dL)	16.8	14.1
Platelets (x10^9/L)	194	274
White cells (x10^9/L)	10.10	9.54
Neutrophils (%)	66.2	69.4
Eosinophils (%)	2	8.6
Lymphocytes (%)	21.3	17.2
Basophils (%)	0.6	0.9
Monocytes (%)	9.9	3.9
Aspartate aminotransferase (U/L)	21	16
Alanine aminotransferase (U/L)	19	23
Creatinine (µmol/L)	89	97
Glucose (mmol/L)	4.7	4.4

for acid-fast bacilli. HIV serology was negative, and no other site of the infection was found. Drug therapy was initiated with rifampicin, isoniazid, ethambutol, and pyrazinamide for 2 months. Cultures of bronchial aspirates were positive within 14 days; later, cultures of two gastric aspirates were positive for acid-fast bacilli. An *M. tuberculosis* subsp. *canetti* isolate was identified, which was susceptible to all primary antituberculous drugs. The treatment was then extended for 4 months with rifampicin and isoniazid. The patient's response to treatment was favorable.

The following methods were used to identify the etiologic agent. First, the samples were decontaminated with N-acetyl-L-cysteine/NaOH. Acid-fast bacilli were detected by auramine staining, the positive smears also were stained with Ziehl-Nielsen stain. The samples were then seeded onto Löwenstein-Jensen and Coletsos slants and also into a liquid system, the BBL Mycobacterial Growth Indicator Tube (MGIT, BD Diagnostic Systems, Sparks, MD).

The mycobacteria were identified by using a specific DNA probe (Gen-Probe, Gen-Probe Incorporated, San Diego, CA) and by performing the usual biochemical tests (nitrate reduction, 68°C catalase resistance, niacin production).

The Pasteur Institute of Paris used two methods for typing: restriction fragment length polymorphism (RFLP) analysis and spoligotyping. In RFLP analysis, after digestion of the *M. tuberculosis* strain's genomic DNA with *Pvu*II restriction enzyme and agarose gel migration, the DNA was transferred on a membrane, according to the Southern method, and then hybridized with an insertion sequence IS*6110* probe (4). In the spoligotyping method, after DNA direct repeat amplification, the labeled polymerase chain reaction product was used as a probe to hybridize with 43 synthetic spacer oligonucleotides (DNA sequences derived from the direct repeat [DR] region of *M. tuberculosis,* H37Rv and *M. bovis* BCG P3), which were attached to a carrier membrane (5). The sensitivity to antituberculous drugs was determined by the indirect proportion method.

MGIT results were positive for the two cultures in 9 and 12 days, respectively. On Löwenstein-Jensen slants, the cultures were positive in 12 and 14 days, respectively. The white, smooth, and glossy colonies were characteristic of *M. tuberculosis* subsp. *canetti* (Figure 1). The two strains had the same phenotypic and genotypic pattern; 68°C catalase was negative, and they reduced nitrate, as do other *M. tuberculosis* species, but they did not produce niacin. The DNA probe, Gen-Probe, confirmed that these strains belonged to the *M. tuberculosis* complex.

These strains contained two copies of IS*6110*. Spoligotyping showed that they shared only 2 of the 43 oligonucleotides reproducing the spacer DNA sequences of *M. tuberculosis,*

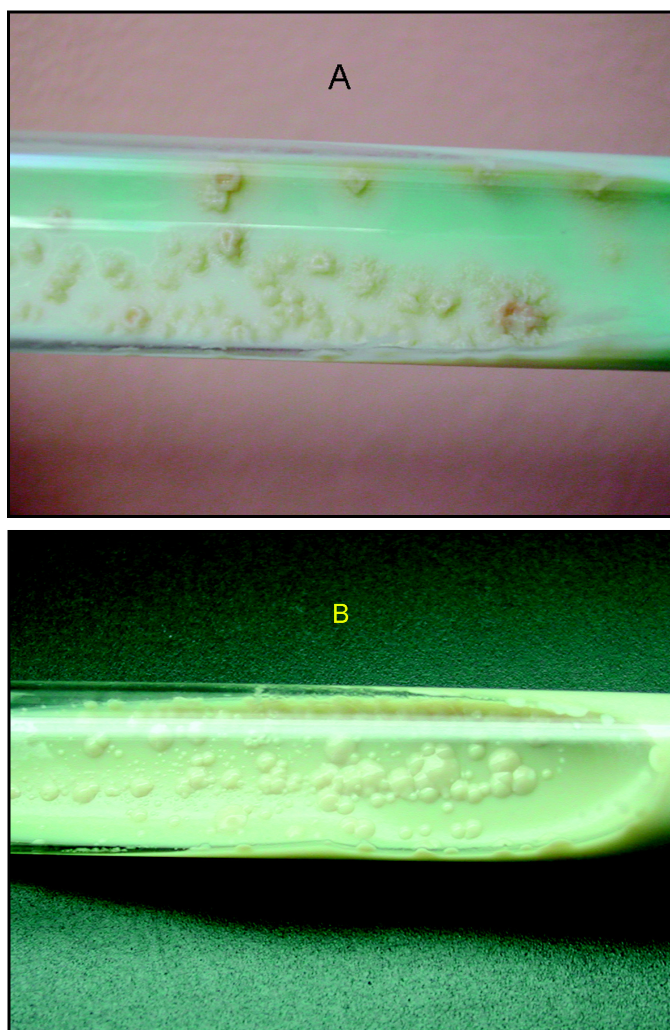

Figure 1. Colony morphology on Löwenstein-Jensen slants, showing *M. canetti* and *M. tuberculosis* strains. (A) Colonies of *M. tuberculosis* are rough, thick, wrinkled, have an irregular margin, and are faintly buff-colored. (B) *M. canetti* exhibits smooth, white and glossy colonies.

H37Rv and *M. bovis* BCG P3. This profile is characteristic of *M. tuberculosis* subsp. *canetti* (Figure 2).

Conclusion

In 1997, van Soolingen reported a case of lymph node TB in a 2-year-old Somali child on the child's arrival in the Netherlands in 1993 (2). In 1998, Pfyffer described abdominal lymphatic TB in a 56-year-old Swiss man (who lived in Kenya) with stage C2 HIV infection (3). These strains of *M. canetti* (So93 from the Somali child and NZM 217/94 from the Swiss man) have been studied extensively. In culture they grow faster than other strains in the *M. tuberculosis* complex. The So93 strain expands by one rough colony for every 500 smooth colonies. They appear smooth, white, and glossy because of the high amount of lipooligosaccharides in the membrane (6); the So93 rough colonies lack this amount (2).

Two copies of the IS*6110* insertion sequence were found in the NZM 217/94 and So93 genome. This fingerprint matched none of the 5,000 other strains preserved in the laboratory of van Soolingen (Bilthoven, the Netherlands) (2). The strains we observed also showed two copies of IS*6110*.

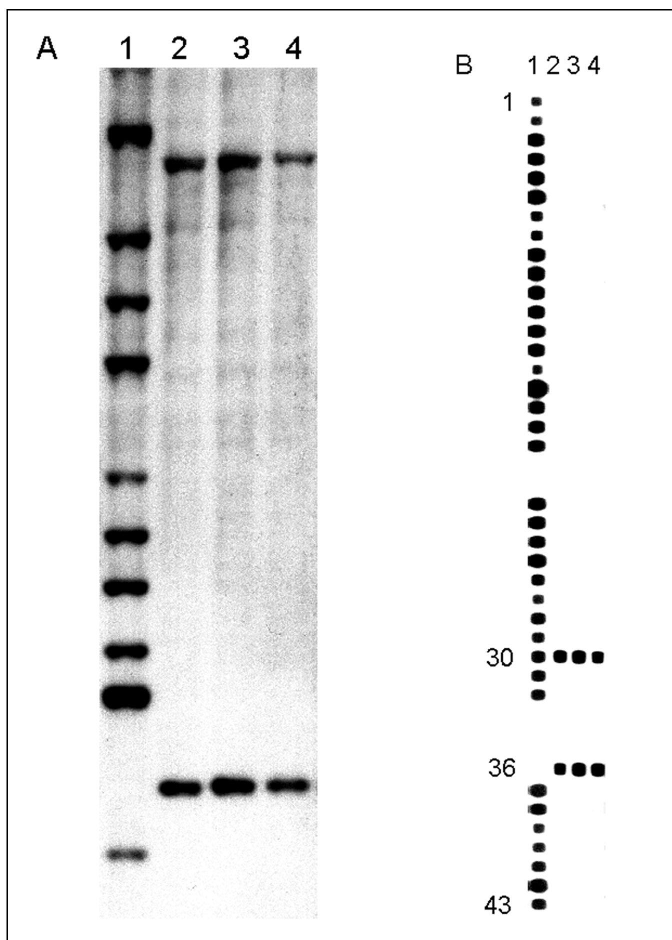

Figure 2. (A) IS*6110* hybridization patterns of *Pvu*II-digested genomic DNA. Lane 1, *Mycobacterium tuberculosis Mt* 14323 (reference strain). Lane 2, *M. canetti* strain NZM 217/94. Lanes 3 and 4, the strains isolated from French legionnaires with pulmonary tuberculosis (TB). (B) Spoligotyping patterns. Lane 1, *M. tuberculosis* H37Rv (reference strain). Lane 2, *M. canetti* strain NZM 217/94. Lanes 3 and 4, the strains isolated from French legionnaires with pulmonary TB.

So93, NZM 217/94, and our two strains share only 2 of 43 identical repeated sequences that have been observed by spoligotyping. Study of the IS*6110* RFLP patterns and of the spacer DNA sequences of the DR locus confirmed that *M. tuberculosis, M. bovis, M. africanum*, *M. microti,* and M. *canetti* represent a closely related group of mycobacteria that are clearly distinct from other mycobacterial species. In the *M. tuberculosis* complex*, M. canetti* appears to be the most divergent strain (2).

We believe that this is the first published report of pulmonary disease caused by *M. canetti*. Our two cases confirm that *M. canetti* is able to involve lungs, like any other member of the *M*. t*uberculosis* complex and is able to affect immunocompetent subjects. The clinical features of these two pulmonary cases of TB caused by *M. canetti* are not specific.

TB caused by *M. canetti* appears to be an emerging disease in the Horn of Africa. A history of a visit to the region should cause this strain to be considered promptly. As travel to this area becomes more frequent, and mycobacterial identification techniques improve, the number of diagnosed cases will likely increase.

Acknowledgments

We thank Michel Fabre for the photographs and Jan Eskandari for his translation of this article.

Dr. Miltgen is assistant head of the Pneumology Department at the Hôpital d'Instruction des Armées of Marseilles, where he specializes in tropical diseases.

References

1. Shinnick TM, Good RC. Mycobacterial taxonomy. Eur J Clin Microbiol Infect Dis 1994;13:884–901.
2. van Soolingen D, Hoogenboezem T, De Haas PEW, Hermans PWM, Koedam MA, Teppema KS, et al. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int J Syst Bacteriol 1997;47:1236–45.
3. Pfyffer GE, Auckenthaler R, van Embden JDA, van Soolingen D. *Mycobacterium canetti*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. Emerg Infect Dis 1998;4:631–4.
4. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
5. Kamerbeek J, Schouls L, Kolk A, Van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997;35:907–14.
6. Daffe M, McNeil M, Brennan PJ. Novel type-specific lipooligosaccharides from *Mycobacterium tuberculosis*. Biochemistry 1991;30:378–88.

Address for correspondence: Jean Miltgen, Service de pneumologie, Hôpital d'instruction des armées Laveran, BP 50, 13998 Marseille, France; fax: 04 91 61 73 37; e-mail: jean.miltgen@wanadoo.fr

Robert Louis Stevenson's sarcophagus, on top of Mount Vaea, Upolu, Western Samoa. Photo courtesy of David Morens

At the Deathbed of Consumptive Art

David M. Morens

Under the wide and starry sky,
Dig the grave and let me lie.
—R. L. Stevenson, “Requiem”

For more than a century, readers have pondered the strange beginning to one of the most haunting poems in the English language, “Requiem.” Who has not wondered how a poet can seem to welcome his own death? Scotsman Robert Louis Stevenson died of a disease so poorly understood in his day that over a few decades its preferred name changed three times, from "phthisis" to "consumption" to "tuberculosis.” A century later, we have reliable scientific facts on this "old" poet-killing disease—we know for one that *Mycobacterium tuberculosis* now infects 1.9 billion people, nearly a third of the world’s population (nearly 2 million deaths each year). But human suffering is still difficult to quantify.

Two million deaths each year. How can we grasp such statistics of misery? Reduced to numbers stacked up in columns and cut up in pie charts, tuberculosis patients don’t seem like us. They live in faraway places, come from obscure cultures, speak incomprehensible languages, have disreputable comorbid disease, or exhibit antisocial behavior. They are “the other.” Maybe they do not even exist.

To grasp the human suffering brought about by tuberculosis, we may need to recall the past when, incurable and incomprehensible, the disease had to be deciphered by metaphors—metaphors that changed as societal views of the disease changed over time. We may need to recall the lives of dying artists and the work they created and let art paint their faces, sculpt their shapes and contours, and compose leitmotivs. Perhaps such *past* images will help fix the gazes of *today’s* victims, whose faces we do not seem to be able to see.

The arts (the novel, the play, the poem, the musical composition, the operatic production) have always helped us understand, given us perspective, invoked compassion, and argued a purpose and meaning for life. We listened to the molto adagio of Barber’s string quartet (opus 11), and came together at President Roosevelt’s death. We read the poems of Walt Whitman and taught our children about the soldier’s sacrifice. We studied the works of Petrarcha and Guillaume de Machaut and revisited the terror of Black Death, nearly 700 years ago. Now, we must turn to art again to grasp human suffering because

scientific knowledge of the disease seems to have displaced our interest in the patient.

Tuberculosis and the Arts

Poetry

In his modest room, a 25-year-old "lapsed" medical student in 1820s Great Britain wakes from a sudden fevered sweat and finds a single drop of blood on the sheet. He has known many patients who spit such bright blood. "It's arterial blood…that blood is my death warrant, I must die," he confides to a friend. One of England's greatest poets, the medical student John Keats, never wrote specifically about phthisis. But his life and his works became a metaphor for generations of patients, a metaphor that helped transform the physical disease "phthisis" into its spiritual offspring, "consumption".

Keats' life was defined by tuberculosis. At 14, he nursed his mother as she died of phthisis. A few years later he watched his brother die of phthisis, and by age 23, he had symptoms of this "hereditary" ailment himself. Yet, as the best remedies (bleeding, diet, red wine, opium) failed, as his work was savaged by critics and he was forced from the woman he loved (he could not marry because of the disease), Keats feverishly wrote his greatest poems: Ode to a Nightingale, Ode on a Grecian Urn, Ode on Melancholy.

He died only months after he first spit blood. Autopsy found his lungs completely destroyed. He was 26. To be dead himself within the year, Percy Bysshe Shelley, another young poet with phthisis, compared Keats to "a pale flower by some sad maiden cherished,/And fed with true-love tears, instead of dew/The bloom, whose petals nipped before they blew/Died on the promise of the fruit, is waste;/The broken lily lies—the storm is overpast."

Shelley's tribute expressed what would become the central metaphor of consumption in the 19th century, the idea that the phthisic body is consumed from within by its passions, "the bloom…Died on the promise of the fruit." Shelley also likened Keats to an eagle that "outsoared the shadow of our night," and "could scale Heaven, and could nourish in the sun's domain." These romantic notions contrasted sharply with the actual demise of tuberculosis patients, poets or not. As death neared, Keats himself contradicted Shelley. In "On Seeing the Elgin Marbles" (published posthumously), Keats wrote, "My spirit is too weak—mortality/Weighs heavily upon me like unwilling sleep,/And each imagin'd pinnacle and steep/Of godlike hardship, tells me I must die/Like a sick Eagle looking at the sky."

In succeeding decades, Keats' illness came to exemplify the so-called spes phthisica, a condition believed peculiar to consumptives in which physical wasting led to euphoric flowering of the passionate and creative aspects of the soul. The prosaic human, it was said, became poetic as the body expired from consumption, genius bursting forth from the fevered combustion of ordinary talent, the body burning so that the creative soul could be released. Keats' great poetic output during his last year was considered a direct consequence of consumption.

Spes phthisica, which sought to make sense of the senseless and give purpose to purposeless suffering and death, came to be viewed as a prerequisite for creative genius. French author Alexandre Dumas fils wrote, "It was the fashion to suffer from the lungs; everybody was consumptive, poets especially; it was good form to spit blood after any emotion that was at all sensational, and to die before reaching the age of thirty." Dumas' colleague, the poet Théophile Gautier, wrote, "…I could not have accepted as a lyrical poet anyone weighing more than ninety-nine pounds." A subsequent alleged decline in the arts was even blamed on decline in tuberculosis incidence.

Opera and the Novel

In the 19th century, it seemed as if everyone was slowly dying of consumption. Without explanation, the good and the bad, the young and the old, all shared the consumptive fate. Consumption came to be viewed not in medical terms (medicine had little to offer anyway), but in popular terms, first as romantic redemption, then as reflection of societal ills. The consumptive prostitute, for example, could be a moral deviant redeemed by suffering and death. Redemption occurred not in the confessional, nor in the "last rites," but in consumption's physical sacrifice. Alternatively, the prostitute could be merely a woman victimized by a corrupt society.

Opera perhaps provides the most powerful examples of romantic redemption through tuberculosis. The pallor and wasting, the burning sunken eyes, the perspiration-anointed skin—all hallmarks of the disease—came to represent haunted feminine beauty, romantic passion, and fevered sexuality, notions reinforced by the excess of consumption deaths in young women. Of several operas that deal explicitly with consumption, three (written within a 6-year interval but produced successively over the last half of the 19th century) show the evolution in thinking about the disease.

Verdi's La traviata [The Lost One] is based on La dame aux camélias [The Lady of the Camellias], Dumas' tale chronicling the life of Parisian beauty Rose Alphonsine Plessis (also known as Marie Duplessis, 1824–1847). Violetta, the heroine, is a courtesan whose loveliness is stereotypically enhanced and made unforgettable by progressive consumption. So strong was the expectation that women dying of consumption looked like beautiful ghosts that, at La traviata's 1853 premiere, the audience broke out laughing when Violetta, played by the ample soprano Fanny Salvini-Donatelli, was told she had only hours to live.

The plot of La traviata develops around the consequences of Violetta's scandalous past, which prevents her marriage to an honorable man whose family objects. In the libretto (by Francesco Maria Piave), Violetta links Alfredo's acceptance to a cure for her consumption. Referring to their separation, she asks, "Do you know the fatal ill, That attacks my life—to kill? That already draws the end? And you'd part me from my friend!"

Eventually realizing that redemption is possible only through death, Violetta withdraws to her former world to suf-

fer and die for love: "If he knew the sacrifice Whereby love I atone—Knew that to my last breath I loved just him alone…." In taking her life, consumption also serves as a vehicle for atonement. Violetta dies redeemed in the eyes of Alfredo and his father, her labored breaths fading as she pleads: "Look at me! I breathe for you still."

Produced only 28 years later (1881; the year before Robert Koch's "discovery" of *Mycobacterium tuberculosis*), Les contes d'Hoffmann [Tales of Hoffmann] exhibits an important shift in thinking about consumption. In Hoffmann, the beautiful consumptive Antonia is treated by the charlatan physician Dr. Miracle. At the time of Keats' death, in 1821, little could be done to treat phthisis; the physician's role was in prognosis. By 1881, the "medicalization" of consumption was in full swing, with diets, nostrums, regimens, and activity lists (all of little or no benefit).

By portraying Antonia as a victim of quackery, Hoffmann satirizes medical impotence. Further, the opera links Antonia's consumption to her mother's. Heredity as a possible cause of consumption (a popular concept before Koch's discovery) appealed to the opera's audience because it absolved the patient from guilt or shame. Heredity as cause may have also appealed to the opera's composer Jacques Offenbach—both he and his son were consumptive. The libretto (by Jules Barbier and Michel Carré) seems also to presage a new medical concept, a mycobacterial neurotoxin postulated as the cause of spes phthisica after Koch's discovery.

In Hoffmann, Dr. Miracle asks, "Well, this trouble she inherited From her mother? Still progressing. We'll cure her." As he forces her to sing, in what amounts to a musical exorcism, he notes, "The pulse is unequal and fast, bad symptom [sic]. Sing." The ghost of Antonia's mother, conjured by Dr. Miracle, calls to her, and she sings frantically, an artistic representation of her consumptive decline. Nearing death, she asks, "What ardor draws and devours me? I give way to a transport that maddens. What flame is it dazzles my eyes. A single moment to live, And my soul flies to Heaven." Antonia is merely a victim, not burdened by sins that need to be atoned for. Unlike Violetta, she dies maddened by disease, without comprehending why and without making any moral choice.

By 1896, the cause of consumption had been discovered. Tuberculosis or TB (as the disease was now becoming known to everyone) had also been definitively linked to poverty and industrial blight, child labor, and sweatshops. A contagious disease and shameful indicator of class, it was no longer easily romanticized in conventional artistic terms. Public health efforts to isolate the infected and control their behavior were everywhere.

Giacomo Puccini's La bohème [The Bohemian Life] (1896) presents tuberculosis as social commentary and features characters new to opera, street artists living with poverty and disease. The opera was fashioned from a book by Louis-Henri Murger, a young writer who also died of tuberculosis. The heroine, Mimì, is based on Murger's lover Lucile Louvet. A seamstress portrayed as a fevered beauty whose allure is heightened by physical decline, Mimì loves a poor poet named Rodolfo. He first loves her in return but then abandons her, fearing he cannot provide for her and (possibly) fearing contagion.

In the libretto (by Giuseppe Giacosa and Luigi Illica), Mimì tells Rodolfo, "My story is brief. I embroider silk and satin at home or outside. I love all things that have gentle magic, that talk of love, of spring…the things called poetry. Do you understand me? They call me Mimì. I don't know why. I live all by myself and I eat alone. I look at the roofs and the sky. But when spring comes the sun's first rays are mine. April's first kiss is mine, is mine! A rose blooms in my vase, I breathe its perfume, petal by petal."

Later, in the winter cold, Mimì is reconciled with Rodolfo and dies beside him. In one of opera's most enduring scenes, there is no attempt at metaphorical understanding. Mimì dies literally. No one is saved, no one is redeemed, and no larger point is made. Opera and tuberculosis have entered a new era, recognizable today, in which tragedy is seen as experiencing loss but is not necessarily understood in an artistic or philosophical sense. In the moments before her death, Rodolfo tells Mimì that she is "beautiful as the dawn." "You've mistaken the image:" she corrects him, "You should have said, beautiful as the sunset." Mimì dies surrounded by a philosopher, a poet, a painter, a musician, and a singer—the arts have become powerless against tuberculosis.

Even though new, La bohème's approach of pointing the finger at society as the cause of human suffering was not unprecedented. An earlier example from literature is Victor Hugo's Les misérables [The Outcasts], published in 1862. In the novel, the hounded protagonist (Jean Valjean) finds redemption in the adoption of a child of a dying woman (Fantine) who had been forced by poverty into prostitution. Like Mimì's in La bohème, Fantine's death from consumption is portrayed as a consequence of social ills. Society, Hugo says, victimizes good people by putting them in hopeless circumstances.

Of Fantine's death Hugo writes with a simple beauty that is eloquent even in translation. Like the "afterthought" the world had regarded the poor woman's entire life, the paragraph describing her death is inserted almost randomly, in retrospect: "We all have one mother—the earth. Fantine was restored to this mother. [She] was buried in the common grave of the cemetery, which is for everybody and for all, and in which the poor are lost. Happily, God knows where to find the soul. Fantine was laid away in the darkness with bodies which had no name; she suffered the promiscuity of dust. She was thrown into the public pit. Her tomb was like her bed."

In the same romantic context is Hugo's earlier novel Notre-Dame de Paris (1831), which featured the grotesque and deformed "hunchback" Quasimodo. In those days, the principal cause of severe spinal deformity was Pott's disease, tuberculosis of the spine. What Hugo knew of this condition can only be speculated, but as a leading poet, he would unquestionably have been familiar with the life of one of England's

greatest poets, Alexander Pope (1688–1744), a hunchback severely deformed in childhood by Pott's disease. Bent to a height of 4½ feet, at age 47, Pope famously described in "An Epistle from Mr. Pope, to Dr. Arbuthnot" "this long Disease, my Life" Quasimodo's physical "otherness" is Hugo's device for insisting we recognize his humanity, without judgment, in all its dimensions, good and ill.

Notre-Dame de Paris carries a simple message, one expunged from many films and operas made of the book. Hugo writes that long after the events depicted in the novel, workmen discovered two entwined skeletons, one of a young woman, the other of a man with a twisted spine. After La Esmeralda's execution, the reader surmises, Quasimodo had ended his life beside the woman who had found him so repugnant. In death, his disfigurement and her beauty alike were dissolved in dust. All of Quasimodo's actions, misunderstood by everyone else, are explained by love, as pure in the deformed, slow-witted bell-ringer as in the beautiful saint.

Operatic and literary examples show the romantic transformation of tuberculosis. Only rarely in the period before Koch's discovery was the disease portrayed realistically in artistic works. The most powerful example of realistic portrayal may be Tolstoy's Anna Karénina (1877). Apparently based on Tolstoy's own brother Nikolai, the novel's character Nikolai Levin suffers an agonizing and vividly depicted death. "A desire for his death was now felt by everyone who saw him; by the hotel servants and the proprietor…by Maria Nikolayevna, Levin, and Kitty. [I]n rare moments, when opium made him find momentary oblivion…[he] expressed what he felt more intensely in his heart than all the others: 'Oh, if only it was all over!' or 'When will this end?' There was no position in which he could lie without pain, there was not a moment in which he could forget…And that was why all his desires were merged into one—a desire to get rid of all the sufferings and their source, the body." Unlike the prostitute whose death provided Jean Valjean's redemption, Maria was rescued from prostitution by Nikolai and nursed him when he was abandoned. As Nikolai entered the long descent into suffering and death, Maria's love was his only human connection.

Science and Public Health

The Romantic Era of Consumption (1821–1881)

At the beginning of this period, the stethoscope was invented and used to diagnose phthisis, and statistics were compiled by population-oriented proto-epidemiologists (e.g., Louis-François Benoiston de Châteauneuf) and by clinical proto-epidemiologists (e.g., Pierre-Charles-Alexandre Louis). In the 1820s, contagion was beginning to coalesce into a modern concept, although it was not imagined in chronic conditions like phthisis. Science was beginning to understand the disease clinically and epidemiologically (not yet etiologically). It was left to the arts to make sense of misery and death in ways that turned otherwise senseless suffering into human dignity and hope, allowing consumption to reveal the innate worth even of prostitutes, the impoverished, and the socially outcast.

Discovery of *Mycobacterium tuberculosis* (1882–1952)

At the end of the 19th century, romantic notions about tuberculosis were replaced by scientific ideas and products: vaccines and therapies, rest cures, tonics, pneumonectomies, lobectomies, thoracoplasties, "artificial" pneumothoracies, phrenic nerve crushings, plombage, pneumatic cabinet treatments, and antiseptic injections into the pleural spaces. Surgeons packed the pleural cavity with fat, paraffin, and even Ping-Pong balls. In contrast to the arts, science and medicine were unapologetically prosaic.

Over a 70-year span, tuberculosis literally transformed Western society. Tuberculosis patients were excluded from many occupations. Married patients had to sleep in separate beds from uninfected spouses and were counseled to avoid sex and especially to not have children. Public health nurses visited door to door, sanatoria were built by the hundreds, and hospitals added tuberculosis wings. Cold water hydrotherapy, alcohol massages, and brisk rubdowns with coarse towels were prescribed. Millions of spittoons and cuspidors were placed in homes and public places. Patients' bed linens were changed daily and were boiled and laundered separately. Handkerchiefs (for those who could afford them—cut-up pieces of muslin for the rest) were stuffed in pockets before leaving home, then disinfected or burned at night. Japanese "paper handkerchiefs" became popular, leading eventually to the modern "facial tissue." Tuberculous women had to forego corsets and brassieres in favor of loose-fitting clothes. Life insurance policies added clauses canceling benefits for the tubercular, and hotels and landlords refused to serve them. Compulsory registration, immigration bans, and even interstate travel restrictions were debated. Suicides in towns with sanatoria increased.

Prevention efforts were put in place. National antituberculosis programs were led by U.S. presidents. Ubiquitous "ad campaigns" featured catchy tunes some called "jingles." Architects designed alcoves to hide concealed spittoons in middle-class homes. Babies were no longer allowed to play on the floor, and mothers were told not to kiss children on the mouth. Some churches abandoned the "common" communion cup. "TB" and "x-ray" became household words. Streets were watered down before sweeping to prevent aerosols. Long "trailing" dresses went out of fashion because they dragged on the ground and picked up potentially infectious dust. Store candies and bakery loaves had to be wrapped; public libraries and their books were regularly disinfected.

Modern Times (1952–2002)

In the late 1940s and early 1950s, trucks cycled through neighborhoods to administer chest x-rays, and schools provided "tine tests." Patients sat in big sunrooms off hospital wards, semi-erect in lounge chairs, wrapped like mummies on an ocean cruise. Family members disappeared for long periods—off to sanatoria for "rest cures." People bought Christ-

mas Seals, and spoke in a slight hush when they mentioned "TB." As a medical student in the 1960s, I remember the second patient I ever examined. Born in 1879, she had undergone a right phrenic nerve crush, sometime around 1910, surviving tuberculosis to the age of 90.

The 1950s represented the cusp of a new era in which drug treatment would end tuberculosis visibility in industrialized countries. People still read Thomas Mann's The Magic Mountain (1924), which was set in the real-life sanatorium in Davos, Switzerland, but less to learn about tuberculosis and more to understand the "Sanatorium Movement." The modern tuberculosis era began around 1952 with antituberculosis chemotherapy, but more die of the disease today than in the 19th century. Medicine and public health have answers but lack the will to apply them. Tragedy now derives not from unknown evils out of our control but from known evils we do not control.

Art is redemptive. Even in modern times, when artistic expression seems to have moved from tuberculosis to other causes of human suffering, dramatic irony occasionally sneaks backstage to comment on the empty theater of tuberculosis progress. In 1967, a woman born in India, acclaimed a great actress in such roles as Anna Karénina and Scarlett O'Hara (Gone with the Wind), died of tuberculosis. Starring at the time in a long-running play, Vivian Hartley (stage name Vivian Leigh) began to experience cough, then fever. Within a few months she was dead. The accidental irony was pure "theater." The play was Chekhov's Ivanov. Vivian Leigh, a wealthy woman dying of tuberculosis, was starring in a play about a wealthy woman dying of tuberculosis, written by a wealthy physician dying of tuberculosis. One may imagine that when art fails to imitate life, life may have no recourse but to imitate art.

Odysseys

In the early 1980s, I realized a boyhood dream, to climb an isolated peak called Mount Vaea, on the remote South Pacific island of Upolu, in what is now Western Samoa. Around that time, one of my patients, born on a Pacific island and admitted to my American hospital for tuberculosis, had walked up to the nursing station for some trivial purpose and in the middle of a conversation with the nurse, had suddenly opened his mouth to an explosion of arterial blood that shot across the room, overflowed like a waterfall to the linoleum floor, and caught him in its plumes as he sank downward and died—suffocated or drowned.

A century earlier, in 1880, coughing blood and dying more slowly, Robert Louis Stevenson fled his consumptive sentence. He escaped to Davos, to rugged sea-swept Speyside, and to sunny Marseille. All the while, he worked on the great books we still read all over the world, Treasure Island, Kidnapped—novels of escape.

Surviving a succession of hemorrhages, Stevenson climbed hills and sailed oceans, following the medical advice of his time. In New York, he stayed at Saranac Lake, America's most famous sanatorium. In San Francisco, he was farther from home than he had ever been. Still the hemorrhages and wasting continued. Escaping ever farther, he went to Hawaii, lived in a grass hut, and told stories to the King and Queen. He went on to Sydney, Australia, where he had additional hemorrhages, and at last to Upolu. Tuberculosis nevertheless continued to consume him. He wrote every day, suffering sudden hemorrhages, waiting to be "weaned from the passion of life."

The doctors said it was a stroke that felled him in December 1894. The 44-year-old Stevenson was placed on a bed in the middle of the room. Family and friends beside him, his Samoan "family" kneeling in a circle all around, the great Tusitala, "teller of tales," took a last breath. A council of chiefs was called, appointing 40 strong men to carry out Stevenson's final wish. They carried his body up to the top of the highest mountain and buried him in the place he had chosen.

Although I have visited the place several times—once dragging a famous epidemiologist barefoot up the mountain—it was the first time, 20 years ago, that I remember most clearly. I reflected then that in 1891, when Stevenson had escaped to Upolu, tuberculosis was still a disease of wealthy industrial nations. To run from it, Stevenson had to leave behind everything else in his life: family, friends, mother country, home, comfort. He fled to the very end of the earth.

Now, more than a century later, tuberculosis has escaped to the places where its victims once sought refuge, the one-time colonies of Western nations. The disease destroys the poor and underprivileged as it once destroyed the wealthy—95% of cases and 98% of deaths occur in the developing world. Its ravages are worst in the Pacific Islands. I wonder if the victims, who are too often beyond the reach of Western cures, have their own operas, poems, or plays to explain their predicament. I wonder where they escape. Do their flights, like Stevenson's, lead only to resignation and surrender?

On top of Mount Vaea, at 1,500 feet, a small promontory shows the blue Pacific in a broad sweep. Facing "The Magic Mountain" half a world away, you see a blue curtain above and below the horizon. At night, the sky is indeed wide and starry. The Southern Cross sparkles wildly, too low and too close to be celestial. The sarcophagus, placed at the highest point, is of white plaster, like the walls of the European hospitals and sanatoria Stevenson had fled. The plaque is of shiny brass, so that even in the moonlight, sometimes even in the tropical starlight alone, it is possible to read the words engraved on the side of the tomb. It is Stevenson's "Requiem," the poetry of his last resigned step in a long but unsuccessful flight.

Under the wide and starry sky,
Dig the grave and let me lie.
Glad did I live and gladly die,
And I laid me down with a will.

This be the verse you grave for me:
Here he lies where he longed to be;
Home is the sailor, home from the sea,
And the hunter home from the hill.

Acknowledgments

I thank Philip K. Wilson for assistance with research, for many helpful suggestions, and for his usual insights; and Stephen B. Greenberg, Ratna Soetjahja Morens, and the staff of the National Library of Medicine, National Institutes of Health, for additional research assistance.

Dr. Morens is professor of epidemiology and tropical medicine at the University of Hawaii School of Medicine. He is currently serving at the National Institute of Allergy and Infectious Diseases, National Institutes of Health. His research interests include viral disease pathogenesis, epidemiology, and the relationship between biomedical research and public health.

Sources

Barber S. Quartet, strings, Op. 11, B minor [musical score]. New York: G. Schirmer; 1939.

Barbier J, Carré M. Les contes d’Hoffmann, opéra en quatres actes [libretto]. Paris: Calmann Lévy; 1853.

Barnes DS. The making of a social disease: tuberculosis in nineteenth-century France. Berkeley (CA): University of California Press; 1995.

Bates B. Bargaining for life: a social history of tuberculosis, 1876–1938. Philadelphia: University of Pennsylvania Press; 1992.

Benoiston de Châteauneuf L-F. De l'influence de certaines professions sur le développement de la phthisie pulmonaire. Annales d'hygiène publique et de médecine légale 1831;6:5–49.

Bushnell OA. The gifts of civilization: germs and genocide in Hawaii. Honolulu: University of Hawaii Press; 1993.

Chekhov AP. Five plays [English translation]. Oxford: Oxford University Press; 1997.

Coleridge ST, Shelley PB, Keats J. The poetical works of Coleridge, Shelley, and Keats, complete in one volume. Philadelphia: J. Grigg; 1831.

Dubos R, Dubos J. The white plague: tuberculosis, man and society. Boston: Little, Brown & Co.; 1952.

Dumas A (fils). La dame aux camélias. Paris: D. Giraud et J. Dagneau; 1852.

Giacosa G, Illica L. La Bohème (Scène de la vie de Bohème de Henry Murger) [libretto]. Milano: Ricordi; 1896.

Guillaume de Machaut. Jugement deu roy de Navarre. In: Guillaume de Machaut. Les œuvres de Guillaume de Machault. Paris: Techener; 1849.

Hoffmann ETA. Rath Krespel. In: Hoffmann ETA. Die Serapion-brüder. Berlin:G. Reimer; 1821.

Hugo V-M. Notre-Dame de Paris. Paris: J. Olaye; 1831.

Hugo V-M. Les misérables. Paris: Pagnerre; 1862.

Hugo V-M. Les misérables [English translation]. New York: G.W. Carleton and Co.; 1880.

Hunt L. Lord Byron and some of his contemporaries: with recollections of the author’s life, and of his visit to Italy. 2nd ed. London: H. Colburn; 1828.

Jacobson AC. Genius: some revaluations. New York: Greenberg; 1926.

Keats J. The poetical works. New edition. London: E. Moxon; 1846.

Keats J. Letters of John Keats to Fanny Brawne, written in the years MDCCCXIX and MDCCCXX and now given from the original manuscripts with introduction and notes by Harry Buxton Forman. London: Reeves and Turner; 1878.

Keers RY. Pulmonary tuberculosis: a journey down the centuries. London: Ballière Tindall; 1978.

Knopf SA. Tuberculosis. In: Stedman TL, editor. Twentieth century practice: an international encyclopedia of modern medical science by leading authorities of Europe and America. Vol. XX. Tuberculosis, yellow fever, and miscellaneous. General index. New York: William Wood and Co.; 1900. p. 3–396.

Knopf SA. The social aspects of tuberculosis [pamphlet]. Publisher not indicated; 1908.

Koch R. Die Aetiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis. Beiträge zur Biologie der Pflanzen 1876;2:277–310,11 figures.

Louis P-C-A. Recherches anatomico-pathologiques sur la phthisie. Précédées du rapport fait à l'Académie royale de médecine par. MM. Bourdois, Royer-Collard et Chomel. Paris: Gabon et Cie; 1825.

MacMurray JM, editor. The book of 101 opera librettos. New York: Black Dog & Leventhal; 1996.

Main PA. The origin and history of tuberculosis. In: Covacevich J, Pearn J, Case D, Chapple I, Phillips G, editors. History, heritage and health: proceedings of the Fourth Conference of the Australian Society of the History of Medicine; 1995 July 2–9, Norfolk Island. Brisbane: The Society; 1996. p. 149–51.

Mann T. Der zauberberg. Berlin: S. Fischer; 1924.

Mikelova L. La Traviata and the New York Times, 1859 to 1985. Opera reviews and the cultural response to tuberculosis. In: Cruse P, editor. The proceedings of the Annual History of Medicine Days, Faculty of Medicine, University of Calgary. Calgary: University of Calgary; 1998. p. 229–36.

Mitchell M. Gone with the wind. New York: MacMillan; 1936.

Moorman LJ. Tuberculosis and genius. Chicago: University of Chicago Press; 1940.

Morse MM. Through the opera glass: a critical analysis of tuberculosis as presented in opera. Pharos 1998;61:11–14.

Murger L-H. Scènes de la [vie de] Bohême. Paris: Michel Lévy frères; 1851.

Ott K. Fevered lives: tuberculosis in American culture since 1870. Cambridge (MA): Harvard University Press; 1997.

Petrarcá F. Opera del preclarissimo poeta miser Francesco Petrarcha con li comenti sopra li Triumphi: Soneti: & Canzone historiate & nouamente corette per Miser Nicolo Perozone co molte acute & excellente additione. Milano: I. A. Scinzenzeler; 1512.

Piave FM [libretto], Verdi G [score]. La traviata; opera tragica [piano-vocal score]. Napoli: Stapilimento musicale Partenopeo; 1853.

Pope A. The works of Alexander Pope, esq. Edinburgh: J. Balfour; 1764.

Reibman J. Phthisis and the arts. In: Rom WM, Garay SM, editors. Tuberculosis. Boston: Little, Brown, and Co.; 1996:21–34.

Rosenktantz BG, editor. From consumption to tuberculosis. New York: Garland Publishing; 1994.

Rothman SM. Living in the shadow of death: tuberculosis and the social experience of illness in American history. New York: Basic Books; 1994.

Ryan F. The forgotten plague: how the battle against tuberculosis was won—and lost. Boston: Little, Brown & Co.;1992.

Shelley PB. The poetical works of Percy Bysshe Shelley. In four volumes. London: Edward Moxon;1839

Sontag S. Illness as metaphor. New York: Farrar, Straus & Giroux; 1978.

Stevenson RL. Ordered South. MacMillan’s Magazine. Volume XXX. 30 May 1874. p. 68–73.

Stevenson RL. Kidnapped, being memoirs of the adventures of David Balfour in the year 1751. London: Cassell; 1886.

Stevenson RL. Treasure Island. London: Cassell; 1883.

Stevenson RL. Underwoods. Book I. In English. London: Chatto and Winders; 1887.

Tolstoy LN. Anna Karénina [English translation]. New York: T.Y. Crowell & Co.; 1886.

Tolstoy LN. Anna Karenina. New York: Penguin Books; 1961.

Tomes N. The gospel of germs: men, women, and the microbe in the American life. Cambridge (MA): Harvard University Press; 1998.

Virchow R. Einige Bemerkungen über die von Herrn Dr. Hartsen aufgeworfene Frage über die Zulässigkeit des Geschlechtsgenusses bei Phthisiskern. Archiv für pathologische Anatomie und Physiologie und für klinische Medicin 1870;49:577–9.

Whitman W. Leaves of grass. Philadelphia: David McKay; 1884.

Address for correspondence: David Morens, NIH, NIAID, Room 3149, 6700-B Rockledge Drive, MSC 7630, Bethesda, MD 20892-7630, USA; fax:301 480 1594; e-mail: dm270q@nih.gov

Mycobacterium haemophilum: Emerging or Underdiagnosed in Brazil?

To the Editor: *Mycobacterium haemophilum* was first described in 1978 by Sompolinsky et al. (1) as the cause of cutaneous infections in a patient with Hodgkin disease. Since then, fewer than 100 cases have been reported worldwide, mostly among immunocompromised patients (2), although *M. haemophilum* infection has also been described in immunocompetent patients as the cause of cervical, submandibular, and perihilar lymphadenopathy in children and of pulmonary nodules in an adult (3–5). Cases have been reported from United States, Australia, Canada, France, Israel, and the United Kingdom, but to date no reports have originated in South America.

The most frequent clinical sign of *M. haemophilum* infection in adults is a skin or joint lesion. Less common sites for isolation of *M. haemophilum* include the respiratory tract, blood, bone marrow, bone, and central venous catheters (2,6). *M. haemophilum* is unique among *Mycobacterium* species owing to its special growth requirements: it grows best at 30°C and requires an iron supplement (hemin or ferric ammonium citrate).

We report here the characterization of three strains of *M. haemophilum* isolated from patients living in three states in two distinct regions of Brazil, Rio de Janeiro and São Paulo (southeast region) and Bahia (northeast region). The first strain was detected in Rio de Janeiro in December 2000 from a blood culture of a 67-year-old man who had received a kidney transplant in 1988 at the age of 55 years and was undergoing immunosuppressive treatment with prednisolone and mycophenolate mofetil. The second strain was detected in São Paulo in March 2001 in a 43-year-old HIV-seropositive man from a biopsied specimen of a nasal ulcer. A direct acid-fast stain showed many acid-fast bacilli. At time of diagnosis, the patient's CD4+ cell count was 8/mm^3 and his viral load was 290.000 copies/mL. The third isolate was detected in Bahia in a 30-year-old HIV-seropositive man who had osteomyelitis in an elbow. A direct acid-fast stain showed rare acid-fast bacilli.

The isolate from the Rio de Janeiro patient grew only in Myco/F Lytic media (Becton Dickinson Microbiology Systems, Sparks, MD) plus blood in primary isolation and subculture; it failed to grow on chocolate agar at 30°C after 6 weeks. The isolates from São Paulo and Bahia showed a slight growth in 12B media on primary isolation; this growth was likely supported by the iron provided by the biopsied tissue. Subcultures on chocolate agar showed good growth after 2–3 weeks at 30°C. The isolates did not grow on Middlebrook 7H10 agar without hemin and grew on the same media when supplemented with 60 MM of hemin. Both strains showed a negative catalase reaction.

The species of all isolates was identified through polymerase chain reaction amplification of the gene encoding for the 65-kDa heat shock protein, followed by restriction analysis with the enzymes *Bst*EII and *Hae*III as described by Telenti et al. (7), with minor modifications. The three isolates showed the same restriction pattern as that obtained for *M. haemophilum* American Type Culture Collection 29548 prototype strain. Isolates from Rio de Janeiro and São Paulo were also molecularly characterized as previously described by Roth et al. (8), corroborating *M. haemophilum* species identification.

To our knowledge, these *M. haemophilum* isolates are the first to be reported in Brazil. These three patients came from cities 429–962 km apart, demonstrating the dispersion of *M. haemophilum* infection in Brazil. Given the specific requirements of *M. haemophilum* for its growth in culture, our findings suggest that its true incidence in Brazil is greatly underestimated. Consequently, we strongly recommend that clinical laboratories in Brazil include an iron-supplemented medium, such as chocolate agar, incubated at 30ºC, for primary isolation of *Mycobacterium* spp in samples from selected patients.

Jorge Luiz Mello Sampaio,*† Venâncio Avancini Ferreira Alves,‡ Sylvia Cardoso Leão,† Vanda Dolabela de Magalhães,§ Marinês Dalla Valle Martino,§ Caio Marcio Figueiredo Mendes,*Antonio Carlos de Oliveira Misiara,¶ Kozue Miyashiro,* Jacyr Pasternak,§ Eliana Rodrigues,‡ Ronaldo Rozenbaum,# Carlos Alberto Sant´Anna Filho, Sônia Regina Marques Teixeira,* Adriano Cunha Xavier,†† Mauro Silvério Figueiredo,* and José Paulo Gagliardi Leite††,‡‡**

*Fleury - Centers for Diagnostic Medicine, São Paulo, Brazil; †Federal University of São Paulo, São Paulo, Brazil; ‡Oswaldo Cruz Hospital, São Paulo, Brazil; §Albert Einstein Hospital, São Paulo, Brazil; ¶Sírio Libanês Hospital, São Paulo, Brazil; #Servidores do Estado Hospital, Rio de Janeiro, Brazil;**Aliança Hospital, Salvador, Bahia, Brazil; ††Lâmina Laboratory, Rio de Janeiro, Brazil; ‡‡Oswaldo Cruz Institute, Rio de Janeiro, Brazil

References

1. Sompolinsky D, Lagziel A, Naveh D, Yankilevitz T. *Mycobacterium haemophilum* sp. nov., a new pathogen of humans. Int J Syst Bacteriol 1978;28:67–75.
2. Shah MK, Sebti A, Kiehn E, Massarella SA, Sepkowitz KA. *Mycobacterium haemophilum* in immunocompromised patients. Clin Infect Dis 2001;33:330–7.
3. Samra Z, Kaufmann L, Zeharia A, Ashkenazi S, Amir J, Bahar J, et al. Optimal detection and identification of *Mycobacterium haemophilum* in specimens from pediatric patients with cervical lymphadenopathy. J Clin Microbiol 1999;37:832–4.
4. Armstrong KL, James RW, Dawson DJ, Francis PW, Masters B. *Mycobacterium haemophilum* causing perihilar or cervical lymphadenitis in healthy children. J Pediatr 1992;121:202–5.
5. White DA, Kiehn TE, Bondoc AY, Massarella SA. Pulmonary nodule due to *Mycobacterium haemophilum* in an immuno-

competent host. Am J Respir Crit Care Med 1999;160:1366–8.
6. Ward MS Lam KV, Cannell PK, Herrmann RP. Mycobacterial central venous catheter tunnel infection: a difficult problem. Bone Marrow Transplant 1999;24:325–9.
7. Telenti A, Marchesi F, Balz M, Bally F, Bottger EC, Bodmer T. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J Clin Microbiol 1993;31:175–8.
8. Roth A, Reischl U, Streubel A, Naumann L, Kroppenstedt RM, Habicht M, et al. Novel diagnostic algorithm for identification of mycobacteria using genus-specific amplification of the 16S-23S rRNA gene spacer and restriction endonucleases. J Clin Microbiol 2000;38:1094–104.

Address for correspondence: Jorge Luiz Mello Sampaio, Fleury–Centers for Diagnostic Medicine, Microbiology Section, Av. General Waldomiro de Lima 508, São Paulo, Brazil 04344-070; fax: 55 11 50147601; e-mail: jorge.sampaio@fleury.com.br.

Children and Multidrug-Resistant Tuberculosis in Mumbai (Bombay), India

To the Editor: India has the highest number of tuberculosis (TB) cases in the world. Each year in India, over 2 million new cases of TB are diagnosed, and approximately 500,000 persons die of the disease (1). During the last decade, multidrug-resistant TB has burgeoned in India, resulting in an extremely large number of multidrug-resistant TB cases, second only to the number of cases noted in Latvia (2). Since 1993, in response to this epidemic, the government of India has implemented the Revised National Tuberculosis Control Program, which is based on directly observed treatment (short course) principles (1).

Mumbai (formerly Bombay), India, is a densely populated metropolis with a population of approximately 12 million, 4.8 million (40%) of whom reside in overcrowded slums. Since 1990, a resurgence of TB has occurred, characterized by a 70% to 140% increase in the rate of TB-related deaths among adults aged 25–44 years (3). A vital factor contributing to this phenomenon is HIV infection. A recent review of autopsy reports from Mumbai showed that 85 (59%) of 143 adult patients with AIDS were diagnosed with pulmonary TB (4), indicating that the disease is the most common opportunistic infection for persons with AIDS. Commensurate with the increase in TB cases is a surge in the prevalence of multidrug-resistant TB in adult patients. Two reference mycobacterial laboratories in private hospitals in Mumbai have reported a high prevalence of multidrug-resistant TB strains; 56 (11%) of 521 cases in 1991–1995, and 58 (58%) of 100 cases in 1994–1995 (5,6).

The crisis of multidrug-resistant TB in adults in Mumbai has been well documented (7). However, little attention has been directed at children also affected by the resurgent TB epidemic. We think that TB is developing in more children in Mumbai today than a decade earlier. Moreover, close proximity to adult patients with multidrug-resistant TB makes children prone to developing primary multidrug-resistant TB, a vulnerability documented in a South African study (8). Similarly, disseminated TB is occurring in large numbers of children living in overcrowded slums in Mumbai with a consequent high death rate (9); we attribute many of these deaths in children to primary multidrug-resistant TB. However, this conclusion is difficult to document, as most affected children are sputum-negative for acid-fast bacilli. Contact tracing to detect the adult source of infection is routinely undertaken; often we can trace the source of infection. Because the facilities for culture and susceptibility testing are not available at affordable rates, proving that the adult contact has multidrug-resistant TB is not feasible in most cases.

The AIDS epidemic in adults in Mumbai has adversely affected the epidemic within the population of children with TB. HIV infection in young adults has resulted in a large number of HIV-infected infants, the result of a lack of any large-scale program aimed at preventing vertical transmission. To combat the growing problem with HIV-infected infants, India's National AIDS Control Organization is performing feasibility studies for implementing interventions to prevent mother-to-child transmission of HIV infection. Clinical trials with nevirapine are currently being conducted at five major public hospitals in Mumbai.

Multidrug-resistant TB frequently develops in adult AIDS patients (7). Accordingly, many pediatric AIDS patients in Mumbai are also developing primary multidrug-resistant TB. Since most families cannot afford antiretroviral therapy, HIV-infected children in whom TB is diagnosed are prescribed a four-drug TB treatment (consisting of isoniazid, rifampicin, pyrazinamide, and ethambutol) and co-trimoxazole for *Pneumocystis carinii* pneumonia prophylaxis. Although deaths in these children are being attributed to AIDS, we think that many of these deaths are related to multidrug-resistant TB.

To combat the TB epidemic, the Revised National Tuberculosis Control Program directly observed treatment strategy has been implemented as part of a public health program. However, most patients receive treatment from private physicians and thus remain outside the purview of the strategy. Private physicians seldom refer their patients to centers offering directly observed treatments because of potential for loss of income (3). In 1991, Uplekar and Shepard (10) reported that 100 private physicians in the Dharavi slums in Mumbai prescribed 80 different anti-TB regimens; most were both inappropriate and expensive. Since private physicians have not yet been involved in the gov-

ernment-run Revised National Tuberculosis Control Program, the situation today remains the same.

Although children are included in the national control program, they do not receive the benefit of directly observed treatment strategy. The Revised National Tuberculosis Control Program does not provide drugs in syrup form or permit breaking of tablets, making the administration of accurate pediatric doses impossible. Most children with TB are also sputum-smear negative for acid-fast bacilli. Doctors must rely on clinical acumen when deciding whether or not to start TB treatment. This lack of a method for definitive diagnosis of TB in children makes treatment centers reluctant to enlist pediatric cases; as a result, these children attend general pediatric outpatient clinics every 28 days to obtain their TB medication. Directly observed treatment strategy is not followed in the general outpatient clinics. Hence, compliance with treatment depends on the motivation and perseverance of the parents. Frequently, one or more of the drugs is out of stock, and parents must use their own small resources to purchase the necessary medication. To avoid long waits in the crowded general pediatric outpatient clinics, some parents intermittently purchase the anti-TB drugs from local chemists, who supply the drugs without a current prescription. This practice leads to frequent defaulting and inadequate treatment.

Since children in general have paucibacillary TB, secondary multidrug-resistant TB is considered less likely to develop in them, even when the treatment is inadequate. However, Karande et al. (11) describe a 12-year-old boy in Mumbai with secondary multidrug-resistant TB. He had received multiple courses of inadequate treatment with various anti-TB treatment regimens for 9 years. The TB gradually progressed in severity and was disseminated with the bacterial load increasing sufficiently for multidrug-resistant TB to develop. We suggest that this case is not unusual and that many children in Mumbai are dying of multidrug-resistant TB because directly observed treatment strategy regimens are unavailable.

The multidrug-resistant TB crisis on Mumbai's children warrants immediate attention and action. We suggest that the directly observed treatment strategy should be made child-friendly with anti-TB drugs made available in suitable pediatric formulations. Private physicians require education and involvement in the treatment strategy. BACTEC (BD Diagnostic Systems, Sparks, MD) culture and susceptibility testing to detect multidrug-resistant TB should be made available at affordable rates. Transmission of HIV to newborns would be reduced by universally implementing a prevention program for mother-to-child transmission at subsidized rates. Immediate action on these suggestions will lower the incidence of both TB and multidrug-resistant TB and reduce the number of deaths from these diseases.

Sunil Karande*† and Sandeep B. Bavdekar‡§

*Lokmanya Tilak Municipal Medical College, Mumbai, India, †Lokmanya Tilak Municipal General Hospital, Mumbai, India, ‡Seth Gordhandas Sunderdas Medical College, Mumbai, India, and §King Edward VII Memorial Hospital, Mumbai, India

References

1. Agarwal SP. TB across the globe (2). Tuberculosis in India—the past and prospects for the future. Scott Med J 2000;45:11–3.
2. Pablos-Mendez A, Raviglione MC, Laszlo A, Binkin N, Rieder HL, Bustreo F, et al. Global surveillance for antituberculosis-drug resistance, 1994–1997. World Health Organization– International Union against Tuberculosis and Lung Disease Working Group on Anti-Tuberculosis Drug Resistance Surveillance. N Engl J Med 1998;338:1641–9.
3. Hira SK, Srinivas Rao AS, Thanekar J. Evidence of AIDS-related mortality in Mumbai, India. Lancet 1999;354:1175–6.
4. Lanjewar DN, Duggal R. Pulmonary pathology in patients with AIDS: an autopsy study from Mumbai. HIV Med 2001;2:266–71.
5. Chowgule RV, Deodhar L. Pattern of secondary acquired drug resistance to antituberculosis drug in Mumbai, India—1991–1995. Indian J Chest Dis Allied Sci 1998;40:23–31.
6. Udwadia ZF, Hakimiyan A, Rodrigues C, Jillisgar T, Mehta A. A profile of drug-resistant tuberculosis in Bombay. Chest 1996;110:228.
7. Udwadia ZF. India's multidrug-resistant tuberculosis crisis. Ann N Y Acad Sci 2001;953:98–105.
8. Schaaf HS, Gie RP, Kennedy M, Beyers N, Hesseling PB, Donald PR. Evaluation of young children in contact with adult multidrug-resistant pulmonary tuberculosis: a 30-month follow-up. Pediatrics 2002;109:765–71.
9. Karande S, Bhalke S, Kelkar A, Ahuja S, Kulkarni M, Mathur M. Utility of clinically-directed selective screening to diagnose HIV infection in hospitalized children in Bombay, India. J Trop Pediatr 2002;48:149–55.
10. Uplekar MW, Shepard DS. Treatment of tuberculosis by private general practitioners in India. Tubercle 1991;72:284–90.
11. Karande S, Kelkar A, Jagiasi A, Kulkarni M. Acquired multidrug-resistant tuberculosis in an immunocompetent adolescent. Pediatr Infect Dis J 2002;21:577–8.

Address for correspondence: Dr. Sandeep B. Bavdekar, A2-9, Worli Seaside Cooperative Housing Society, Khan Abdul Gaffar Khan Road, Worli, Mumbai 400 018, India; fax: 91-22-414 34 35; e-mail: drsbavdekar@vsnl.com

Certificate of Knowledge in Travel Medicine Examination

The International Society of Travel Medicine (ISTM) will offer its first international Certificate of Knowledge in Travel Medicine examination on May 7, 2003, before the opening of the 8th ISTM Conference in New York City, on May 7–11. Those passing the exam will receive a Certificate in Travel Health. The exam is open to all licensed travel medicine practitioners, including physicians, nurses, and pharmacists.

To obtain more information on preparation for taking the exam, registration, and a Candidate Bulletin of Information, please access the ISTM website at www.istm.org. Applicants may also contact Brenda Bagwell, ISTM Secretariat, P.O. Box 871089, Stone Mountain, GA, 30087-0028. USA. Telephone: 1-770-736-7060; fax: 1-770-736-6732; e-mail: exam@istm.org

Cestode Zoonoses: Echinococcosis and Cysticercosis: An Emergent and Global Problem

Philip Craig and
Zbigniew Pawlowski, editors

Vol. 341 NATO Science Series,
IOS Press, Amsterdam
410 pages, hardcover
ISBN: 1-58603-220-8
Price: US $100

This book is a collection of short articles written by the participants of a research workshop held in Poznan, Poland, in September 2000. The workshop, supported by NATO Scientific Affairs, focused on the three major larval cestode diseases of humans: *Taenia solium* neurocysticercosis, *Echinococcus granulosus* cystic hydatidosis, and *E. multilocularis* alveolar hydatidosis. The format and depth of the articles are variable, but readers familiar with these parasites will find the book to be a convenient collection of new information on the subject. A shortcoming is that the book's preface and summary are each limited to a single page.

Perhaps most interesting for readers of the Emerging Infectious Diseases Journal are the reviews of epidemiologic data related to the emergence or reemergence of these three diseases. In sub-Saharan Africa, for example, neurocysticercosis has emerged as being more widely distributed than previously assumed and is a major cause of epilepsy. Surgery for pediatric cystic echinococcosis in Kyrgystan increased threefold during the period 1993–1998 (reaching 6 cases/100,000), suggesting new transmission probably related to worsening economic conditions after the collapse of the former Soviet Union. Surveillance for human cases of alveolar echinococcosis (which can have a mortality rate of 90% if untreated) is being strengthened in western Europe, given that *E. multilocularis* infection rates in foxes have increased in recent years. The book contains other valuable updates on diagnostics, immunology and vaccines, imaging and clinical management, geographic information systems and ecology, veterinary medicine, and community-based control programs. Readers with an interest in helminthology will find this book most useful.

Frank O. Richards, Jr.

The Carter Center, Atlanta, Georgia, USA

Address for correspondence: Frank O. Richards, Jr., The Carter Center, 453 Freedom Parkway, Atlanta, GA 30307, USA; fax: 770-420-5100; e-mail: sdsulli@emory.edu

Immunology of Infectious Diseases

By Stefan H.E. Kaufmann,
Alan Sher, and Rafi Ahmed

American Society for
Microbiology Press
Washington, D.C.
520 pages
ISBN: 1-55581-214-7
Price: US $115.95

Whether an infectious disease agent is an "old acquaintance" or a new, emerging threat, the immune system's battle against it is usually the first line of defense it encounters. With vaccines and effective treatments often unavailable, the immune system's efforts to eradicate infectious agents or infected cells are frequently the only means to combat them. Understanding the immune system—as well as the infectious agent's tactics to undermine it—is of vital importance to the researcher and clinician. This textbook attempts to provide just this information.

Immunology of Infectious Diseases is a textbook in the best sense of the word, presenting its contents in a clear, structured manner. Instead of encylopedic coverage of every infectious disease agent known, a set of paradigmatic infections were selected on the basis of the depth of available knowledge. The book is divided into eight sections, each of which addresses a particular aspect of the host-infectious agent interaction, describing it in separate chapters for bacteria, fungi, parasitic eukaryotes, and viruses. So instead of discussing all aspects of viral diseases, the reader learns about the innate immune response to the various pathogens, chapter by chapter, in the respective section. Emphasis is thereby placed on the immune system's "point of view" about an infectious process, rather than on the microbe's.

After an introduction to the various classes of infectious disease agents, the book describes the immune responses directed against the different types of infections, proceeding from the innate to the acquired (adaptive). Discussion of the pathology of infections not eradicated by the immune system early on and the cunning strategies of the infectious microbes to evade immune attacks is followed by sections on immunogenetics and exploration of the immune system's interventions against two high-incidence infections, tuberculosis and AIDS.

Although the infections discussed in this book are not emerging ones in the strictest sense, the example of AIDS shows just how fast an infectious disease that was emerging, seemingly restricted to a subset of the population only two decades ago, can grow into a pandemic in a highly mobile, dense population at the end of the 20th century. Even tuberculosis, the "wasting disease" dreaded by our grandfathers' generation, which scientists believed to be under control, can be regarded as an emerging disease: *Mycobacterium tuberculosis* has stepped into the limelight again in the wake of HIV, which renders a growing number of people immunocompromised.

As infection and immune reaction are so intricately intertwined, this book is valuable reading to anyone interested in infectious diseases in humans. Maybe in the future prions will have to be included as a new type of infectious agent whose rise we are just now witnessing. (Information on prion pathology is still hotly debated, and data on routes of transmission and immune system reactions are still scarce.)

The book is a handy and manageable length. In contrast to many standard textbooks of immunology, Immunology of Infectious Diseases is text oriented. Except for a central insert of color plates, figures are in black and white only. Each chapter contains an alphabetical list of references. Because of its accessible modular structure, this textbook is easy to navigate, rendering it easy to use. The book is suitable for anyone with a background in cell biology and basic immunology. Advanced undergraduate students and postgraduates with a grasp of the main groups of leukocytes and immune effector mechanisms, as well as specialists in other subdisciplines, will find this textbook to be a highly useful and readable introduction to the immune system's mainstay, the battle against infection.

Thomas Lazar

Institute of Transfusion Medicine, Humboldt University Medical School, Berlin, Germany

Address for correspondence: Thomas Lazar, Institute of Transfusion Medicine (ITM), Charité Humboldt University Medical School, Augustenburger Platz 1, Berlin, Germany 13353; e-mail: tlazar@web.de

Correction, Vol. 8, No. 7

In Smallpox Research Activities: U.S. Interagency Collaboration, 2001 by James LeDuc et al., an error occurred in the text on Page 744, left column, under Diagnostic Tests, line 25. The information provided between lines 10 and 25 should be attributed to S. Ibrahim, U.S. Medical Research Institute of Infectious Diseases as unpublished data.

The corrected article appears online at http//www.cdc.gov/ncidod/eid/vol8no7/02-0032.htm

We regret the omission.

EMERGING INFECTIOUS DISEASES

Full text free online at www.cdc.gov/eid

The print journal is available at **no charge** to public health professionals

☐ YES, I would like to receive Emerging Infectious Diseases.

Please print your name and business address in the box and return by fax to 404-371-5449 or mail to
EID Editor
CDC/NCID/MS D61
1600 Clifton Road, NE
Atlanta, GA 30333

Moving? Please give us your new address (in the box) and print the number of your old mailing label here______________

EID Online
www.cdc.gov/eid

About the Cover

Amedeo Modigliani (1884-1920). Self-Portrait, 1919.
Oil on canvas 100 cm x 65 cm.
Museu de Arte Contemporanea da Universidade de Sao Paulo, Brazil

Modigliani was born in Livorno, Italy, where he grew up in a Jewish ghetto. He studied art in Florence, and in 1906 he moved to Paris, where he met Pablo Picasso and other leading artists of his era. In Paris, he was influenced by fauvism, the avant-garde art movement promoting a strong, emotional, and nonrealistic use of color, and by his friend the Romanian sculptor Constantin Brancusi, known for his artistic search of pure form. Modigliani was also influenced by African carvings and masks, particularly in his early work, which was mostly sculpture (1).

In his brief life, which even in childhood was marked by ill health, Modigliani was able to grow as an artist and attain his own distinctive style. He is known for his graceful, simplified, and sympathetic portrayal of the human form. His paintings, mostly portraits and studies of the human figure, are characterized by fine sinuous lines and have a simple, spare, and flat appearance, which gives them an almost classical effect. The figures are elongated, the faces oval, and the shapes ethereal, reminiscent at times of Sandro Botticelli (see cover Vol.7, No.3, Emerging Infectious Diseases). The portraits (more than 200 from 1916 to 1919), unburdened by detail, rely on color and shape for emotional and psychological insight and emit a "curious sense of pathos" (1).

Modigliani's fauvist contemporaries had moved away from the conventional and sentimental in art. They were not interested in the representation of observed reality or even in passion mirrored on a face. They were after "radical simplicity," the "genius of omission." Expression to them was achieved through form and spatial depth, the arrangement of line and color on a flat plane and the empty spaces around them (2). In this artistic climate, Modigliani would not have been interested in tuberculosis as a subject for his art, nor would he have painted a conventional portrait of this disease that consumed his adult life and eventually killed him at age 36.

The famous self-portrait on this cover of Emerging Infectious Diseases, painted in 1919, just one year before Modigliani's death, inspired writers who were captivated by its romanticism to speculate broadly about its meaning. Even though many of the interpretations are mostly conjecture, the length and thinness of the face, as well as its pallor and eerie calmness, may well be due to tuberculosis (3).

In this portrait, so reminiscent of the African masks that had fascinated him not for their intense expression but for their formal simplicity and coherence, Modigliani seems to have captured the essence of his subject, himself. He turned to his fauvist roots for the striking hues so typical of tuberculous complexion, to his friend Brancusi's sculptures for the studied serenity, and to his own emotional capacity for the depths of darkness welled in those stylized eyes. But whether he intended it or not, the master portrait painter, Modigliani, in this self-portrait of hollowed cheeks and sealed lips, painted more than his face. He painted the face of tuberculosis.

Polyxeni Potter

1. Amedeo Modigliani—Self portrait: famous art reproductions 2002 [cited 2002 September 17]. Available from URL: http.www.famousartreproductions.com/modiglianibiography.html
2. Janson HW, Janson AF. History of art. New York: Harry N. Abrams, Inc.; 2001.
3. Chretien J. Tuberculosis: the illustrated history of a disease. Hauts-de-France; 1998.

EMERGING INFECTIOUS DISEASES
A Peer-Reviewed Journal Tracking and Analyzing Disease Trends Vol.8, No.12 December 2002

Upcoming Issue

For a complete list of articles included in the December issue, and for articles published online ahead of print publication, see http://www.cdc.gov/ncidod/eid/upcoming.htm

Look in the December issue for the following topics:

Outpatient Antibiotic Use and Prevalence of Antibiotic-Resistant Pneumococci in France and Germany: Sociocultural Perspective

Dengue Hemorrhagic Fever in Infants: Research Opportunities Ignored

Antimicrobial Resistance in *Streptococcus pneumoniae*, Taiwan

Vector Competence of California Mosquitoes for *West Nile virus*
Efficacy of Killed Virus Vaccine, Live Attenuated Chimeric Virus Vaccine, and Passive Immunization for Prevention of West Nile Virus Encephalitis in Hamster Model

Mass Vaccination Campaign Following a Community Outbreak of Meningococcal Disease

Antimicrobial Resistance of *Escherichia coli* O26, O103, O111, O128, and O145 from Animals and Humans

Using Automated Health Plan Data to Assess Infection Risk from Coronary Artery Bypass Surgery

Rift Valley Fever Outbreak, Kingdom of Saudi Arabia: Isolation and Genetic Characterization of the virus from *Aedes vexans arabiensis*

Upcoming Infectious Disease Conferences

November 8-9, 2002

EMBL/EMBL Science and Society Infectious Diseases: Challenges, Threats and Responsibilities
Heidelberg, Germany
E-mail: conferences@embl-heidelberg.de
Website: http://www.embl-heidelberg.de/Conferences/SciSoc02/

November 9-13, 2002

American Public Health Association 130th Annual Meeting and Exposition
Philadelphia, PA
Contact: APHA Convention Services (202) 780-5600
Website: http://www.apha.org/meetings

November 9-10, 2002

American Society of Tropical Medicine and Hygiene (ASTMD) Courses and Meetings 2002
Updates in Wilderness and Extreme Medicine
Adam's Mark Hotel, Denver, Colorado
Contact: ASTMH (847) 480-9592
E-mail: astmh@astmh.org
Website: http://www.astmh.org

November 10-14, 2002

American Society of Tropical Medicine and Hygiene (ASTMD) Courses and Meetings 2002
American Society of Tropical Medicine and Hygiene
51st Annual Meeting
Adam's Mark Hotel, Denver, Colorado
Contact: ASTMH (847) 480-9592
E-mail: astmh@astmh.org
Website: http://www.astmh.org

November 19-23, 2002

3rd World Congress of Pediatric Infectious Diseases - WSPID
Santiago, Chile
Website: http://www.kenes.com/wspid

January 26-30, 2003

National Hepatitis Coordinator Conference
San Antonio, TX
Contact: Valerie Curry (404) 371-5365
Website: http://www.cdc.gov/ncidod/diseases/hepatitis/coordinators/index.htm

January 28-29, 2003

2003 Symposium on Statistical Methods "Study Design and Decision Making in Public Health"
Atlanta, GA
Contact: Ram B. Jain (404)639-8867 or Betsy Cadwell (404)639-8693
Website: http://apps.nccd.cdc.gov/DRHAbstract2/

EMERGING INFECTIOUS DISEASES www.cdc.gov/eid

JOURNAL BACKGROUND AND GOALS

What are "emerging" infectious diseases?

Infectious diseases whose incidence in humans has increased in the past 2 decades or threatens to increase in the near future have been defined as "emerging." These diseases, which respect no national boundaries, include

- New infections resulting from changes or evolution of existing organisms.
- Known infections spreading to new geographic areas or populations.
- Previously unrecognized infections appearing in areas undergoing ecologic transformation.
- Old infections reemerging as a result of antimicrobial resistance in known agents or breakdowns in public health measures.

Why an "Emerging" Infectious Diseases journal?

The Centers for Disease Control and Prevention (CDC), the agency of the U.S. Public Health Service charged with disease prevention and health promotion, leads efforts against emerging infections, from AIDS, hantavirus pulmonary syndrome, and avian flu, to tuberculosis and *West Nile virus* infection. CDC's efforts encompass improvements in disease surveillance, the public health infrastructure, and epidemiologic and laboratory training.

Emerging Infectious Diseases represents the scientific communications component of CDC's efforts against the threat of emerging infections. However, even as it addresses CDC's interest in the elusive, continuous, evolving, and global nature of these infections, the journal relies on a broad international authorship base and is rigorously peer-reviewed by independent reviewers from all over the world.

What are the goals of Emerging Infectious Diseases?

1) Recognition of new and reemerging infections and understanding of factors involved in disease emergence, prevention, and elimination. Toward this end, the journal

   - Investigates factors known to influence emergence: microbial adaptation and change, human demographics and behavior, technology and industry, economic development and land use, international travel and commerce, and the breakdown of public health measures.
   - Reports laboratory and epidemiologic findings within a broader public health perspective.
   - Provides swift updates of infectious disease trends and research: new methods of detecting, characterizing, or subtyping pathogens; developments in antimicrobial drugs, vaccines, and prevention or elimination programs; case reports.

2) Fast and broad dissemination of reliable information on emerging infectious diseases. Toward this end, the journal

   - Publishes reports of interest to researchers in infectious diseases and related sciences, as well as to public health generalists learning the scientific basis for prevention programs.
   - Encourages insightful analysis and commentary, stimulating global interest in and discussion of emerging infectious disease issues.
   - Harnesses electronic technology to expedite and enhance global dissemination of emerging infectious disease information.

Editorial Policy and Call for Articles

Emerging Infectious Diseases is a peer-reviewed journal established expressly to promote the recognition of new and reemerging infectious diseases around the world and improve the understanding of factors involved in disease emergence, prevention, and elimination.

The journal has an international scope and is intended for professionals in infectious diseases and related sciences. We welcome contributions from infectious disease specialists in academia, industry, clinical practice, and public health, as well as from specialists in economics, demography, sociology, and other disciplines. Inquiries about the suitability of proposed articles may be directed to the Editor at 404-371-5329 (tel), 404-371-5449 (fax), or eideditor@cdc.gov (e-mail).

Emerging Infectious Diseases is published in English and features the following types of articles: Perspectives, Synopses, Research Studies, Policy and Historical Reviews, Dispatches, Commentaries, Another Dimension, Letters, Book Reviews, and News and Notes. The purpose and requirements of each type of article are described in detail below. To expedite publication of information, we post journal articles on the Internet as soon as they are cleared and edited.

Chinese, French, and Spanish translations of some articles can be accessed through the journal's home page at http://www.cdc.gov/eid.

Instructions to Authors

Manuscript Preparation. For word processing, use MS Word. Begin each of the following sections on a new page and in this order: title page, keywords, abstract, text, acknowledgments, biographical sketch, references, tables, figure legends, appendixes, and figures. Each figure should be in a separate file.

Title Page. Give complete information about each author (i.e., full name, graduate degree(s), affiliation, and the name of the institution in which the work was done). Clearly identify the corresponding author and provide that author's mailing address (include phone number, fax number, and e-mail address). Include separate word counts for abstract and text.

Keywords. Include up to 10 keywords; use terms listed in Medical Subject Headings Index Medicus.

Text. Double-space everything, including the title page, abstract, references, tables, and figure legends. Printed manuscript should be single-sided, beginning with the title page. Indent paragraphs; leave no extra space between paragraphs. After a period, leave only one space before beginning the next sentence. Use 12-point Times New Roman font and format with ragged right margins (left align). Italicize (rather than underline) scientific names when needed.

Biographical Sketch. Include a short biographical sketch of the first author—both authors if only two. Include affiliations and the author's primary research interests.

References. Follow Uniform Requirements (www.icmje.org/index.html). Do not use endnotes for references. Place reference numbers in parentheses, not superscripts. Number citations in order of appearance (including in text, figures, and tables). Cite personal communications, unpublished data, and manuscripts in preparation or submitted for publication in parentheses in text. Consult List of Journals Indexed in Index Medicus for accepted journal abbreviations; if a journal is not listed, spell out the journal title. List the first six authors followed by "et al." Do not cite references in the abstract.

Tables and Figures. Create tables within MS Word's table tool. Do not format tables as columns or tabs. Send graphics in native, high-resolution (200 dpi minimum) .TIF (Tagged Image File), or .EPS (Encapsulated Postscript) format. Graphics should be in a separate electronic file from the text file. For graphic files, use Arial font. Convert Macintosh files into the suggested PC format. Figures, symbols, letters, and numbers should be large enough to remain legible when reduced. Place figure keys within the figure. For more information see EID Style Guide (http://www.cdc.gov/ncidod/EID/style_guide.htm).

Manuscript Submission. Include a cover letter indicating the proposed category of the article (e.g., Research, Dispatch) and verifying that the final manuscript has been seen and approved by all authors. Submit an electronic copy (by e-mail) to the Editor, eideditor@cdc.gov.

Manuscript Types

Perspectives. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author. Articles in this section should provide insightful analysis and commentary about new and reemerging infectious diseases and related issues. Perspectives may also address factors known to influence the emergence of diseases, including microbial adaptation and change, human demographics and behavior, technology and industry, economic development and land use, international travel and commerce, and the breakdown of public health measures. If detailed methods are included, a separate section on experimental procedures should immediately follow the body of the text.

Synopses. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author—both authors if only two. This section comprises concise reviews of infectious diseases or closely related topics. Preference is given to reviews of new and emerging diseases; however, timely updates of other diseases or topics are also welcome. If detailed methods are included, a separate section on experimental procedures should immediately follow the body of the text.

Research Studies. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch of first author—both authors if only two. Report laboratory and epidemiologic results within a public health perspective. Although these reports may be written in the style of traditional research articles, they should explain the value of the research in public health terms and place the findings in a larger perspective (i.e., "Here is what we found, and here is what the findings mean").

Policy and Historical Reviews. Articles should be under 3,500 words and should include references, not to exceed 40. Use of subheadings in the main body of the text is recommended. Photographs and illustrations are encouraged. Provide a short abstract (150 words) and a brief biographical sketch. Articles in this section include public health policy or historical reports that are based on research and analysis of emerging disease issues.

Dispatches. Articles should be 1,000–1,500 words and need not be divided into sections. If subheadings are used, they should be general, e.g., "The Study" and "Conclusions." Provide a brief abstract (50 words); references (not to exceed 15); figures or illustrations (not to exceed two); and a brief biographical sketch of first author—both authors if only two. Dispatches are updates on infectious disease trends and research. The articles include descriptions of new methods for detecting, characterizing, or subtyping new or reemerging pathogens. Developments in antimicrobial drugs, vaccines, or infectious disease prevention or elimination programs are appropriate. Case reports are also welcome.

Commentaries. Thoughtful discussions (500–1,000 words) of current topics. Commentaries may contain references but should not include figures or tables.

Another Dimension. Thoughtful essays, short stories, or poems on philosophical issues related to science, medical practice, and human health. Topics may include science and the human condition, the unanticipated side of epidemic investigations, or how people perceive and cope with infection and illness. This section is intended to evoke compassion for human suffering and to expand the science reader's literary scope. Manuscripts are selected for publication as much for their content (the experiences they describe) as for their literary merit.

Letters. This section includes letters that present preliminary data or comment on published articles. Letters (500–1,000 words) should not be divided into sections, nor should they contain figures or tables. References (not more than 10) may be included.

Book Reviews. Short reviews (250–500 words) of recently published books on emerging disease issues are welcome. The name of the book, publisher, and number of pages should be included.

Announcements. We welcome brief announcements (50–150 words) of timely events of interest to our readers. (Announcements may be posted on the journal Web page only, depending on the event date.)

Conference Summaries. Summaries (500–1,000 words) of emerging infectious disease conferences may provide references to a full report of conference activities and should focus on the meeting's content rather than on individual conference participants.