Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak

Jennifer L. Gardy, Ph.D., James C. Johnston, M.D., Shannan J. Ho Sui, Ph.D.,
Victoria J. Cook, M.D., Lena Shah, M.Sc., Elizabeth Brodkin, M.D.,
Shirley Rempel, R.N., Richard Moore, Ph.D., Yongjun Zhao, D.V.M.,
Robert Holt, Ph.D., Richard Varhol, M.Sc., Inanc Birol, Ph.D., Marcus Lem, M.D.,
Meenu K. Sharma, Ph.D., Kevin Elwood, M.D., Steven J.M. Jones, Ph.D.,
Fiona S.L. Brinkman, Ph.D., Robert C. Brunham, M.D.,
and Patrick Tang, M.D., Ph.D.

ABSTRACT

From the British Columbia Centre for Disease Control (J.L.G., J.C.J., V.J.C., L.S., S.R., K.E., R.C.B., P.T.); University of British Columbia (J.L.G., V.J.C., K.E., R.C.B., P.T.); Canada's Michael Smith Genome Sciences Centre; (R.M., Y.Z., R.H., R.V., I.B., S.J.M.J.), Canadian Field Epidemiology Program, Public Health Agency of Canada (L.S.); and Health Canada (M.L.) — all in Vancouver; Simon Fraser University, Burnaby, BC (S.J.H.S., S.J.M.J., F.S.L.B.); Fraser Health Authority, Surrey, BC (E.B.), and National Microbiology Laboratory, Winnipeg, MB (M.K.S.) — all in Canada. Address reprint requests to Dr. Tang at British Columbia Centre for Disease Control, 655 W. 12th Ave., Vancouver, V5Z 4R4, Canada, or at patrick.tang@bccdc.ca.

Drs. Gardy, Johnston, and Ho Sui contributed equally to this article.

BACKGROUND

An outbreak of tuberculosis occurred over a 3-year period in a medium-size community in British Columbia, Canada. The results of mycobacterial interspersed repetitive unit–variable-number tandem-repeat (MIRU-VNTR) genotyping suggested the outbreak was clonal. Traditional contact tracing did not identify a source. We used whole-genome sequencing and social-network analysis in an effort to describe the outbreak dynamics at a higher resolution.

METHODS

We sequenced the complete genomes of 32 *Mycobacterium tuberculosis* outbreak isolates and 4 historical isolates (from the same region but sampled before the outbreak) with matching genotypes, using short-read sequencing. Epidemiologic and genomic data were overlaid on a social network constructed by means of interviews with patients to determine the origins and transmission dynamics of the outbreak.

RESULTS

Whole-genome data revealed two genetically distinct lineages of *M. tuberculosis* with identical MIRU-VNTR genotypes, suggesting two concomitant outbreaks. Integration of social-network and phylogenetic analyses revealed several transmission events, including those involving "superspreaders." Both lineages descended from a common ancestor and had been detected in the community before the outbreak, suggesting a social, rather than genetic, trigger. Further epidemiologic investigation revealed that the onset of the outbreak coincided with a recorded increase in crack cocaine use in the community.

CONCLUSIONS

Through integration of large-scale bacterial whole-genome sequencing and social-network analysis, we show that a socioenvironmental factor — most likely increased crack cocaine use — triggered the simultaneous expansion of two extant lineages of *M. tuberculosis* that was sustained by key members of a high-risk social network. Genotyping and contact tracing alone did not capture the true dynamics of the outbreak. (Funded by Genome British Columbia and others.)

MYCOBACTERIUM TUBERCULOSIS IS AN IMportant infectious disease even in developed countries with extensive control programs. This is the case in British Columbia, Canada, where the 2007 incidence rate of 6.4 cases per 100,000 population exceeded the national average of 4.7 cases per 100,000 population.[1]

In May 2006, a case of smear-negative pleural tuberculosis was diagnosed in an adult in a medium-size community in British Columbia. A second case, manifested as disseminated tuberculosis, was reported in an infant in July 2006. Reverse contact tracing identified nine additional cases between August and October 2006, when the British Columbia Centre for Disease Control (BCCDC) initiated an epidemiologic investigation.

Cases were identified throughout the remainder of 2006 and peaked in 2007, resulting in a 2007 incidence rate of 72 cases per 100,000 population in the region. By December 31, 2008, a total of 41 cases of tuberculosis had been identified, with all cultured isolates having an identical pattern of mycobacterial interspersed repetitive unit–variable-number tandem repeats (MIRU-VNTRs). Traditional contact tracing and social-network analysis were used to identify key persons, places, and behaviors that contributed to the spread of the outbreak. Epidemiologic data were later complemented with whole-genome sequencing of 36 *M. tuberculosis* isolates from the region.

METHODS

CASE DEFINITION AND OUTCOMES

Laboratory-confirmed cases of tuberculosis were defined as those with *M. tuberculosis* complex present on culture. Clinical cases were defined as those for which *M. tuberculosis* or other pathogens could not be cultured but that were characterized by a radiologic, pathologic, or therapeutic response that was consistent with active tuberculosis.[2] Of the 41 cases, 37 were laboratory confirmed, and 4 were clinical.

Therapeutic outcomes were recorded and classified according to standardized treatment outcomes of the World Health Organization.[3] All isolates were susceptible to all first-line antituberculosis antibiotics.

EPIDEMIOLOGIC INVESTIGATION

Case finding was carried out from May to October 2006 by local nurses with support from the BCCDC. Traditional contact-tracing approaches were used, in accordance with Canadian guidelines.[2] In October 2006, two BCCDC investigators were used to assist local health authorities in investigation of the outbreak. The work was deemed part of public health's response to the outbreak and was not subject to institutional review.
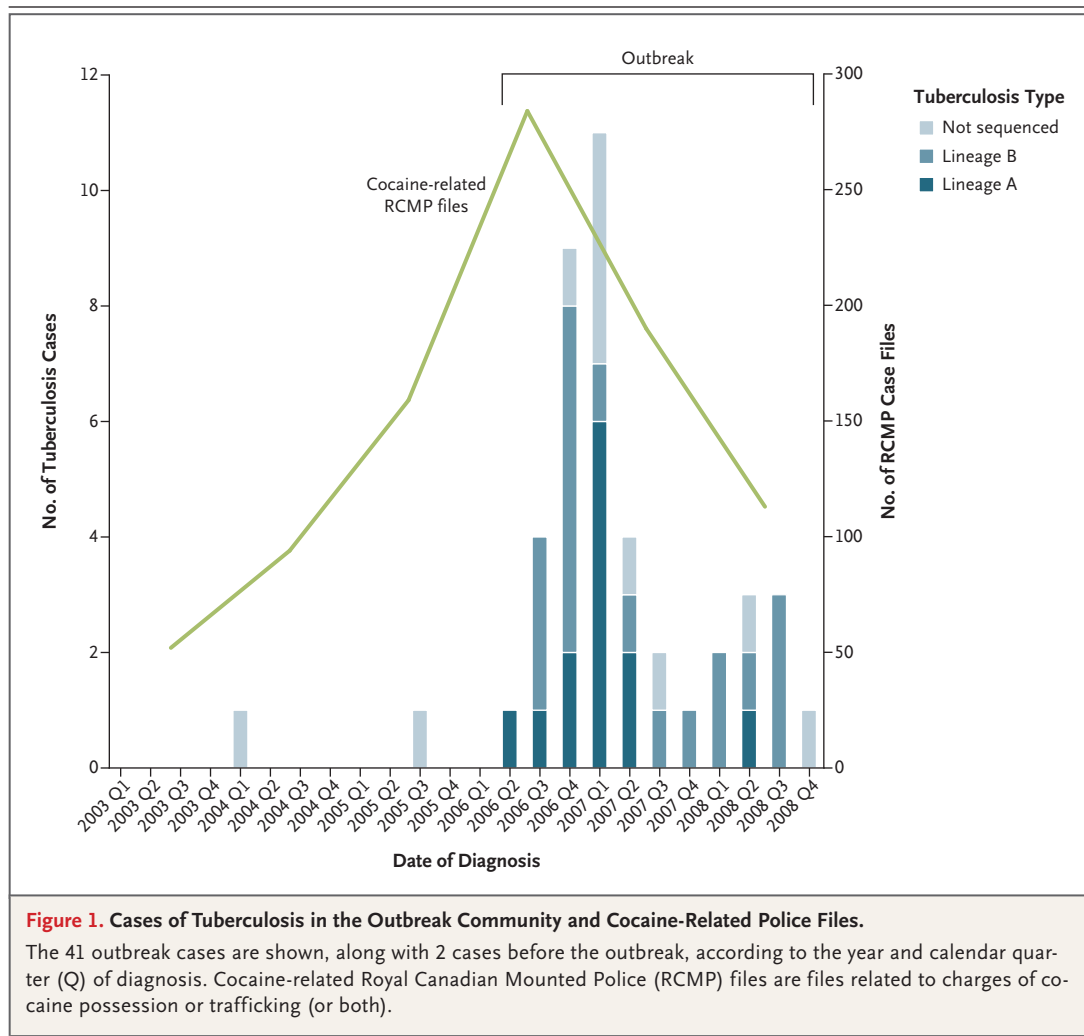
A social-network questionnaire (SNQ) (see the Supplementary Appendix, available with the full text of this article at NEJM.org) was developed to identify shared socialization settings and to prioritize case finding. The SNQ focused on drug and alcohol use, residential and travel history, places of social aggregation, and identification of contacts in the context of high-risk behaviors and locations, through the use of name-generator questions such as "Who else hangs out there?" The SNQ was administered by trained interviewers in the form of an open-ended interview and was used retrospectively to examine 9 of the 11 cases diagnosed before October 31, 2006 (1 of which occurred in an infant, so the parent was interviewed). With respect to the remaining 2 cases, the patients died before October 2006.

The field investigation concluded in November 2006. Case finding with the use of both traditional contact tracing and the SNQ was continued by the local health authorities. The resulting data were collated into a single network visualized in Cytoscape software[4] (see the Supplementary Appendix).

GENOMIC INVESTIGATION

M. tuberculosis isolates were cultured at the BCCDC Public Health Microbiology and Reference Laboratory. Restriction-fragment–length polymorphism (RFLP) analysis based on insertion sequence 6110 was performed on the first 13 outbreak isolates, and 24-loci MIRU-VNTR analysis was performed on all 37 laboratory-confirmed, culture-positive isolates at the National Reference Centre for Mycobacteriology of the National Microbiology Laboratory in Winnipeg, Manitoba.

A total of 36 *M. tuberculosis* isolates (32 of the 37 outbreak isolates and 4 historical isolates with identical MIRU-VNTR patterns) were sequenced (Genome Analyzer II sequencer, Illumina). The

Figure 1. Cases of Tuberculosis in the Outbreak Community and Cocaine-Related Police Files.

The 41 outbreak cases are shown, along with 2 cases before the outbreak, according to the year and calendar quarter (Q) of diagnosis. Cocaine-related Royal Canadian Mounted Police (RCMP) files are files related to charges of cocaine possession or trafficking (or both).

number of 50-bp reads ranged from 9.3 million to 21.6 million after filtering, and reads were deposited in the National Center for Biotechnology Information's Sequence Read Archive under accession number SRA020129. Reads were aligned to the reference genome (of *M. tuberculosis* strain CDC1551), with an average of 99.21% of the reference genome being covered by at least one read. Alignment identified 204 single-nucleotide polymorphisms (SNPs) unique to the 36 isolates. The SNPs were concatenated and used to construct a distance matrix, which was hierarchically clustered to generate a dendrogram. Phylogenetic analysis of the SNPs was also carried out, with the use of maximum-likelihood methods and Bayesian Markov chain Monte Carlo (MCMC) methods. A subgroup of SNPs, not in-

cluding those from repetitive genes, was also analyzed. Further details are provided in the Supplementary Appendix.

RESULTS

CLINICAL AND EPIDEMIOLOGIC CHARACTERISTICS OF THE PATIENTS

Between May 2006 and December 2008, a total of 41 cases of tuberculosis were diagnosed in the outbreak community (Fig. 1), elevating the annual incidence rate for the region by a factor of more than 10. The majority of the patients were adults (mean age, 36 years) presenting with pulmonary tuberculosis (68%) or pleural tuberculosis (24%) (Table 1). Two pediatric cases were diagnosed, including one in an infant presenting with dis-

seminated tuberculosis. None of the patients were born outside Canada, and none were positive for the human immunodeficiency virus.

Outcomes were recorded through December 31, 2009, with a minimum of 12 months of follow-up for all patients. The majority of patients had favorable outcomes, with 35 patients (85%) meeting the World Health Organization criteria for "cure or treatment completed."[5] One patient (2%) did not complete treatment, and 1 patient (2%) had a relapse of pulmonary tuberculosis. Four deaths (10% of patients) were recorded during the follow-up period, including one death from complications of disseminated tuberculosis and one death from hepatic failure during receipt of antituberculosis medication. Two patients died from unrelated causes (motor-vehicle accident and drug overdose).

SOCIAL-NETWORK ANALYSIS TO IDENTIFY A SOURCE PATIENT

Traditional contact tracing was used early in the outbreak for purposes of secondary case finding and source-case determination. However, a single source case could not be clearly identified, since the contact network was biased toward the pediatric case (MT0009) (Fig. 2A); this reflected the use of reverse contact tracing, in which the contacts of the child were screened in an effort to find the source of the child's infection.

A social-network approach was subsequently used. On the basis of a network derived from interviews of patients from early in the outbreak, one case (MT0001) was identified as the most likely source case (Fig. 2B). MT0001 was the fifth case identified in the outbreak; the patient was an adult with cavitary, smear-positive pulmonary tuberculosis that had been symptomatic and untreated for at least 8 months before the detection of the first case. MT0001 was connected to all but two early cases through direct contact or a shared social setting. Use of the SNQ improved subsequent case-finding efforts by revealing previously unreported social interactions and identifying several locations frequented by infectious patients, including two hotels, a meal center, two community centers, and a series of crack houses.

As in past studies involving SNQs,[6] use of the SNQ in this study identified demographic characteristics associated with an increased risk of tuberculosis transmission in low-incidence regions. Many SNQ respondents reported transient

Table 1. Characteristics of 41 Patients with Laboratory-Confirmed or Clinical Tuberculosis in a British Columbia Community from May 2006 through December 2008, According to Tuberculosis Lineage.*

Characteristic	Lineage A (N=13)	Lineage B (N=19)	Not Known (N=9)
Age — yr			
Median (range)	37 (21–55)	33 (1–71)	38 (4–52)
Mean (range)	36 (21–55)	34 (1–71)	32 (4–52)
Female sex — no. (%)	6 (46)	13 (68)	4 (44)
Canadian-born — no. (%)	13 (100)	19 (100)	9 (100)
Tuberculosis type — no. (%)			
Pulmonary	10 (77)	12 (63)	6 (67)
Pleural	3 (23)	4 (21)	3 (33)
Extrathoracic	0	1 (5)	0
Disseminated	0	2 (11)	0
Sputum smear positive for AFB — no. (%)	5 (38)	8 (42)	1 (11)
Culture-positive tuberculosis — no. (%)	13 (100)	19 (100)	5 (56)
Treatment outcome — no. (%)			
Death	1 (8)	3 (16)	0
Cure or treatment completed	11 (85)	16 (84)	8 (89)
Treatment not completed	0	0	1 (11)
Relapse	1 (8)	0	0
Risk factor — no. (%)			
Alcohol use	11 (85)	13 (68)	3 (33)
Crack cocaine use	7 (54)	12 (63)	6 (67)
Crack cocaine or alcohol use	11 (85)	16 (84)	7 (78)
HIV seropositivity	0	0	0

* AFB denotes acid-fast bacillus, and HIV human immunodeficiency virus.

living arrangements, crack cocaine use (61%), and alcohol use (66%) (Table 1). The social network was highly transitive (one of several measures of interconnectedness within a social network) (see the Supplementary Appendix), with each patient reporting, on average, social contact with six other patients.

WHOLE-GENOME SEQUENCING TO IDENTIFY GENETIC LINEAGE

RFLP and MIRU-VNTR typing of the outbreak isolates yielded identical patterns, suggesting a clonal outbreak (Fig. 3). Review of RFLP data from the same geographic area in the years preceding the outbreak indicated that the same clone had been detected sporadically in the region since 1995.
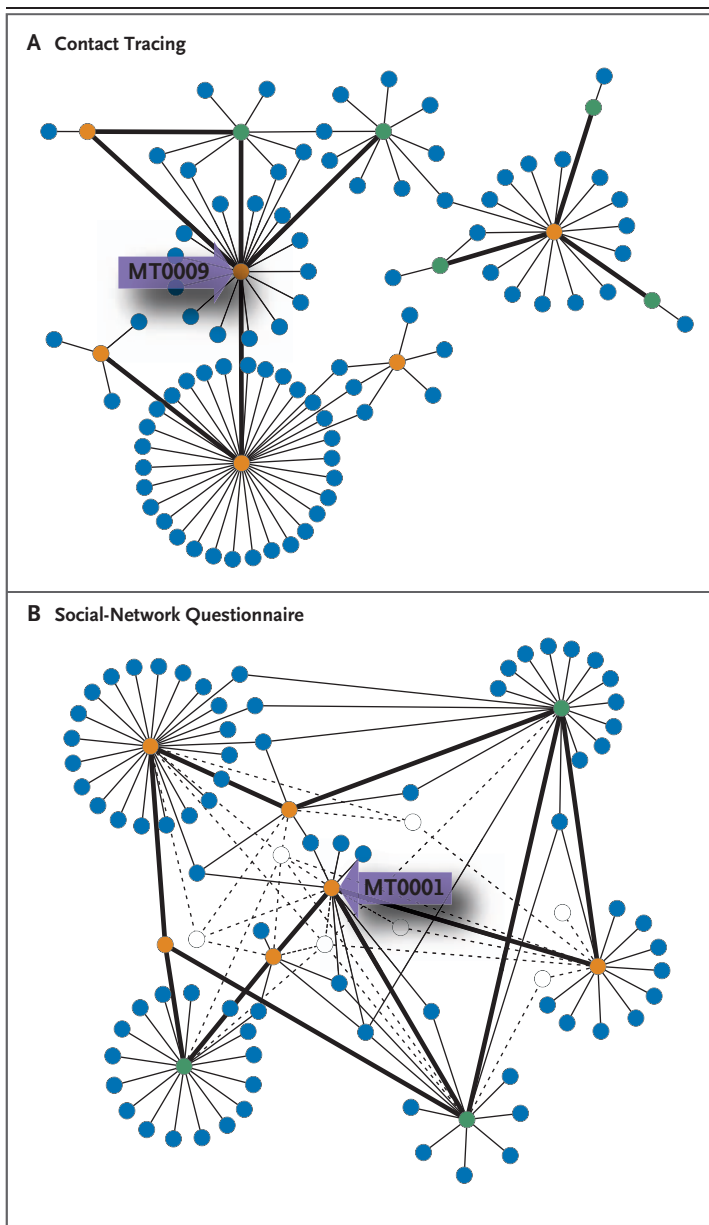
Figure 2. Community Networks of the First 11 Outbreak Cases, Derived from Traditional Contact Tracing and Social-Network Questionnaires.

Panel A shows the network based on contact tracing of the first 11 cases; Panel B shows the network based on the social-network questionnaires administered to 8 patients and the parent of 1 pediatric patient (the remaining 2 patients died before administration of the questionnaire). Orange circles indicate smear-positive outbreak cases, green circles smear-negative outbreak cases, blue circles named contacts, and white circles named places of social aggregation. Purple arrows indicate the most likely source case (identified on the arrow) suggested by the network structure. Heavy lines connect active cases to one another, light lines connect active cases to named contacts, and dashed lines connect active cases to places.

To examine transmission dynamics of the outbreak at a higher resolution, we analyzed the 204 SNPs that differentiated the 36 isolates from each other, most of which were nonsynonymous (44%, with lesser percentages in the other categories), in agreement with previous reports.[7,8] Unexpectedly, hierarchical clustering of the SNPs revealed two distinct lineages of *M. tuberculosis* cocirculating within the population (Fig. 3), suggesting not one outbreak but two simultaneous outbreaks. Thirteen outbreak isolates clustered to form lineage A, whereas 19 outbreak isolates
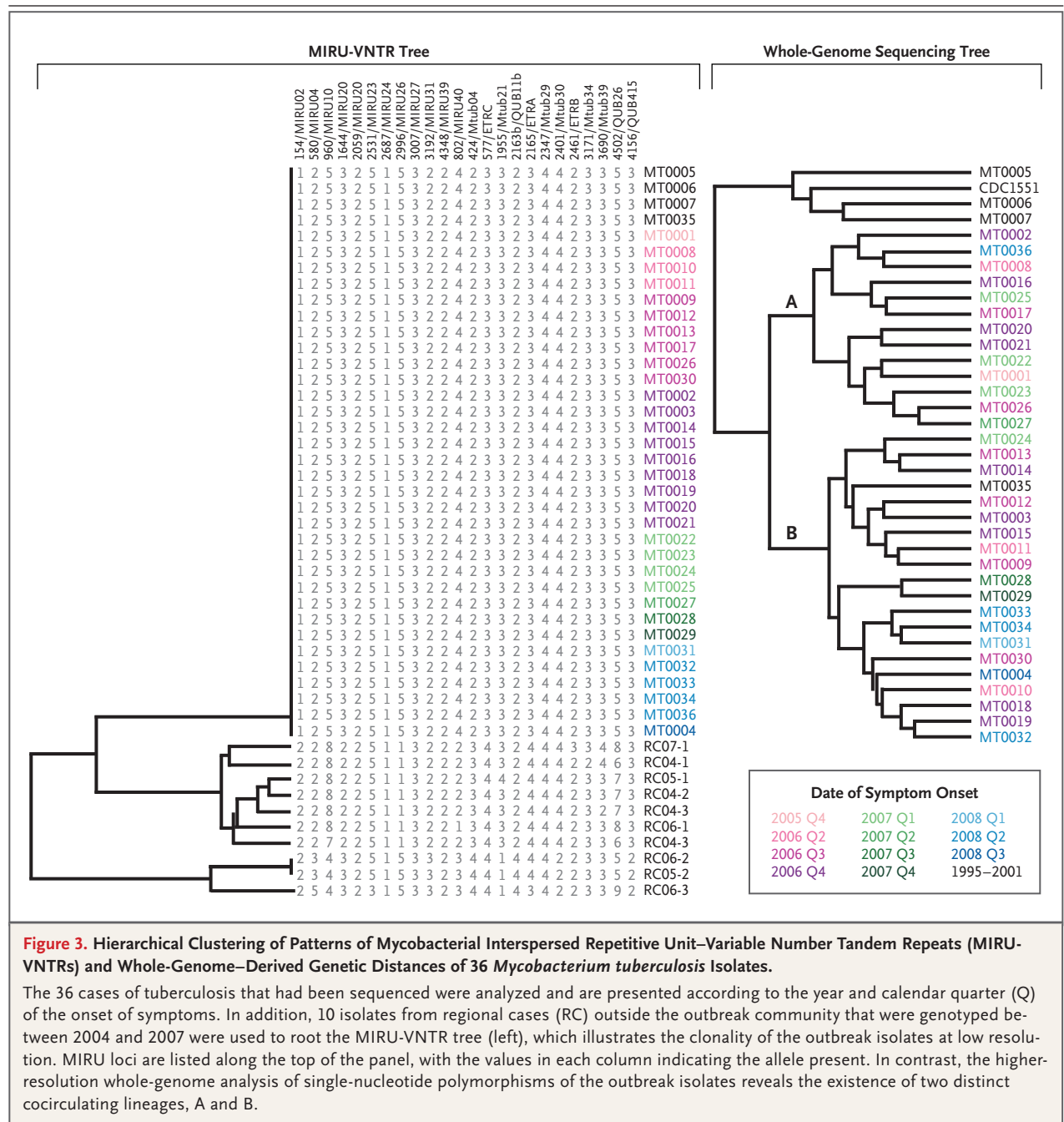
formed lineage B. The presence of two lineages was observed regardless of the similarity measure implemented during hierarchical clustering and was also observed when clustering was replaced with a Bayesian MCMC phylogenetic method. A maximum-likelihood method yielded a nearly identical tree, with 28 of 32 outbreak isolates belonging to the expected lineage. Four isolates (MT0031, MT0033, MT0034, and MT0036) whose lineage membership varied in the maximum-likelihood tree may reflect cases of coinfection with lineage A and B strains, resulting in a mixed signal that was difficult to resolve through maximum-likelihood approaches. Analysis of a data set filtered to remove SNPs in repetitive regions, as well as those appearing in a single isolate, recapitulated the dual-lineage observation.

No significant differences in demographic characteristics, clinical presentation, outcomes, or risk factors were detected between patients with lineage A tuberculosis and those with lineage B tuberculosis (Table 1). Examination of the complete community network revealed that most outbreak cases (80%) involved direct contact with both lineages, suggesting a single heterogeneous social network rather than two independent subnetworks associated with each lineage (see the Supplementary Appendix).

ORIGINS OF THE OUTBREAK

The historical isolates MT0005 (from 1995), MT0006 (from 2000), and MT0007 (from 2001) were most similar to lineage A (in maximum-likelihood and MCMC analyses), whereas the historical isolate MT0035 (from 2001) clustered with lineage

Figure 3. Hierarchical Clustering of Patterns of Mycobacterial Interspersed Repetitive Unit–Variable Number Tandem Repeats (MIRU-VNTRs) and Whole-Genome–Derived Genetic Distances of 36 *Mycobacterium tuberculosis* Isolates.

The 36 cases of tuberculosis that had been sequenced were analyzed and are presented according to the year and calendar quarter (Q) of the onset of symptoms. In addition, 10 isolates from regional cases (RC) outside the outbreak community that were genotyped between 2004 and 2007 were used to root the MIRU-VNTR tree (left), which illustrates the clonality of the outbreak isolates at low resolution. MIRU loci are listed along the top of the panel, with the values in each column indicating the allele present. In contrast, the higher-resolution whole-genome analysis of single-nucleotide polymorphisms of the outbreak isolates reveals the existence of two distinct cocirculating lineages, A and B.

B (in clustering, maximum-likelihood, and MCMC analyses), indicating that divergence of the two lineages preceded the current outbreak. The observation that both lineages had been present in the region for at least 5 years before the outbreak suggested that a social or environmental factor operating before 2006 triggered the outbreak, rather than a genetic change in the organism.

Epidemiologic follow-up revealed that of the 41 patients with tuberculosis during the outbreak, 25 (61%) had a self-reported history of crack cocaine use. Cocaine-related police files in the outbreak community mirrored the epidemic curve, peaking in 2006 before declining over the subsequent years (Fig. 1), and a number of crack houses proliferated in the region during the

same period (Murray K, Royal Canadian Mounted Police: personal communication), suggesting that the use of crack cocaine may have played a role in triggering and sustaining the outbreak.

CONSTRUCTION OF PUTATIVE TRANSMISSION NETWORKS

By applying three rules, social contacts linked to putative transmission events can be identified from a social network: first, transmission moves forward in time; second, a person with only a single previously infectious contact must have been infected by that contact; and third, in cases in which a person has multiple upstream infectious contacts, smear-positive contacts are more likely to transmit disease than smear-negative contacts.

In our study, on the basis of MIRU-VNTR genotyping alone, in which all isolates are assumed to be genetically equivalent, the majority of putative transmission events (13 events) were linked to MT0001 (Fig. 4A). Of these, 9 are biologically impossible if the genomic data are considered, since MT0001, a case of lineage A tuberculosis, could not have been the source of infection with lineage B.

To construct a more accurate transmission network, the social-network data were visualized in the context of the dendrogram generated from the genomic data (Fig. 4, and the Supplementary Appendix). Social relationships between patients with tuberculosis of the other lineage were removed (since lineage A cases could not have caused lineage B cases and vice versa), and the three rules of transmission were applied.

Social-network analysis had identified MT0001 as the most probable source case. This result was recapitulated in the genomics-based transmission network, with MT0001 acting as a "superspreader," probably causing six cases within lineage A. In lineage B, two cases, MT0010 and MT0011, were most likely to have been responsible for four and three subsequent cases, respectively. Clinically, the three source cases were characterized by smear-positive disease with pronounced delays (2 to 10 months) between the onset of symptoms and diagnosis.

Examining clades of closely related isolates revealed that within a clade, most patients did not have social relationships with other patients in the clade; rather, relationships consisted of a shared contact who was the source of transmission (e.g., Z to X and Z to Y), probably within a
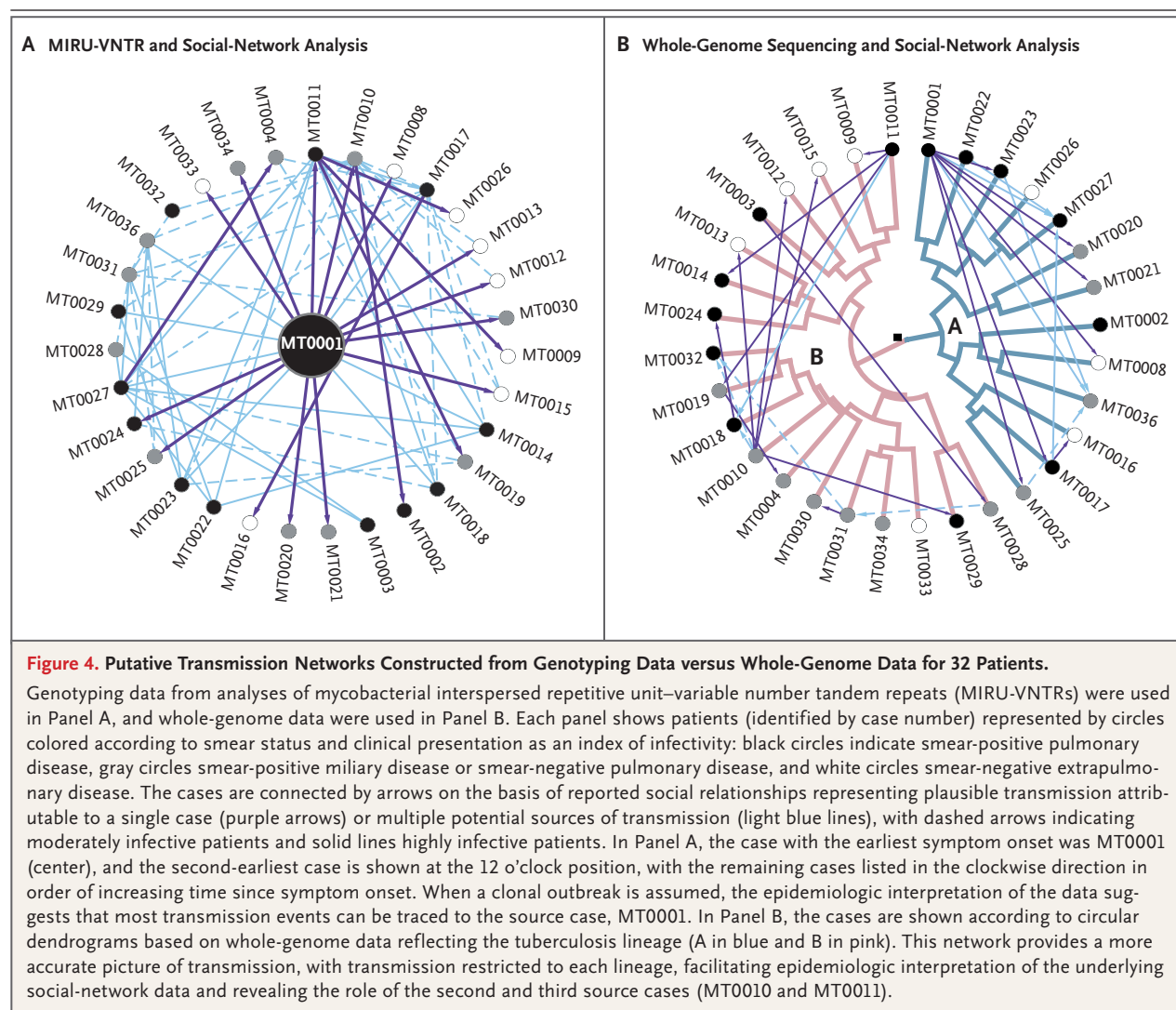
short time frame. Pairwise genetic distances between strains from source patients and strains from their exposed contacts may be affected by multiple variables, including length of time between infection and method of specimen collection. Strains present in patients with a high bacillary burden for several months before diagnosis and specimen collection have most likely undergone microevolution, manifested as greater genetic diversity in strains isolated from persons they infect after longer periods.

DISCUSSION

Between May 2006 and December 2008, a total of 41 cases of tuberculosis were diagnosed in a British Columbia community struggling with the challenges of alcoholism, drug use, and transient housing arrangements. Because of the recognized limitations of contact tracing,[9] field epidemiologists used social-network analysis early in the outbreak. Social-network analysis can improve case finding in vulnerable populations through structured interviews that identify high-risk behaviors, places of social aggregation, and persons not specifically named in traditional contact tracing.[10-12] In the outbreak we studied, social-network analysis outperformed contact tracing in identifying a probable source case as well as several locations and persons subsequently targeted for follow-up.

Throughout the field investigation, it was assumed that the outbreak was clonal because all isolates exhibited identical MIRU-VNTR patterns. Identical genetic fingerprints are frequently used to identify clusters of tuberculosis cases that may be epidemiologically linked[13] and as a surrogate for identity at the whole-genome level. Follow-up genome sequencing revealed this assumption to be false, showing that conventional molecular epidemiologic techniques cannot capture the full level of genetic diversity within a single genotype of *M. tuberculosis*, which may comprise multiple distinct lineages. The higher-resolution SNP patterns afforded by whole-genome sequencing revealed that the outbreak was the coalescence of two outbreaks, each with its own causative lineage of *M. tuberculosis*.

Association of historical isolates with each lineage suggested that divergence of the two lineages from a common ancestor occurred before 2001. Retrospective review of contact-tracing data around the four sequenced historical isolates did

Figure 4. Putative Transmission Networks Constructed from Genotyping Data versus Whole-Genome Data for 32 Patients.

Genotyping data from analyses of mycobacterial interspersed repetitive unit–variable number tandem repeats (MIRU-VNTRs) were used in Panel A, and whole-genome data were used in Panel B. Each panel shows patients (identified by case number) represented by circles colored according to smear status and clinical presentation as an index of infectivity: black circles indicate smear-positive pulmonary disease, gray circles smear-positive miliary disease or smear-negative pulmonary disease, and white circles smear-negative extrapulmonary disease. The cases are connected by arrows on the basis of reported social relationships representing plausible transmission attributable to a single case (purple arrows) or multiple potential sources of transmission (light blue lines), with dashed arrows indicating moderately infective patients and solid lines highly infective patients. In Panel A, the case with the earliest symptom onset was MT0001 (center), and the second-earliest case is shown at the 12 o'clock position, with the remaining cases listed in the clockwise direction in order of increasing time since symptom onset. When a clonal outbreak is assumed, the epidemiologic interpretation of the data suggests that most transmission events can be traced to the source case, MT0001. In Panel B, the cases are shown according to circular dendrograms based on whole-genome data reflecting the tuberculosis lineage (A in blue and B in pink). This network provides a more accurate picture of transmission, with transmission restricted to each lineage, facilitating epidemiologic interpretation of the underlying social-network data and revealing the role of the second and third source cases (MT0010 and MT0011).

not yield direct contacts between patients with historical cases and those with outbreak cases of the same lineage, but all these patients were known to frequent common locations in the community. Had social-network analysis and location-based screening been performed for the historical cases rather than contact tracing alone, the present outbreak might have been prevented.

The highly transitive nature of the outbreak community's social network made transmission patterns difficult to identify with the use of epidemiologic techniques alone. Our findings show that genome sequencing allows the larger social network to be divided into subnetworks associated with specific genetic lineages of disease. Removal of social relationships that could not have led to transmission according to the genomic data greatly

reduced the complexity of the network, increased the usefulness of the existing epidemiologic data, and allowed for the identification of individual transmission events and three coprimary source patients. The source patients all had smear-positive tuberculosis, were highly connected within the social network, and had been symptomatic for several months before treatment. In the social network we studied, the contribution of these "superspreader" cases to the overall prevalence of disease played a larger role than secondary transmission through person-to-person contact.

The simultaneous reappearance of two extant lineages suggests that a social or environmental factor, not a genetic change in the organism, most likely triggered the outbreak. Our observations suggest that a rise in crack cocaine use

within the community, peaking in 2006, may have been this trigger. Previous work has established that social networks can mediate behaviors related to addictive substances.[14,15] However, a lack of longitudinal network data precludes delineation of the spread of crack cocaine in the network described here. Surveys of crack cocaine users describe findings that are consistent with high rates of tuberculosis exposure,[16,17] and several tuberculosis outbreaks have been reported in crack-using communities.[11,12] A constellation of impaired host immunity, lack of access to health care, and crowded, poorly ventilated environments most likely predisposes users to both exposure to and development of tuberculosis.[12,18,19] Poverty and the highly transitive social networks established by crack users may also contribute to tuberculosis transmission.

In summary, we describe the integration of social-network analysis with high-resolution bacterial genome sequencing to enhance the investigation of a tuberculosis outbreak. Our data show that epidemiologic inferences are greatly improved by interpretation in the context of whole-genome sequencing and that a single *M. tuberculosis* genotype contains enough genetic diversity to refine epidemiologic linkages on the basis of genetic lineage or clade membership.

Although whole-genome sequencing may not currently be feasible as a routine tool for outbreak investigation, data generated by this and other sequencing projects will identify the most informative panel of markers of genomic variability in *M. tuberculosis*, facilitating the development of higher-resolution genotyping methods. As the cost of whole-genome sequencing continues to decrease and next-generation sequencing platforms become integrated into public health practice, combined microbial genomic and epidemiologic approaches like those described here will become an important and tractable first step toward a systems approach to tuberculosis control.[20]

REFERENCES

1. Tuberculosis in Canada 2007 (catalog no. HP37-5/2007E-PDF). Ottawa: Minister of Public Works and Government Services Canada, 2009.

2. Canadian tuberculosis standards, 6th ed. (catalog no. HP40-18/2007E-PDF). Ottawa: Tuberculosis Prevention and Control, Public Health Agency of Canada, and the Canadian Lung Association/Canadian Thoracic Society, 2007.

3. An expanded DOTS framework for effective tuberculosis control. Geneva: World Health Organization, 2002. (WHO/CDS/TB/2002.297.)

4. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13: 2498-504.

5. Treatment of tuberculosis: guidelines for national programmes, 4th ed. Geneva: World Health Organization, 2009. (WHO/HTM/TB/2009.420.)

6. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. Int J Tuberc Lung Dis 2008;12:480-92.

7. Gutacker MM, Mathema B, Soini H, et al. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis 2006;193:121-8.

8. Dos Vultos T, Mestre O, Rauzier J, et al. Evolution and diversity of clonal bacteria: the paradigm of Mycobacterium tuberculosis. PLoS One 2008;3(2):e1538.

9. Asghar RJ, Patlan DE, Miner MC, et al. Limited utility of name-based tuberculosis contact investigations among persons using illicit drugs: results of an outbreak investigation. J Urban Health 2009; 86:776-80.

10. Cook VJ, Sun SJ, Tapia J, et al. Transmission network analysis in tuberculosis contact investigations. J Infect Dis 2007; 196:1517-27.

11. McElroy PD, Rothenberg RB, Varghese R, et al. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. Int J Tuberc Lung Dis 2003;7:Suppl: S486-S493.

12. Malakmadze N, González IM, Oemig T, et al. Unsuspected recent transmission of tuberculosis among high-risk groups:

implications of universal tuberculosis genotyping in its detection. Clin Infect Dis 2005;40:366-73.

13. Nguyen LN, Gilbert GL, Marks GB. Molecular epidemiology of tuberculosis and recent developments in understanding the epidemiology of tuberculosis. Respirology 2004;9:313-9.

14. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behaviour in a large social network. Ann Intern Med 2010;152:426-33.

15. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008;358:2249-58.

16. Howard AA, Klein RS, Schoenbaum EE, Gourevitch MN. Crack cocaine use and other risk factors for tuberculin positivity in drug users. Clin Infect Dis 2002;35:1183-90.

17. Grimes CZ, Hwang LY, Williams ML, Austin CM, Graviss EA. Tuberculosis infection in drug users: interferon-gamma release assay performance. Int J Tuberc Lung Dis 2007;11:1183-9.

18. Story A, Bothamley G, Hayward A. Crack cocaine and infectious tuberculosis. Emerg Infect Dis 2008;14:1466-9.

19. Crack cocaine use among persons with tuberculosis — Contra Costa County, California, 1987–1990. MMWR Morb Mortal Wkly Rep 1991;40:485-9.

20. Fenner L, Egger M, Gagneux S. Annie Darwin's death, the evolution of tuberculosis and the need for systems epidemiology. Int J Epidemiol 2009;38:1425-8.

Copyright © 2011 Massachusetts Medical Society.

Great Blue Heron   Peter Pak, M.D.