Expanded Geographical Distribution of the N Family of *Mycobacterium tuberculosis* Strains within the United States


Seattle Biomedical Research Institute and Seattle-King County Public Health Department, Seattle, Washington; Epidemic Intelligence Service, Epidemiology Program Office, Centers for Disease Control and Prevention, Atlanta, Georgia; and The Public Health Research Institute at the International Center for Public Health, Newark, New Jersey

Received 15 September 2003/Returned for modification 10 November 2003/Accepted 19 November 2003

The N and W-Beijing families of *Mycobacterium tuberculosis* are phylogenetically closely related. The ability of the W-Beijing family to rapidly cause widespread disease is well described; however, few outbreaks involving the N family have been reported outside the New York City, N.Y., area. During 2002 to 2003, Seattle, Wash., experienced a rapidly expanding tuberculosis outbreak involving 38 persons in a 23-month period. The outbreak strain, SBRI9, exhibited the genotypic properties of the N family. Its IS6110 insertion fragment length polymorphism pattern was identical or nearly identical to those of two N family strains that were responsible for clusters of tuberculosis cases, including a large nosocomial outbreak, in New York City and New Jersey from 1989 to 1990. It was also identical to strains involved in late 1990s tuberculosis cases in Michigan, Maryland, and Arkansas. Further monitoring of the N family may show that it shares with the W-Beijing family the propensity to spread rapidly, suggesting that this characteristic evolved prior to the divergence of the two genetic lineages.

Genotypic analysis of *Mycobacterium tuberculosis* isolates has been used to confirm tuberculosis (TB) outbreaks (17), to investigate TB transmission in communities (3, 24), to identify cases of exogenous reinfection (23), to confirm cross-contamination in the laboratory (10), and to study clonal expansion of strains (14). Clonal expansion has been studied most extensively in the W-Beijing family of *M. tuberculosis* strains (25). A review of the epidemiology and phylogenetic reconstruction of the W-Beijing family was published recently (2).

The widespread geographic distribution and frequent TB outbreaks associated with the W-Beijing family suggest that these strains are successful human pathogens that spread more aggressively than other strains (2, 16, 20, 25). The factors responsible for this success are not yet known. Studies have documented clonal expansion of multidrug-resistant (MDR) as well as pan-susceptible W family strains across the United States (4, 14). Characterization of the development of MDR during dissemination of one clone, strain W14 (14), improved our understanding of how MDR *M. tuberculosis* develops and how it can be controlled. By understanding how the W-Beijing family evolved, we may better understand the specific characteristics that make some strains of *M. tuberculosis* spread more rapidly than others.

W and Beijing strains share certain genotypic characteristics, and they are viewed as members of the same phylogenetic lineage. These characteristics are (i) an IS6110 restriction fragment length polymorphism (RFLP) pattern similar to that of strain W, with 15 to 26 IS6110 copies; (ii) a spoligotype pattern, S00034, characterized by the absence of spacers 1 to 34 and positive hybridization to spacers 35 to 43 in the direct repeat region; (iii) the occurrence of two mutations, one in *katG* codon 463 (Leu) and one in gyrA codon 95 (Thr), that define principal genetic group 1; (iv) an IS6110 insertion in the origin of replication between the *dnaA* and *dnaN* genes; and (v) one or two IS6110 insertions in the NTF chromosomal region (Beijing strains have one insertion, whereas W MDR strains from New York City, N.Y., have two insertions in a head-to-tail orientation) (3, 4, 14, 18, 21).

In contrast to the case for the W-Beijing family, the public health importance of closely related groups has not been extensively documented. Doing so would help us to understand the characteristics that make the strains of the W-Beijing family spread so aggressively and how they evolved. One ancestral branch related to the W-Beijing family, the N family, was described by Kurepina et al. (14). N family strains lack IS6110 insertions in the NTF chromosomal region and have IS6110-based RFLP patterns distinct from those of W-Beijing strains, but they share other genotypic characteristics with W-Beijing strains.

In 2002, Public Health Seattle and King County, Wash., reported a 100% increase in TB among its homeless population, suggesting that an outbreak was occurring. Strain typing of *M. tuberculosis* isolates from homeless TB patients identified an outbreak cluster, which was determined to be caused by an N family strain that is closely related or identical to strains responsible for TB cases in New York City, New Jersey, and Michigan. This is the first report of N family clusters outside the New York City area.

**MATERIALS AND METHODS**

Strain typing of isolates from the Seattle outbreak. *M. tuberculosis* isolates from patients suspected to be involved in the Seattle, Wash., outbreak were sent
to the Seattle Biomedical Research Institute for genotyping from 1 August 2002 through 1 January 2003. Standard spoligotyping methods (12) were used to rapidly assess whether isolates were likely related to the outbreak strain. Subsequently, standard IS6110-based RFLP methods (22) were performed on all isolates, with priority given to isolates with matching spoligotype patterns. Further genetic analysis was performed on 16 isolates with identical spoligotype and RFLP patterns. Matching spoligotypes were assigned designations according to standard nomenclature (9). RFLP fingerprint images were captured by using an Expression 1680 instrument (Epson) and analyzed by using BioNumerics software (Applied Maths, Sint-Martens-Latem, Belgium) on a Dell workstation.

Genotypic analysis of the Seattle outbreak strain. Strains were analyzed by PCR for the presence of IS6110 in the dnaN-dnaA intergenic region of the origin of replication. Primers were originally designed for use as a PCR-generated probe for region A in dnaN-dnaA (region A [5'-GCCACCGTGTCAGCGCTC CAA] and region Ar [5'-GGGCTTCGGTGTACCTCAT]) were used as previously described (23). Amplification was performed in a 100-μl reaction mixture containing 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.2 mM deoxynucleoside triphosphates (dNTPs) (200 μM each), 1 μM primer (each), 50 ng of DNA template, and 1 U of Taq polymerase (Sigma Chemical Co., St. Louis, Mo.) were prepared for a final reaction volume of 50 μl. PCR cycling times were (i) 4 min at 94°C, (ii) 30 cycles of 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C, and (iii) 7 min at 72°C. Amplification products were analyzed on a 1% agarose gel. A 537-bp product was expected in amplicons that did not have an IS6110 insertion in the dnaN-dnaA region, compared to a ~2,000-bp product expected in those that carried the insertion. Amplification products were confirmed by automated sequencing. Products were purified with the QIAQuick Spin PCR purification kit (Qiagen, Valencia, Calif.) and sequenced with an ABI PRISM 377 DNA sequencer (Applied Biosystems, Foster City, Calif.). Sequences were analyzed by BLAST comparison to the National Center for Biotechnology Information nonredundant sequence database (www.ncbi.nlm.nih.gov/BLAST/).

The principal genetic group of the outbreak strains was determined by PCR amplification and DNA sequencing of target-specific regions within the genes katG and gyrA. Primers katGf (5'-TGCAAGTCCGCCGCTTAGC)-katG (5'-ACGCAACCGGACACGAG) and gyrAf (5'-CATCTACGATCGATCGA)- gyrA (5'-GGGCTTCGGTGTACCTCAT) were used as previously described (13). The PCR mixture contained 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.2 mM dNTPs (200 μM each), 50 ng of DNA template, and 1 U of Taq polymerase (Sigma), for a final volume of 50 μl. PCR cycling times for katG were as follows: (i) 4 min at 94°C, (ii) 35 cycles of 1 min at 94°C, 1 min at 60°C, and 2 min at 72°C, and (iii) 15 min at 72°C. Those for gyrA were as follows: (i) 4 min at 94°C; (ii) 30 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C; and (iii) 15 min at 72°C. Products were analyzed on a 1% agarose gel, and the appropriate product sizes were detected (katG, 1,429 bp; gyrA, 320 bp). The products were then purified and sequenced as described above.

A multiplex PCR approach was used to determine whether there was an IS6110 in the NTF-1 region of the outbreak strains of M. tuberculosis. PCR primers and parameters were previously described by Pilkaitys et al. (18). Briefly, a 50-μl reaction mixture containing 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl2, 0.2 mM dNTPs (200 μM each), 1.0 μM primer (each), 50 ng of DNA template, and 1 U of Taq polymerase was amplified with the following parameters: (i) 4 min at 94°C; (ii) 30 cycles of 1 min at 94°C, 1.75 min at 60°C, and 2 min at 72°C; and (iii) 7 min at 72°C. Products were analyzed on a 1.5% agarose gel. The expected band sizes were 523 and 223 bp for Beijing family strains and 523 bp for the positive control (H37Rv).

Seattle outbreak investigation. The medical charts of all homeless persons diagnosed with TB in 2002 were reviewed to determine each patient's site of disease and results of the sputum examination and human immunodeficiency virus (HIV) test. Contact investigations were reviewed to determine epidemiological associations among patients and results of tuberculin skin tests in contacts. Drug susceptibility testing was performed on all M. tuberculosis isolates.

Identification of related strains involved in TB clusters in other states. Spoligotyping and RFLP images were independently compared to those in the Centers for Disease Control and Prevention (CDC) 1996 to 2000 National TB Genotyping and Surveillance Network (NTBGSN) database (7) and the Public Health Research Institute (PHRI) TB Center database in Newark, N.J.

RESULTS

Genotypic characteristics of the Seattle outbreak strain. Spoligotyping and RFLP analysis were performed on 123 isolates from King County. Spoligotype analysis of 16 isolates from epidemiologically linked homeless persons (K. H. Lofy, M. McConnell, S. Goldberg, N. Mills, T. Kuss, G. Cangelsi, J. Milan, L. Diem, J. Pang, K. Hawthorne, and P. McElroy, presented at the International Meeting of the American Thoracic Society, Seattle, Wash., 2003) revealed positive hybridization to spacers 35 to 43 of the direct repeat locus (spoligopattern S00034). IS6110 RFLP analysis of these 16 isolates showed a matching 15-band pattern. The number of bands is at the low end of range of the IS6110 copy numbers expected for W-Beijing family isolates (Table 1). The Seattle outbreak strain RFLP pattern was designated SBR19.

As shown by PCR analysis, strain SBR19 has an IS6110 insertion in the origin of replication, between the dnaA and dnaN genes (Fig. 1A). Sequence analysis showed that it has the mutations in the katG and gyrA genes (Kat463Leu and Gyr95Thr, respectively) that categorize it as a member of principal genetic group 1. These characteristics are consistent with the W-Beijing family (2). However, strain SBR19 does not have an IS6110 insertion in the NTF region (Fig. 1B). Taken together, these characteristics are consistent with the N family (Table 1).

Characteristics of the Seattle outbreak. Seventeen (57%) of 30 TB cases among homeless people diagnosed in 2002 were linked by molecular strain typing to strain SBR19. As of 14 November 2003, 21 (78%) of 27 TB cases among homeless people diagnosed in 2003 matched the outbreak strain, bringing the total number of cases involving this strain to 38 in a 23-month period. Further molecular and epidemiological analysis is in progress. Isolates from all outbreak cases have been susceptible to isoniazid, rifampin, ethambutol, pyrazinamide, and streptomycin. Seven (41%) of the patients diagnosed in 2002 were HIV infected.

Of 242 high-risk outbreak contacts who completed tuberculin skin testing and were entered in the database as of 15 April 2003, 95 (39%) had positive skin tests with ≥5 mm of induration (Lofy et al., presented at the International Meeting of the American Thoracic Society). In contrast, during the first 10 months of 2002, before program resources were diverted from routine screening to intensive screening activities, routine screening resulted in only 52 positive results (7%) (≥10 mm) among 798 homeless persons in Seattle completing tuberculin testing. One outbreak case (HIV negative) had previously been treated successfully for active disease with a different M. tuberculosis strain that exhibited a 3-band IS6110 RFLP, indicating that this patient was exogenously reinfected with the 15-band SBR19 strain (data not shown).

Related strains involved in TB clusters in other states. Comparison of the 15-band SBR19 RFLP pattern to those in the
bands are nonspecific products, which were seen previously in some strains analyzed under these conditions (18).

New Jersey (H11005) respectively. Beijing strains are expected to have one IS6110 insertion in the NTF region, resulting in a 223-bp PCR product. W strains, which were outbreak isolates, indicating an IS6110 insertion in the origin of replication. (B) Multiplex PCR for IS6110 insertions in the NTF region (18). Lanes 1 and 2, Beijing family control strains. Lanes 3 to 5, Seattle outbreak isolates. Lanes 6 and 7, non-W-Beijing control strains KY and H37Ra, respectively. Beijing strains are expected to have one IS6110 insertion in the NTF region, resulting in a 223-bp PCR product. W strains, which were not included in this experiment, have a second insertion resulting in a 175-bp product. The 500-bp product is an internal PCR control, and larger bands are nonspecific products, which were seen previously in some strains analyzed under these conditions (18).

Comparison of the SBR19 RFLP pattern to those in the NTBGSN database found 10 RFLP pattern matches with isolates obtained from TB patients diagnosed during 1996 to 2000 in Michigan (n = 7), Arkansas (n = 1), Maryland (n = 1), and New Jersey (n = 1); the national fingerprint number assigned by the NTBGSN to this fingerprint pattern is 00787. All 10 cases included pulmonary disease. Six (diagnosed in 1996 to 1997) of the seven Michigan cases, including two in infants, were in a single zip code. Of the seven cases, it was apparent that in two cases adults transmitted the disease to one or two of their younger relatives, accounting for five of the seven cases. The additional two cases did not have enough epidemiological information to link them to any of the other cases.

**DISCUSSION**

This is the first report of a large TB outbreak due to the M. tuberculosis N family outside the New York City area. Initial characterization of the N family showed that it is very closely related to the W-Beijing strain family (Table 1). Kurepina et al. (14) originally identified 10 N family strains, 3 of which, strains N, N2, and N3, were recovered from a total of 197 patients as of 1998. Strains N and N3 were both pan-susceptible; however, strain N2 was an MDR strain responsible for a nosocomial outbreak in 1989 to 1990 (8). Since 1992, 77 N2 strains have been entered into the PHRI TB Center database. Strain N4, which shares an identical IS6110 RFLP pattern with strain SBR19 and is pan-susceptible like the Seattle strain, has been isolated from two or three cases per year in New York City and New Jersey from 1992 to 2002 (n = 24). Our findings show that this strain family has spread to other states as well (Table 2).

Due to the significant public health threat posed by the MDR W strain identified in New York City in the early 1990s (4), the clonal expansion and molecular nature of this group have been extensively characterized (3, 14, 18). The global distribution of the Beijing family has also been extensively reported (2). However, most reports identified Beijing family strains based on subsets of the characteristics in Table 1, along with other characteristics relevant to their studies. For example, the W-Beijing genotype is commonly defined on the basis of its spoligotype pattern, the high-copy-number IS6110 pattern, and in some cases the presence of the IS6110 insertion(s) in the dnaA-dnaN region (5, 16, 24, 25). These genetic markers are also common to members of the N family. It is the presence or absence of the IS6110 insertions(s) in the NTF chromosomal region, combined with distinct IS6110 RFLP patterns, that distinguishes the W-Beijing and N family strain lineages.

Given that most studies do not distinguish this branch from the W-Beijing strains, it is possible that members of the N family have caused case clusters that were attributed to the W-Beijing family. In an attempt to refine the molecular epidemiology of W-Beijing strains and related lineages, a DNA “polyprobe” which can identify and distinguish members of the W-Beijing...
family and related strain families, including the N family, has been developed (N. Kurepina, E. Shashkina, B. Mathema, E. Likhoshvay, J. Bifani, K. Kremer, D. van Soolingen, and B. Kreiswirth, Abstr. 3rd Meet. Concerted Action Project, Prague, Czech Republic, p. 40, 2003).

Prior to the present report, it was not known whether there are epidemiological links between N family cases in other states and the Seattle outbreak. None of the Michigan patients with isolates matching the Seattle outbreak strain were homeless.

A region in the United States is defined by the CDC as having a low incidence of TB if it has ≤3.5 cases per 100,000 (11). The Seattle metropolitan area is a moderate-incidence region, with case rates of 6 per 100,000 throughout the 1990s (6). One case associated with the Seattle 2002 outbreak was an exogenous reinfection of an HIV-negative individual, something that is rarely reported for moderate-incidence settings (19, 23). Bandera et al. (1) reported exogenous reinfection in a region of Italy described as low incidence. However, that region had a case rate of 17.5 per 100,000, five times higher than the CDC definition of low incidence in the United States. Sixteen percent of case patients in that study with recurrent TB had exogenous reinfection, and 60% of the reinfection strains (n = 3 of 5) were also responsible for case clusters within the community. It is difficult to discern the extent to which host factors, rather than pathogen characteristics, contribute to exogenous reinfection and aggressive expansion of TB outbreaks. However, these observations are consistent with highly virulent strains of M. tuberculosis. Recent data have shown that a W-Beijing strain, HN878, was more virulent in mice and rabbits than strains CDC1551 and H37Rv (15; G. Kaplan, presented at the Keystone Symposium Tuberculosis: Integrating Host and Pathogen Biology, Taos, N.Mex., 26 January 2003). Analysis of the N family may show that it shares similar characteristics, which would indicate that these characteristics developed prior to the divergence of the N and W-Beijing family lineages. Such observations, combined with comparative genomic approaches, may help to identify the genetic basis for the rapid geographic spread of certain M. tuberculosis strains.

ACKNOWLEDGMENTS

This work was supported by the Seattle Biomedical Research Institute, the Seattle-King County Department of Public Health, the Washington State Department of Health, the Centers for Disease Control and Prevention, and grant AI25767 from the National Institutes of Health (to G.A.C.).

BioNumerics software was a kind gift of the Firland Foundation. We thank Kaeryn Lewis, Linda Haba, Jeanette Frazier, Carolyn Wallis, Christy Huynh, and the Washington State Department of Health, especially Trang Kuss, Kim Field, and Joseph Aharchi.

REFERENCES


