Molecular Identification of Streptomycin Monoresistant *Mycobacterium tuberculosis* Related to Multidrug-Resistant W Strain

Pablo Bifani,*† Barun Mathema,* Martha Campo,* Soraya Moghazeh,* Beth Nivin,† Elena Shashkina,* Jeffrey Driscoll,§ Sonal S. Munsiff,‡ Richard Frothingham,¶ and Barry N. Kreiswirth*

*Public Health Research Institute Tuberculosis Center, New York, NY, USA; †New York University School of Medicine, New York, New York, USA; ‡New York City Department of Health, New York, New York, USA; §New York State Department of Health, Albany, New York, USA; ¶Durham Veterans Affairs Medical Center, Durham, North Carolina

A distinct branch of the *Mycobacterium tuberculosis* W phylogenetic lineage (W14 group) has been identified and characterized by various genotyping techniques. The W14 group comprises three strain variants: W14, W23, and W26, which accounted for 26 clinical isolates from the New York City metropolitan area. The W14 group shares a unique IS6110 hybridizing banding motif as well as distinct polymorphic GC-rich repetitive sequence and variable number tandem repeat patterns. All W14 group members have high levels of streptomycin resistance. When the streptomycin resistance rpsL target gene was sequenced, all members of this strain family had an identical mutation in codon 43. Patients infected with the W14 group were primarily of non-Hispanic black origin (77%); all were US-born. Including HIV positivity, 84% of the patients had at least one known risk factor for tuberculosis.

With the advent of molecular techniques, tuberculosis (TB) investigators have a powerful tool to further the understanding of the transmission and phylogenetic properties of *Mycobacterium tuberculosis*. Molecular techniques have been used to discriminate exogenous versus endogenous disease (1-3), investigate suspected outbreaks (2,4-6) and cases of laboratory cross-contamination (7-9), study transmission within a defined geographic setting (10,11), and demonstrate the occurrence of exogenous superinfection in immunocompetent and immunocompromised patients (12,13).

Genotyping has facilitated identification and characterization of the W strain, a multidrug-resistant (MDR) clone associated primarily with nosocomial transmission in hospitals and detention facilities in New York City (NYC) in the early 1990s (14-17). In such studies, molecular markers were used to confirm and characterize the W strain outbreak and to elucidate a plausible evolutionary scenario for the sequential acquisition of multidrug resistance (14,17,18). In a recent study, when molecular techniques were applied in a population-based setting, genotyping identified a drug-susceptible group of isolates (W4) that represents a distinct branch of the W phylogenetic lineage. Members of this W4 group define a previously unidentified cluster of cases in a community in northern New Jersey; the cluster likely resulted from both historical and recent transmission (19).

Although the W-MDR strain from NYC is clearly distinct from the drug-susceptible W4 group, both groups of isolates evolved from a common ancestor (20).

Strains that define the W family have several distinguishing genotypes in common: 1) they belong to principal genetic group 1 (17,21); 2) they have similar spoligotype patterns, characterized by a deletion of spacers 1-34 and the corresponding repeat in the direct repeats (DR) region (19, 22-24), an alteration that defines spoligopattern S00034 and closely related spoligotypes; 3) they contain a unique insertion in the origin of replication (19,20) and in the NTF locus (20,25). Strains grouped in the W or Beijing family are prevalent in China, Southeast Asia (26,27), Russia, and other former Soviet regions (unpub. data) and have recently been reported in South Africa (23).

By multiple genetic techniques, we investigated a cluster of NYC *M. tuberculosis* isolates (W14 group) that are resistant to at least streptomycin and share identical or closely related DNA fingerprint patterns. Results from multiple typing methods were used to show the relatedness of the three IS6110 fingerprint patterns, W14, W23, and W26 and their evolution.

Materials and Methods

Mycobacterial Isolates and Patients

A total of 26 isolates from 26 patients were identified and grouped based on their IS6110 fingerprint patterns. The IS6110 fingerprint archive at the Public Health Research Institute Tuberculosis Center (PHRI TB Center) includes fingerprint patterns from >13,000 *M. tuberculosis* clinical

---

*Address for correspondence: Barry N. Kreiswirth, Public Health Research Institute Tuberculosis Center, New York, New York 10016; USA; fax: 212-578-0853; e-mail: barry@phri.org*
specimens characterized since 1992. More than 80% of the isolates were cultured from NYC and New Jersey cases; remaining samples are from seven other states and international sources.

The W14 group isolates are part of a NYC convenience sample, which represents 44% (n=6,655) of the total number of culture-positive TB cases reported in 1992-1999 (~15,000). Isolates were collected for numerous outbreak, surveillance, and research studies (14,17,28-31). Basic clinical, demographic, and routine contact-tracing information was obtained for 22 of the 26 patients; data for 21 cases were obtained from the NYC Tuberculosis Control Program surveillance database; the one exception was from the New Jersey Department of Health and Senior Services. Isolates originated from 16 institutions, including NYC hospitals and correctional facilities. Laboratory cross-contamination was ruled out since none of the clinical specimens were cultured or processed during the same time period, and all patients had well-documented TB.

**IS6110 DNA Fingerprinting and Pattern Interpretation**

M. tuberculosis isolates were cultured on Lowenstein-Jensen slants and grown at 37°C for 3 to 5 weeks. IS6110 DNA fingerprint analyses were performed according to a standard method using both the 5' and 3' fragments of the IS6110 genetic element (32). Hybridization patterns were compared on a Sun Sparc5 Workstation by using the Biolmage Whole Band Analyzer software version 3.4 (Genomic Solutions, Ann Arbor, MI). The Jaccard matching method and an unweighted pair group method that used arithmetic averages and average linkage clustering identified related patterns, in accordance with the protocol of the National Tuberculosis Genotyping and Surveillance Network, Centers for Disease Control and Prevention (CDC). Nomenclature of the DNA fingerprint patterns was as follows: Isolates with identical banding patterns were assigned the same arbitrary letter code (e.g., W, J, AF). IS6110 patterns that resembled but were not identical to one of these patterns were denoted by addition of a number (e.g., W14, W23).

**Variable Number of Tandem Repeats**

Tandem repeat loci ETR-A to ETR-E were amplified by polymerase chain reaction (PCR) and analyzed by gel electrophoresis to generate variable number of tandem-repeat (VNTR) allele profiles, as described by Frothingham and Meeker-O’Connell (35). Each digit of the allele profile represents the number of tandem-repeat copies at a particular locus. The patterns were compared against the PHRI TB Center VNTR database (~500 isolates) and profiles at the Durham Veterans Affairs Medical Centers database (745 isolates, 85 in the W phylogenetic lineage) (19).

**Streptomycin/Isoniazid/Rifampin/Ethambutol and Ethionamide Susceptibility Testing**

Primary drug resistance to streptomycin/isoniazid (INH)/rifampin/ethambutol (SIRE) was determined by using TB susceptibility Quad Plate I and II (Remel No. 3501). M. tuberculosis cultures were determined to be resistant to antimicrobial agents at concentrations ≥1 µg/mL for INH and rifampin, 7.5 µg/mL for ethambutol, and 10 µg/mL for streptomycin. The MIC for streptomycin was determined for representative samples according to standard methodology by using the agar diffusion assay (36). Isolates were subcultured onto 7H10/ADC medium (Difco, Detroit, MI) containing 2 to 500 µg/mL of streptomycin. The NYU Bureau of Laboratories performed pyrazinamide (PZA) testing (37).

**DNA Sequencing Streptomycin (rpsL) and INH (katG) Resistance Genes**

Analyses of the rpsL and katG genes were performed on a MicroGene Clipper 2 Dye Automated DNA Sequencer (Visible Genetics, Toronto, Ontario, Canada). Sequencing in the 5' and 3' directions was carried out simultaneously by using a two-dye system (Cy 5 and Cy 5.5; two-dye filter subsystem) to confirm mutations. The primers used to generate amplicons for sequencing were provided by Visible Genetics.

**W-Strain Family Genotype**

The principal genetic group was determined for each isolate as described (21). In addition, specific IS6110 insertion site mapping probes were used to determine the presence of insertions in the origin of replication and the NTB chromosomal region (20,25). The region flanking the deletion in the DR locus commonly associated with the W family was analyzed by PCR amplification and comparative hybridization (23).

**Results**

**The W Family Strain Collection**

The W14 fingerprint pattern was identified in a database of >13,000 clinical specimens representing approximately 9,000 patients and 3,563 distinct IS6110 fingerprint patterns genotyped at the PHRI TB Center. From the total IS6110 fingerprint patterns archived (N=3,563), 357 similar yet distinct patterns representing 1,498 isolates were grouped into the W family database by using standard matching criteria. Twenty isolates had an identical hybridization pattern (W14); 6 other isolates shared closely related IS6110 fingerprint patterns (3 W23 and 3 W26). These 26 isolates with three IS6110 patterns (W14, W23, and W26)
form the W14 group, a subgroup within the W phylogenetic lineage.

Molecular Features of W14 Group Isolates

The W14 group was originally defined on the basis of IS6110 patterns. The group includes three closely related fingerprints (W14, W23, and W26), with two characteristic common motifs (denoted by brackets in Figure 1A). Motifs A and B are unique to the W14 group. Southern blot hybridization of the Pvu II membrane with the 5'-IS6110 probe confirmed that the difference in IS6110 pattern between W23 and W26 was the product of two independent IS6110 insertions, rather than the outcome of a single RFLP event (Figure 1B).

PGRS probing showed a distinct hybridization pattern (P00026) for all 26 isolates in this group (Figure 1C). The P00026 pattern was used to segregate the W14 group isolates from all other clinical samples, including other W family groups in a PGRS database of >600 isolates.

Results from five chromosomal loci (ETR-A to ETR-E) were combined as a VNTR allele profile. All isolates in the W14 group had the VNTR allele profile 42445. This allele profile was unique to the W14 group when compared with the isolates from the PHRI TB Center VNTR database (n=>500) and an independent collection archived in Durham VA Medical Center database (n=745). Together, these databases include 85 isolates of the W family, including isolates from Asia, Africa, the former Soviet Union, and the United States. However, the allele profile 42445 was closely related to other patterns found in members of the W family. The most common profile for the W family is 42435; all profiles identified to date within the W family differed from this profile by a single allele. These VNTR profiles include 32435, 42436, 42437, 4253, and the W14 group, which is defined by 42445 (19). Unlike other members of the W phylogenetic lineage, the W14 group has a unique spoligotype arbitrarily designated as S00069 (Figure 2). Spoligopattern S00069 differed from S00034 associated with other reported W family members by the absence of spacer number 40. Patterns obtained by rehybridizing the same membranes used for IS6110 and PGRS fingerprinting with the DRab probe suggested that the disruption of spacer 40 was not due to an IS6110 insertion (Figure 3). An IS6110 insertion within spacer 40 would have generated two DR hybridizing bands, including one of a higher molecular weight, as found in strain CDC1551 (Figure 3) (38). Pattern S00069 was unique to the W14 group when compared with the Wadsworth spoligotype database of >2,500 isolates. The disruption of spacer 40 did not affect either chromosomal flanking region. Hence, in agreement with the S00034 spoligotype that groups all W family isolates analyzed, the region upstream from spacer 35 was deleted in isolates with S00069.

All members of the W14 group have high-level resistance (MIC>500 µg/mL) to streptomycin (STRR). In addition, one W14 and one W23 isolate were resistant to INH (INHR), one W14 was INHR and ethambutol (EMBR) resistant, two W14 isolates were ethionamide (ETHR) resistant, and one W14 was ETHR and pyrazinamide (PZR) resistant (Figure 4). Except for the 26 isolates in the W14 group, no W family isolate from a U.S. case in our collection was streptomycin monoresistant.

DNA sequencing of the gene commonly associated with streptomycin resistance, the rpsL gene, identified a single mutation (codon 43: AAG→AGG; Lys→Arg) in all the 26 W14 group isolates in this study. Three isolates were resis-
tant to INH. Sequence analysis of the katG gene in these isolates revealed that the one INH<sup>R</sup>-W23 isolate had a genetic alteration at codon 315 (AGC→ACC; Ser→Thr), while the two INH<sup>R</sup>-W14 isolates had the same single nucleotide insertion generating a frameshift mutation at codon 283 (CTG→ATG), which results in a termination upstream at codon 310, previously unreported (Figure 4).

All 26 isolates in the W14 group were linked to the W family based on all the secondary typing methods used. All had katG codon 463 sequence (CTG; Leu) and gyrA codon 95 sequence (ACC; Thr), placing them in genetic group 1 in the broad evolutionary framework outlined by Sreevatsan et al. (20,21). IS<sub>6110</sub> insertion site mapping showed that all 26 isolates had the A1 insertion in the origin of replication and a single IS<sub>6110</sub> copy in the NTF chromosomal locus (20,25).

Demographic Features of the W14 Cases

For the 22 patients in the W14 group, the mean age was 40 (range: 1 to 83 yrs) (Table). All were US-born, and 17 (77%) were of non-Hispanic black origin. In patients for whom HIV serology data were available, 74% (14 of 19) were HIV seropositive. Including HIV serology, 84% (16 of 19) had at least one of the known risk factors for TB (i.e., HIV, history of incarceration, homelessness, alcohol abuse, or intravenous drug use). A review of all drug-treatment regimen records of the W14 cases showed that only one patient had streptomycin as part of TB chemotherapy.

Investigation of 18 cases from NYC identified 107 contacts (mean of 6 contacts per case). Among contacts, 24% were PPD (purified protein derivative) positive. Investigations conducted by the NYC TB Control Program did not link any patients in this cohort or identify any new contacts with active disease.

Discussion

In this study the combination of multiple and independent molecular typing techniques permitted the identification and sub-grouping of a distinct branch of the W phylogenetic lineage, the W14 group. Twenty isolates with an identical IS<sub>6110</sub> pattern (W14) and an additional six (three W23 and three W26) isolates with closely related IS<sub>6110</sub> fingerprint patterns were identified in a fingerprint archive of >13,000 isolates. The W14 group was first typed to the "W" family fingerprint database of 1,498 isolates and then later subgrouped as a distinct branch of the W family. Five independent molecular techniques were used to confirm the distinctiveness of W14 isolates. The IS<sub>6110</sub> fingerprint pattern had two distinct motifs, unique within the PHRI database (Figure 1A). In addition, VNTR, PGRS, and spoligotyping patterns distinguished the W14 group from all other isolates analyzed by these methods in our collection, which includes >1,000 VNTR results, 600 PGRS images, and >2,500 spoligotype samples (Figure 1-3).

All isolates in the W14 group were resistant to high levels of streptomycin (>500 µg/mL), 20 were mono-streptomycin resistant, and the remaining 6 isolates were poly-resistant. Sequencing analysis confirmed that all isolates shared the same mutation in the rpsL gene in codon 43, a mutation previously associated with high levels of streptomycin resistance (MIC>500; (39)). The genetic alteration that directs the change of lysine to threonine in codon 43 of the rpsL gene is one of the two most frequently reported mutations associated with streptomycin resistance. Together, they account for approximately 53% of streptomycin-resistant
Genotyping data of 187 mono-streptomycin-resistant clinical isolates, including the 83 from NYC, identified the W14 group as the only epidemiologically important cluster of patients. Six additional, unrelated clusters of two cases each were identified in the PHRI fingerprint database. From the PHRI archive, no streptomycin-susceptible isolates with the W14 molecular characteristics have been identified to date from NYC or other locations (samples fingerprinted: n= >12,000; NYC: n= >6,000). Taken together, the molecular data point to the acquisition of streptomycin resistance before dissemination of this strain into the community.

Several strains have continued to acquire additional secondary drug resistances (Figure 4), including INH, ETH, PZA, and EMB (W14: 1 STR<sup>R</sup> INH<sup>R</sup>, 1 STR<sup>R</sup> INH<sup>R</sup> EMB<sup>R</sup>, 2 STR<sup>R</sup> ETH<sup>R</sup>, 1 STR<sup>R</sup> PZA<sup>R</sup> ETH<sup>R</sup>; W23: 1 STR<sup>R</sup> INH<sup>R</sup>). None of the isolates examined acquired resistance to rifampin. In a previous study, we elucidated a plausible pathway for the evolution of the W-MDR strain in NYC (W-index strain) in the early 1990s (17). In that study, the emergence and spread of the W-MDR strain in NYC was believed to have started by acquisition of streptomycin resistance, followed by INH and rifampin resistance, but the outbreak did not occur until the MDR-genotype was developed. In contrast, the W14 group of isolates spread after streptomycin resistance was acquired; subsequently, additional resistance developed, creating a group of poly-resistant variants. Sequencing data in combination with IS6110 Southern blot hybridization were used to demonstrate that INH resistance had developed on two occasions independently, once in W14 and once in a W23 (Figure 4) isolate. The W23 katG substitution on codon 315 is found on the same codon as in the W-MDR index strain but involves a different nucleotide. DNA sequence analysis of different resistance target genes provides molecular markers to speculate the stepwise building of polyresistant strains.

Since the late 1980s, molecular methods have been gradually integrated into the study of TB epidemiology and control. In addition to augmenting conventional epidemiologic investigations, molecular typing has been used to identify previously unrecognized point-source cases and in two reports was used to confirm transmission in a social setting (41,42). More recently, molecular typing with surveillance data uncovered an epidemiologically significant strain cluster (n=43), the W4 group, from New Jersey, that has no apparent common point source or patient links (19). The concordance of demographic and geographic data with molecular methods and the lack of concrete links in the W4 cases suggest a combination of reactivation and recent transmission. Likewise in this study, despite the demographic, geographic and molecular grouping, we could not establish any patient linkage in the W14 group. Nonetheless, the extent and validity of the molecular data demonstrated that these 22 patients were infected with either the same strain (16 W14 cases) or one of two closely related isolates (3 W23 or 3 W26).

This study has a number of limitations. PHRI’s collection from NYC reflects a convenience sample that is approximately 44% of the total number of culture-positive TB cases reported to the NYC TB Control Program in 1992 to 1999. Of the W14 group patients, all were U.S.-born, >70% were HIV positive, and most (73%) were 25 to 50 years of age at diagnosis. Despite the demographic homogeneity in these
patients, no epidemiologic links were established. Furthermore, the limited sampling suggests an actual denominator of W14 group cases larger than reported in this study, which aids in explaining the inability to establish epidemiologic links in the patients. Patient interviews were initially conducted as part of routine contact investigation; patients were not reinterviewed once molecular cluster information was available. Therefore, in-depth investigation of cases that appeared from molecular and surveillance data to be related was not available, which limits the inferences we can make on the observed lack of epidemiologic links.

Given the lack of overt epidemiologic links indicative of recent transmission between cases, the possibility that the W14 strain group is endemic to the region should be considered. However, the streptomycin monoresistance suggests that this strain must have spread after the introduction and application of this drug for TB treatment in 1944. It is surprising that not a single susceptible, related strain has been identified in the region even following extensive molecular analysis of other W variants. If the W14 or its predecessor were endemic, only the streptomycin-resistant variant has managed to spread, again suggesting that this dissemination depended on a single event, and yet, contact investigation has failed to establish a link. We speculate that the W14 cluster with the large young HIV-seropositive population is likely the result of recent transmission; however, the lack of patient-to-patient links suggests a subgroup of cases caused by an endemic strain that developed in NYC sometime after streptomycin was introduced in 1944.

By using multiple independent molecular markers, we made inferences on the strains’ recent evolution. This study highlights the importance of tracking all types of drug-resistant strains to prevent the sequential development of multi-drug-resistant strains. The utility that these methods lend to TB control will rely on the efficiency of integrating both surveillance and contact-tracing information with the molecular data in a proactive manner for more well-informed TB investigations.

Acknowledgments
We thank H. Marasco, A. Ravikovitch, and W. Eisner for assistance with patient and fingerprint databases. We thank Visible Genetics for providing us with the Clipper Sequencer and sequencing kits on a trial basis.

This research was supported in part by CDC’s National Tuberculosis Genotyping and Surveillance Network Cooperative Agreement. This is publication 75 from the Public Health Research Institute, Tuberculosis Center.

Dr. Bifani was a graduate student in the Public Health Research Institute Tuberculosis Center in New York City at the time this article was written. The center was established in January 1992 in response to the reemergence of TB in New York. He is currently at the Institut Pasteur de Lille France working as a postdoctoral fellow.

References


