Genomic Deletions Classify the Beijing/W Strains as a Distinct Genetic Lineage of *Mycobacterium tuberculosis*

Anthony G. Tsolaki,^{1*} Sebastien Gagneux,¹ Alexander S. Pym,¹
Yves-Olivier L. Goguet de la Salmoniere,¹ Barry N. Kreiswirth,²
Dick Van Soolingen,³ and Peter M. Small¹

Division of Infectious Diseases and Geographic Medicine, Stanford University Medical Center, Stanford, California 94305¹;
Tuberculosis Center, Public Health Research Institute, 225 Warren St., Newark, New Jersey 07103-3535²;
and Mycobacteria Reference Department, Diagnostic Laboratory for Infectious Diseases and
Perinatal Screening, National Institute of Public Health and the Environment,
Bilthoven, The Netherlands³

Received 5 November 2004/Returned for modification 14 January 2005/Accepted 3 April 2005

Beijing/W strains of *Mycobacterium tuberculosis* are geographically widespread and hypervirulent. To enhance our understanding of their origin and evolution, we sought phylogenetically informative large sequence polymorphisms (LSPs) within the Beijing/W family. Comparative whole-genome hybridization of Beijing/W strains revealed 21 LSPs, 7 of which were previously unreported. We show that some of these LSPs are unique event polymorphisms that can be used to define and subdivide the Beijing/W family. One LSP (RD105) was seen in all Beijing/W strains and thus serves as a useful marker for the identification of this family of strains. Additional LSPs (RD142, RD150, and RD181) further divided this family into four monophyletic subgroups, demonstrating a deeper population structure than previously appreciated. All Beijing/W strains were also observed to have an intact *pkas15/1* gene that is involved in the biosynthesis of a phenolic glycolipid, a putative virulence factor. A simple PCR assay using these Beijing/W strain-defining deletions will facilitate molecular epidemiological studies and may assist in the identification of the molecular basis of phenotypes associated with this important lineage of *M. tuberculosis*.

Molecular epidemiological studies have identified a genetically related group of *Mycobacterium tuberculosis* strains called Beijing/W that are widespread in many regions of the world (3, 10). Strains from this group have been associated with drug resistance, most notably in an outbreak of multidrug-resistant tuberculosis in New York (1, 4, 5, 35) and more recently in Russian prisons (31, 32). This association with drug resistance led to the hypothesis that strains from the Beijing/W group might be hypermutable (25), but this has recently been refuted (37). In Vietnam, Beijing/W strains were also associated with relapses and treatment failures (17). In the laboratory, these strains have a hypervirulent phenotype in mice and elicit a distinct immune response with a reduced level of cytokines, resulting in a failure to induce Th1 type immunity (19, 20) of which a polyketide synthase-derived phenolic glycolipid (PGL) is thought to play a principal role (26). Beijing/W strains also have a faster growth rate within human monocytes than do control strains (18). The bacterial genetic factors underlying this virulence are unknown, but the genetic relatedness of these Beijing/W strains suggests clonal expansion (19, 25, 29). It has even been hypothesized that vaccination with *Mycobacterium bovis* BCG may protect less well against Beijing/W strains, suggesting that they are escape variants (11, 19, 36).

Beijing/W strains share genetic markers, such as similar

IS6110 restriction fragment length polymorphism patterns and spoligotypes (3, 15). Currently, Beijing/W strains are principally identified by the number of spacers in the direct repeat (DR) region of the *M. tuberculosis* genome, namely, spoligotype S00034, or ST1, which is characterized by the deletion of spacers 1 to 34 (36) and, in a global survey of spoligotypes, type S00034 was found to be the most common (9). These genetic markers have been powerful and widely used in identifying strains for molecular epidemiology studies of tuberculosis outbreaks. Large sequence polymorphisms (LSPs) have been described in numerous studies of organisms from the *M. tuberculosis* complex, giving important insights into their evolution and biology (6, 12, 21, 23, 33). An improved understanding of the degree of LSPs in the Beijing/W family will provide markers that will enable a more refined placing of these strains within the broader genetic population structure of the *M. tuberculosis* species.

In the present study, we seek to classify the Beijing/W strains from a phylogenetic perspective by identifying new LSPs that are unique-event polymorphisms (UEPs). Our data show that the Beijing/W strains are monophyletic within the *M. tuberculosis* species and also reveal further subpopulations within this family. PCR screens with LSPs can be used to facilitate further epidemiological studies to elucidate the prevalence of these strains in the worldwide epidemic of tuberculosis. Furthermore, this new robust classification provides a rational sampling framework for the selection of strains for exploring fundamental biological questions, such as whether there are specific bacterial factors that promote the geographic dispersion of the Beijing/W family.

* Corresponding author. Mailing address: Department of Infectious Diseases and Microbiology, Centre for Molecular Microbiology and Infection, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom. Phone: 44-207-594-3094. Fax: 44-207-594-3095. E-mail: a.tsolaki@imperial.ac.uk.

MATERIALS AND METHODS

Strain selection and study design. A total of 179 *M. tuberculosis* strains were used in the present study. In order to identify the genomic deletions in the Beijing/W family of *M. tuberculosis*, reference strains that represent this family were obtained from the groups that originally described them. Nine strains of *M. tuberculosis* were provided by D. Van Soolingen (Bilthoven, The Netherlands) representing “Beijing” (16, 36) and five strains were provided by B. Kreiswirth (Newark, NJ) representing “W” (4). LSPs identified in these prototype strains were compared to deletions identified in a separate study of 100 randomly selected clinical isolates from San Francisco (33). For verification, genomic deletions that were found to be common to all Beijing/W strains of *M. tuberculosis* were used to screen a further 65 clinical strains in San Francisco that had spoligotype S00034, using real-time PCR.

Identification of genomic deletions. Whole genomic hybridization by using the Affymetrix *M. tuberculosis* GeneChip was used to identify genomic deletions that were absent from the fully sequenced strain H37Rv. The precise location of each deletion was confirmed by PCR and DNA sequencing. A detailed explanation for the methodology can be found in our recent article (33).

Spoligotyping. All strains in the present study were analyzed by spoligotyping using a method that determines the position and number of spacers and DR sequences of the DR region of *M. tuberculosis* as previously described (14).

Rapid identification of Genomic LSPs and SNPs using Real Time-PCR. Rapid identification of genomic deletions in *M. tuberculosis* strains were performed by real-time PCR in 96-well plates with the following TaqMan primers and probes: for *rrs* (ribosomal 16S rRNA, internal control), L_primer CGT TCC CGG GCC TTG TAC, R_primer CGG GTG TTA CCG ACT TTC ATG, and VIC probe CAC CGC CCG TCA CG; for RD105, L_primer AAC GAA CTG CGC ACT GAA CTC, R_primer TCC CGC ACC GGT TGA G, and FAM probe AGA GTG GAC AGT TTC C; for RD207, L_primer TCC GGC TCA CAA GAG GAA GT, R_primer TGT GCC CCA TCG GTG TTC, and FAM probe TTT CGA AGC GTT CGT CA; for RD142, L_primer CGG CAA GAT CAC AAA AAT CCA, R_primer CCC GTA ACG GGT CCA ATG T, and FAM probe TCT TAG TGC AGC CTT C; and for RD150, L_primer TGC TCG GTG GGC AAC TG, R_primer CGG CAC AGC GAG ACG AA, and FAM probe AAC GAC ATC CAC GCA TT. Real-time PCR was also used to detect the single nucleotide polymorphism at *katG463* (CTG codon) and *gyrA95* (ACC codon) (30), with the following TaqMan primers: for *KatG463_SNP*, L_primer CCG AGA TTGCCA GCC TTA AG, R_primer GAA ACT AGC TGT GAG ACA GTC AAT CC, *KatG463_cgg_FAM* probe CAG ATC CGG GCA TC, and *KatG463_ctg_VIV* probe CCA GAT CCT GGC ATC; and for *GyrA95_SNP*, L_primer CCG AGA CCA TGG GCA ACT AC, R_primer GCG CTG CTG GGT ACC TCA TC, *GyrA95_agc_FAM* probe TAC GAC AGC CTG GTG CG, and *GyrA95_acc_VIC* probe TAC GAC ACC CTG GTG CG. Allelic discrimination was performed to detect SNPs and deletions comparing amplification signal with the 16S rRNA internal control by using an ABI 7000 Real Time Thermocycler with standard conditions.

Analysis of *pks15/1* polymorphism. The *pks15/1* locus is described to be polymorphic among members of the *M. tuberculosis* complex (7). This region was analyzed for variability by PCR and sequencing with the oligonucleotide primers *pks1L* CTG GGT TGG CCT GCA CGT GGG CCA TAA and *pks15R* GCC CCC GCA GAG GCG CCG GTT with the following conditions of 40 cycles of amplification: 94°C for 0.40 min, 62°C for 0.40 min, and 72°C for 2.00 min. The PCR products obtained were sequenced by using an ABI 377 autosequencer and standard conditions and sequences analyzed by using SeqMan (DNASTar, Inc., Madison, WI).

RESULTS AND DISCUSSION

Characteristics of Beijing/W genomic deletions. In order to identify LSPs that will identify and further classify the Beijing/W family, we analyzed reference Beijing/W strains and a collection of Beijing/W strains from San Francisco at the whole-genome level. Using the Affymetrix GeneChip microarray, a total of 21 distinct LSPs were identified that were present in the sequenced strain H37Rv but absent from the Beijing/W strains. Of these, 13 deletions were found in the 14 reference strains and a further 8 deletions were found in the Beijing/W strains from San Francisco (12, 33). The LSPs ranged in size from 224 to 11,985 bp and partially or completely deleted 91 *M.*

tuberculosis genes. The genomic address, genes deleted, and PCR primers used to amplify and sequence the LSPs boundaries are given in Table 1. The sequences of the LSPs boundaries were found to be conserved. These LSPs have been designated regions of difference (RD) by using the established convention (6). Fourteen LSPs have been described previously (33), while seven are new (RD139BW, RD144BW, RD168BW, RD200BW, RD213BW, RD220BW, and RD224c). Based on frequency, three classes of LSPs were identified: four common LSPs (RD105, RD149, RD152, and RD207) were found in all Beijing/W strains. Three (RD181, RD142, and RD150) were found variably deleted, and the remaining 14 LSPs were unique to a single strain.

These LSPs have deleted and disrupted genes that may have an influence on gene expression and perhaps even be advantageous to Beijing/W. These genomic regions are intriguing for further study. Among the common LSPs, RD105 is 3,467 bp in size and completely deletes Rv0072 and Rv0073 and truncates Rv0071 and Rv0074. Rv0071 has a derived protein sequence similar to proteins of group II intron maturases (22). A variable repeat was also identified located 21 nucleotides downstream of the ATG start codon of Rv0071 and 25 nucleotides before the RD105 deletion. This is a 9-bp sequence that is repeated five times in H37Rv while in Beijing/W strains variability in the number of repeats was observed ranging from three to nine repeats, with a median of seven repeats (results not shown) and is in frame with the open reading frame of Rv0071. The significance of this repeat region is unknown, but it displays homology to five VDP repeats at the N terminus of hypothetical proteins found in two *Streptococcus* spp. and *M. leprae* and to internalin, a putative peptidoglycan bound protein from *Listeria monocytogenes* (8). The location of this repeat region is in close proximity to Rv0071 and may be phase variable. Phase variation is a common mechanism of transcriptional gene regulation in many other pathogenic bacteria (34). Among the other common LSPs, RD207 is 7,399 bp and is responsible for the partial deletion of the DR region and seven genes, including IS6110 (28). One of these genes, Rv2819c, is of unknown function and was found to be downregulated in the attenuated strain H37Ra (27). RD207 is responsible for the nine-spacer S00034 spoligotype characteristic of Beijing/W strains (36) and predictably occurs in all 14 reference strains. RD149 (or RD3) and RD152 are associated with mobile genetic elements, e.g., prophage and hot spots for IS6110 insertion, respectively (6, 13). Of the unique LSPs, RD139BW and RD213BW are of particular interest. RD139BW deletes three genes involved in the *Kdp* complex, and a gene (*trcS*) that is part of a two-component regulatory system. In the SCID mouse model, mice infected with deletion mutants of *trcS*, died more rapidly than those infected with wild-type bacteria, suggesting a role of *trcS* in virulence (24). RD213BW is 2,119 bp in size and deletes a section of the *pks1* gene. Polymorphisms in the *pks15/1* genes have previously been described, one of which curtails the production of PGL (7). Considering these parallel deletion events in the *pks15/1* region, a positive selection pressure favoring the loss of PGL production might exist under certain circumstances (26). However, most deletions are thought to be slightly deleterious to the fitness of the *M. tuberculosis* (12). In summary, there is anecdotal evidence suggesting that deleted genes in the Beijing/W strains may contribute to virulence, but

TABLE 1. Characteristics of deleted sequences found in *M. tuberculosis* Beijing/W family

Region of difference	Position		Size (bp)	Gene(s) deleted	Sequence		
	Start	Stop			L primer	R primer	
105	79567	83034	3,467	Rv0071, Rv0072, Rv0073, Rv0074	GGAGTCGTTGAGGGGTGTTTCATCAGCTCAGTC	CGCCAAGGCCGCATAGTCACCGGTG	
108	96640	99257	2,617	<i>hycE</i> , Rv0088, Rv0089, Rv0090	CGCCGCGTCTGCCCTCTT	GACCAGGACACCGCGCAGATACG	
110a	254000	257596	3,596	<i>fadD4</i> , <i>nadR</i> , Rv0213c	GAACTGCCCGCTGATCGAAGAATGG	GGTGTGATGCCGCTGTATGTGC	
127	746518	747805	1,287	<i>fabD2</i> , Rv0650	GACTCCGATCTGGAGGAGCTG	CAGGCCGCGCAGAAAGTCCAG	
129	769380	770002	622	<i>end</i> , Rv0669c	GAAAGCCGCTCGTGGTCAGGTTG	GCTGCGGCGGCTCCATC	
139BW ^a	1153020	1156702	3,683	<i>kdpA</i> , <i>kdpB</i> , <i>kdpC</i> , <i>ncS</i>	GATCCTGATCATCCCGTTGCGCTGTGCTT	AGGACCTGACCTGACCGACCTGTGTCAT	
142	1332182	1335033	2,851	Rv1190, Rv1191, Rv1192, <i>sigI</i>	TCGCGAGCAGCAACACGAC	TCACCTTCATTTCCAGCGGCACT	
144BW ^a	1498102	1499472	1,371	<i>dmg</i> , Rv1330c	TGCGCCACCTGCAGCGGTGATGAGTT	GTACAGACCGGTGTGGCCAATGCCGTT	
149	1779264	1788512	9,248	Rv1572c, Rv1573, Rv1574, Rv1575, Rv1575A, Rv1576c, Rv1577c, Rv1578c, Rv1579c, Rv1580c, Rv1581c, Rv1582c, Rv1583c, Rv1584c, Rv1585c, Rv1586c	CATGTACCCCTGGCCCGACGGGTC	CCAGGCGATCTTGACACGCGCACAC	
150	1896862	1899349	2,487	Rv1671, Rv1672c, Rv1673c, Rv1674c	TGTGGCGTGCGTCCGCAATAG	CGGGACGGCAACGGGTGAT	
152	1986636	1998621	11,985	<i>culI</i> , <i>picD</i> , Rv1754c, Rv1756c, Rv1757c, Rv1760, Rv1761c, Rv1762c, Rv1763, Rv1764, Rv1765c, <i>wug22</i>	CCGGGTTGAGCAATCGGATATCAGTGAC	TGGGATAGTTCAGGTGGCCATCGTGGCAT	
165	2153059	2153384	325	Rv1907c	GCCTTCGCTGCCCGTGTCTG	GCTCGCGGGGTGCTGGTGT	
166	2153074	2161125	8,051	<i>ace4a</i> , <i>fadB5</i> , <i>fur4</i> , <i>katG</i> , <i>lppC</i> , Rv1907c, Rv1910c, Rv1913, Rv1914c	CCAGGGTTGGCAGCCGTGCGTCTTGAA	GTACTGCTGTGCGCAAGCGCATAGAGAAGAT	
168BW ^a	2190641	2194794	4,154	Rv1937, <i>phbB</i> , Rv1939, <i>nb4</i> , Rv1941, Rv1942c	TTGTCGTGACCGCGTCGGGCAAGAACA	ACGTTGCGCTGCGGGTAGACCTCATCAATT	
181	2535429	2536140	711	Rv2262c, Rv2263	CGCAACGGCGCGGTGAACTCT	CGGGCGGCTGCGGGGAACCTT	
182a	2546463	2548424	1,961	<i>cyp121</i> , Rv2274c, Rv2275	CGGTGGGGGCTCTGCGTGGTC	AGGCGTTCGGGGATGCGTCTT	
200BW ^a	3004381	3006631	2,251	Rv2687c, Rv2688c, Rv2689c	AGGTGGAATCATTAACGCCGCGCAGCAT	CGGTAGCTGCGTGGCGACACGAA	
207	3120521	3127920	7,399	Rv2814c, Rv2815c, Rv2816c, Rv2817c, Rv2818c, Rv2819c, Rv2820c	GACGAGTTCCGGCTCAAAATGT	CCCCGGGAGGAAACAGAA	
213BW ^a	3292399	3294517	2,119	<i>pksl</i>	GAGGTGGCGTACCGGCTGCTGATTT	ATCGAAGAGTTGCTGTCGCGCACCGTT	
220BW ^a	3489095	3489318	224	None	TTTACCGCGCGAGGTAGAGCGGCTGTT	TGGACCGGCGCAATTCTGCAAGGCCGAAT	
224c ^a	3683267	3686632	3,366	<i>atsB</i> , Rv3300c	GCCTCGACCATCGAGATCCT	CATCGGGTAGCTGTGTGCT	

^a New deletions that were identified in this study.

FIG. 1. Genetic diversity in the Beijing/W family of *M. tuberculosis*. Markers were identified that define this lineage within *M. tuberculosis* and also divide it into distinct subgroups. LSPs present in H37Rv and absent from the interrogated Beijing/W isolates are shown in black. Each row represents an isolate, and each column is a region of difference. All strains had the 7-bp insertion at *pks15/1* and had CTG/ACC codons at *katG463/gyrA95*. Superscript a: strains indicated by a number only were from Tsolaki et al. (33), strains indicated by "Beijing" plus a number were from D. Van Soolingen, and strains indicated by "TN" numbers were from B. N. Kreiswirth.

FIG. 2. Beijing/W family of *M. tuberculosis* is monophyletic. The phylogeny shows that RD105 defines the Beijing/W family and is further subdivided into four groups by RD181, RD150, and RD142. All of the Beijing/W strains contain an intact *pks15/1* gene and may produce functional PGL.

no firm associations were observed. However, the frequency of these deletions suggests that some are useful markers for defining the Beijing/W family of *M. tuberculosis*, which may signify a clonal expansion of these strains through selective pressures. It has been postulated that Beijing/W strains could have emerged as escape variants resulting from insufficient protection by *M. bovis* BCG vaccination (11, 19, 36). This is an intriguing hypothesis, and these new molecular markers will facilitate future studies on the spread the Beijing/W strains and their role in the tuberculosis epidemic.

Genomic deletions place Beijing/W strains in a monophyletic grouping. In our recent work, we described a phylogeny for the *M. tuberculosis* species constructed from the 68 LSPs identified in 100 clinical isolates of *M. tuberculosis* from San Francisco (12). We took this phylogeny and compared the LSPs from the 14 reference Beijing/W strains to observe where they may fit within the *M. tuberculosis* phylogenetic tree. All of the 14 reference strains were found to have RD105, RD207, RD152, and RD149 LSPs in common with those of the East Asia clade of this phylogeny (12). In order to verify these observations, we performed spoligotyping on all 100 isolates from San Francisco and also found that each of the 23 strains in the East Asia phylogenetic clade had the S00034 spoligotype characteristic of the Beijing/W strains (Fig. 1). Furthermore, all of these strains were also found to be principle genetic group 1, having the *katG463/gyrA95* genotype (30), which is another characteristic of Beijing/W strains (3). RD150 was only observed among six strains from East Asia clade (12, 33) and not in the 14 reference strains. Moreover, an additional seven (unique) Beijing/W deletions (RD108, RD110a, RD127, RD129, RD165, RD166, and RD182a) were found only in the strains from East Asia clade, showing that our broader sampling from another data set, refined the population genetic structure within this family (12, 33).

In agreement with previous studies (6, 12), three types of LSPs were identified from this analysis. The first two types include repetitive DNA, mobile genetic elements (e.g., RD149), and LSPs resulting from homologous recombination between IS6110 (e.g., RD152 and RD207). These are likely to occur independently in different strains. In contrast, a third type of LSPs has been identified that are not associated with

either repetitive DNA or mobile genetic elements. This latter type (e.g., RD105, RD142, RD150, and RD181) truncate genes at specific positions and are therefore likely to represent UEPs (6, 12). Based on this definition, the Beijing/W lineage is principally defined by deletion RD105, which occurs in all of the Beijing/W strains. RD149, RD152, and RD207 also occur in all of the Beijing/W strains but are associated with mobile genetic elements and therefore may not be UEPs (Fig. 1). RD142, RD150, and RD181 are variably deleted in Beijing/W strains but are monophyletic for subdivisions of the Beijing/W group. RD142 and RD150 were only observed in strains with RD181 and RD105. These subdivisions are useful for further refining the phylogenetic structure of this lineage and may identify strains within the Beijing/W group with differing phenotypes.

Taken together, these data show that the Beijing/W family of strains is monophyletic and therefore a true clonal lineage within *M. tuberculosis*. The phylogeny of the Beijing/W lineage of *M. tuberculosis* (Fig. 2) is a robust classification, providing specific markers that allow rapid and simple identification. It not only defines the Beijing/W lineage but also subdivides it further, demonstrating a refined population structure within this family. This new classification also provides a robust framework for the selection of representative isolates for further study in the laboratory and in molecular epidemiological studies. To illustrate this and to further validate this new typing tool, we screened by using real-time PCR an additional 65 strains from San Francisco that had spoligotype S00034 for these lineage-defining deletions. All 65 strains had RD105 deleted, thereby confirming its utility in identifying Beijing/W strains. These strains were further examined for the presence of RD181, RD150, and RD142 and were found to be variably deleted 59, 12, and 1 time, respectively (i.e., these sequences were not deleted from all strains). This new classification will be a valuable tool for understanding the contribution of Beijing/W strains in the worldwide pandemic of tuberculosis and the evolution and pathogenesis of the tubercle bacterium. Selection of strains for further study using these Beijing/W-defining LSPs could have important implications for the evaluation of new diagnostics and the efficacy of vaccines in different

geographic regions. Beijing/W is one of a number of genotypes of *M. tuberculosis* influencing the current worldwide epidemic of tuberculosis. It has been proposed that a prevalent genotype in Delhi may be a close relative or progenitor of Beijing/W (2). Deletion analysis of this Delhi group, as well as other genotype families, will refine further their origin and evolution.

***pks15/1* polymorphisms are absent in all groups of the Beijing/W lineage.** A recent study described polymorphisms in the *pks15/1* region of mycobacteria that has given insights into the hypervirulence of Beijing/W strains (7, 36). In *M. tuberculosis* 210 (a Beijing/W strain) and in *Mycobacterium canettii* the *pks15/1* region is an intact open reading frame, whereas in *Mycobacterium bovis* there is a 6-bp deletion and in the H37Rv reference strain there is a 7-bp deletion (7). The *pks15/1* region is involved in the production of a PGL, which is immunomodulatory and associated with a hypervirulent phenotype in the mouse model, including the inhibition of the proinflammatory immune response (26). Loss of this gene could therefore account for the various phenotypes attributed to the Beijing/W strain family.

We investigated the *pks15/1* region in the Beijing/W strains from the present study to determine whether all four subgroups could potentially produce PGL. PCR amplification and sequencing of this *pks15/1* region was performed on the 102 Beijing/W strains (14 reference strains, 23 strains from the East Asia clade [12], and the 65 strains from San Francisco). The results revealed that all four subgroups of Beijing/W strains had an intact *pks15/1* open reading frame (Fig. 1) and potentially could produce PGL. This confirms and expands the findings of a study which used this polymorphism to refine the phylogeny of the *M. tuberculosis* complex (21).

Concluding remarks. Comparative whole-genome hybridization was previously used to provide insights into the evolution of *M. tuberculosis* (6, 23, 33). We applied the same approach to investigate genetic diversity in the Beijing/W family, a set of strains that are geographically widely disseminated. LSPs were found that were associated with this family of strains and that define it as a lineage within *M. tuberculosis*. Interestingly, a group of LSPs were also identified that subdivide the Beijing/W family into at least four exclusive subgroups, raising the possibility that there are phenotypic differences even within the Beijing/W family. Future studies attributing phenotypes to the Beijing/W family should ensure all genetic subgroups are represented. We thus have characterized the Beijing/W family within the broader phylogeny of *M. tuberculosis* and have identified robust markers that will help to elucidate the origin and role of this lineage in the global tuberculosis pandemic.

ACKNOWLEDGMENTS

We gratefully acknowledge Pablo J. Bifani for help in the selection of strains and Natalia Kurepina, Ruth Griffin, Brian Robertson, and Saurabh Johri for helpful comments on the manuscript.

This study was funded by National Institute of Allergy and Infectious Diseases grant AI34238 (to P.M.S.).

REFERENCES

- Agerton, T. B., S. E. Walway, R. J. Blinkhorn, K. L. Shilkret, R. Reeves, W. W. Schluter, B. Gore, C. J. Pozsik, B. B. Plikaytis, C. Woodley, and I. M. Onorato. 1999. Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. *Clin. Infect. Dis.* **29**:85–95.
- Bhanu, N. V., D. van Soolingen, J. D. van Embden, L. Dar, R. M. Pandey, and P. Seth. 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis* **82**:105–112.
- Bifani, P. J., B. Mathema, N. E. Kurepina, and B. N. Kreiswirth. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol.* **10**:45–52.
- Bifani, P. J., B. Mathema, Z. Liu, S. L. Moghazeh, B. Shopsin, B. Tempalski, J. Driscoll, R. Frothingham, J. M. Musser, P. Alcaba, and B. N. Kreiswirth. 1999. Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* **282**:2321–2327.
- Bifani, P. J., B. B. Plikaytis, V. Kapur, K. Stockbauer, X. Pan, M. L. Lutfey, S. L. Moghazeh, W. Eisner, T. M. Daniel, M. H. Kaplan, J. T. Crawford, J. M. Musser, and B. N. Kreiswirth. 1996. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA* **275**:452–457.
- Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeyer, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**:3684–3689.
- Constant, P., E. Perez, W. Malaga, M. A. Laneelle, O. Saurel, M. Daffe, and C. Guilhot. 2002. Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex: evidence that all strains synthesize glycosylated *p*-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the *pks15/1* gene. *J. Biol. Chem.* **277**:38148–38158.
- Drams, S., P. Dehoux, M. Lebrun, P. L. Goossens, and P. Cossart. 1997. Identification of four new members of the internalin multigene family of *Listeria monocytogenes* EGD. *Infect. Immun.* **65**:1615–1625.
- Filioli, L., J. R. Driscoll, D. van Soolingen, B. N. Kreiswirth, K. Kremer, G. Valetudie, D. D. Anh, R. Barlow, D. Banerjee, P. J. Bifani, K. Brudey, A. Cataldi, R. C. Cooksey, D. V. Cousins, J. W. Dale, O. A. Dellagostin, F. Drobniewski, G. Engelmann, S. Ferdinand, D. Gascoyne-Binzi, M. Gordon, M. C. Gutierrez, W. H. Haas, H. Heersma, E. Kassa-Kelembho, H. M. Ly, A. Makristathis, C. Mammina, G. Martin, P. Mostrom, I. Mokrousov, V. Narbonne, O. Narvskaya, A. Nastasi, S. N. Niobe-Eyangoh, J. W. Pape, V. Rasolof-Razanamparany, M. Ridell, M. L. Rossetti, F. Stauffer, P. N. Sufys, H. Takiff, J. Texier-Maugein, V. Vincent, J. H. de Waard, C. Sola, and N. Rastogi. 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J. Clin. Microbiol.* **41**:1963–1970.
- Glynn, J. R., J. Whiteley, P. J. Bifani, K. Kremer, and D. van Soolingen. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. *Emerg. Infect. Dis.* **8**:843–849.
- Hermans, P. W., F. Messadi, H. Guebexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, D. Frommel, et al. 1995. 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.* **171**:1504–1513.
- Hirsh, A. E., A. G. Tsolaki, K. DeRiemer, M. W. Feldman, and P. M. Small. 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc. Natl. Acad. Sci. USA* **101**:4871–4876.
- Ho, T. B., B. D. Robertson, G. M. Taylor, R. J. Shaw, and D. B. Young. 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20-kb variable region in clinical isolates. *Yeast* **17**:272–282.
- Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. van Embden. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**:907–914.
- Kremer, K., J. R. Glynn, T. Lillebaek, S. Niemann, N. E. Kurepina, B. N. Kreiswirth, P. J. Bifani, and D. van Soolingen. 2004. Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J. Clin. Microbiol.* **42**:4040–4049.
- Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakus, J. M. Musser, and J. D. van Embden. 1999. 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.* **37**:2607–2618.
- Lan, N. T., H. T. Lien, B. Tung le, M. W. Borgdorff, K. Kremer, and D. van Soolingen. 2003. *Mycobacterium tuberculosis* Beijing genotype and risk for treatment failure and relapse, Vietnam. *Emerg. Infect. Dis.* **9**:1633–1635.
- Li, Q., C. C. Whalen, J. M. Albert, R. Larkin, L. Zukowski, M. D. Cave, and R. F. Silver. 2002. Differences in rate and variability of intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. *Infect. Immun.* **70**:6489–6493.
- Lopez, B., D. Aguilar, H. Orozco, M. Burger, C. Espitia, V. Ritacco, L. Barrera, K. Kremer, R. Hernandez-Pando, K. Huygen, and D. van Soolingen. 2003. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* **133**:30–37.
- Manca, C., L. Tsenova, A. Bergtold, S. Freeman, M. Tovey, J. M. Musser,

- C. E. Barry III, V. H. Freedman, and G. Kaplan. 2001. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN- α /beta. *Proc. Natl. Acad. Sci. USA* **98**:5752–5757.
21. Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch. 2004. Macro-array and bioinformatic analyses reveal mycobacterial "core" genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**:483–496.
 22. Matsuura, M., J. W. Noah, and A. M. Lambowitz. 2001. Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20**:7259–7270.
 23. Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J. Infect. Dis.* **186**:74–80.
 24. Parish, T., D. A. Smith, S. Kendall, N. Casali, G. J. Bancroft, and N. G. Stoker. 2003. Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infect. Immun.* **71**:1134–1140.
 25. Rad, M. E., P. Bifani, C. Martin, K. Kremer, S. Samper, J. Rauzier, B. Kreiswirth, J. Blazquez, M. Jouan, D. van Soolingen, and B. Gicquel. 2003. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* **9**:838–845.
 26. Reed, M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, and C. E. Barry III. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431**:84–87.
 27. Rindi, L., N. Lari, and C. Garzelli. 2001. Genes of *Mycobacterium tuberculosis* H37Rv downregulated in the attenuated strain H37Ra are restricted to *M. tuberculosis* complex species. *New Microbiol.* **24**:289–294.
 28. Sampson, S. L., M. Richardson, P. D. Van Helden, and R. M. Warren. 2004. IS6110-mediated deletion polymorphism in isogenic strains of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **42**:895–898.
 29. Smith, N. H., J. Dale, J. Inwald, S. Palmer, S. V. Gordon, R. G. Hewinson, and J. M. Smith. 2003. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc. Natl. Acad. Sci. USA* **100**:15271–15275.
 30. Sreevatsan, S., X. Pan, K. E. Stockbauer, N. D. Connell, B. N. Kreiswirth, T. S. Whittam, and J. M. Musser. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. USA* **94**:9869–9874.
 31. Tounghousova, O. S., A. Mariandyshev, G. Bjune, P. Sandven, and D. A. Caugant. 2003. Molecular epidemiology and drug resistance of *Mycobacterium tuberculosis* isolates in the Archangel prison in Russia: predominance of the W-Beijing clone family. *Clin. Infect. Dis.* **37**:665–672.
 32. Tounghousova, O. S., P. Sandven, A. O. Mariandyshev, N. I. Nizovtseva, G. Bjune, and D. A. Caugant. 2002. Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. *J. Clin. Microbiol.* **40**:1930–1937.
 33. Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small. 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proc. Natl. Acad. Sci. USA* **101**:4865–4870.
 34. van der Woude, M. W., and A. J. Baumler. 2004. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* **17**:581–611.
 35. van Rie, A., R. M. Warren, N. Beyers, R. P. Gie, C. N. Classen, M. Richardson, S. L. Sampson, T. C. Victor, and P. D. van Helden. 1999. Transmission of a multidrug-resistant *Mycobacterium tuberculosis* strain resembling "strain W" among noninstitutionalized, human immunodeficiency virus-seronegative patients. *J. Infect. Dis.* **180**:1608–1615.
 36. van Soolingen, D., L. Qian, P. E. de Haas, J. T. Douglas, H. Traore, F. Portaels, H. Z. Qing, D. Enkhsaikan, P. Nymadawa, and J. D. van Embden. 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J. Clin. Microbiol.* **33**:3234–3238.
 37. Werngren, J., and S. E. Hoffner. 2003. Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. *J. Clin. Microbiol.* **41**:1520–1524.