JOURNAL OF BACTERIOLOGY, Sept. 2010, p. 4562–4570 Vol. 192, No. 18
0021-9193/10/$12.00 doi:10.1128/JB.00536-10

Massive Gene Duplication Event among Clinical Isolates of the *Mycobacterium tuberculosis* W/Beijing Family▽†

Pilar Domenech,[1] Gaëlle S. Kolly,[1,2] Lizbel Leon-Solis,[1] Ashley Fallow,[1] and Michael B. Reed[1]*

Research Institute of the McGill University Health Centre, 1625 Pine Ave. West, Montreal, Quebec, Canada H3G 1A4,[1] and Faculté des Sciences de la Vie, Ecole Polytechnique Fédérale de Lausanne, Station 19, 1015 Lausanne, Switzerland[2]

Received 10 May 2010/Accepted 7 July 2010

As part of our effort to uncover the molecular basis for the phenotypic variation among clinical *Mycobacterium tuberculosis* isolates, we have previously reported that isolates belonging to the W/Beijing lineage constitutively overexpress the DosR-regulated transcriptional program. While generating *dosR* knockouts in two independent W/Beijing sublineages, we were surprised to discover that they possess two copies of *dosR*. This *dosR* amplification is part of a massive genomic duplication spanning 350 kb and encompassing >300 genes. In total, this equates to 8% of the genome being present as two copies. The presence of IS*6110* elements at both ends of the region of duplication, and in the novel junction region, suggests that it arose through unequal homologous recombination of sister chromatids at the IS*6110* sequences. Analysis of isolates representing the major *M. tuberculosis* lineages has revealed that the 350-kb duplication is restricted to the most recently evolved sublineages of the W/Beijing family. Within these isolates, the duplication is partly responsible for the constitutive *dosR* overexpression phenotype. Although the nature of the selection event giving rise to the duplication remains unresolved, its evolution is almost certainly the result of specific selective pressure(s) encountered inside the host. A preliminary *in vitro* screen has failed to reveal a role of the duplication in conferring resistance to common antitubercular drugs, a trait frequently associated with W/Beijing isolates. Nevertheless, this first description of a genetic remodeling event of this nature for *M. tuberculosis* further highlights the potential for the evolution of diversity in this important global pathogen.

Mycobacterium tuberculosis—the bacterium responsible for almost 2 million human deaths each year due to tuberculosis (TB)—exhibits a highly clonal population structure and, unlike many pathogenic bacteria, shows little evidence of having recently acquired exogenous DNA from unrelated organisms via horizontal gene transfer and recombination (30). Thus, it seems that the primary mechanism by which *M. tuberculosis* is able to modify existing phenotypes is through an alteration in the complement of genetic material that it already has available. To date, most of the phenotypic variation that has been reported for *M. tuberculosis* involves single nucleotide polymorphisms (SNPs) and deletions, although transposition of conserved insertion sequence elements (e.g., IS*6110*) represents another potential mode for generating variability (1, 14, 16, 22). Recent SNP- and microarray-based whole-genome surveys have demonstrated that *M. tuberculosis* has evolved via clonal expansion into six major lineages that are referred to colloquially as Indo-Oceanic, East Asian, East African-Indian, Euro-American, West African-1, and West African-2. Aside from the widespread Euro-American lineage, each of these lineages tends to exhibit a strong degree of association with a particular geographic area (19, 29, 38).

Of all the major lineages described to date, it is undoubtedly the East-Asian lineage (commonly referred to as the Beijing or W/Beijing lineage) that has received the most attention in recent years. Originally described as the predominant genotype in China, where it is responsible for approximately 50% of TB cases, the W/Beijing lineage now accounts for at least 13% of all isolates worldwide (26). Alarmingly, the W/Beijing genotype appears to be emerging in several diverse regions, including countries of the former Soviet Union, South Africa, and Western Europe (9, 17, 26). Of further concern are the reports associating this recent epidemic spread with the appearance of drug resistance and treatment failure (12, 25). Together, these findings have led to mounting speculation that W/Beijing strains possess unique phenotypic attributes related to an increased ability to cause disease and to survive drug exposure within certain patient settings (3, 26). Despite the fact that the W/Beijing lineage is often treated as a single homogenous unit, these strains can be further classified into five evolutionary sublineages (groups 1 to 5) on the basis of the RD105, RD207, RD181, RD150, and RD142 deletions (14). Recent indications are that the aforementioned epidemiological traits may be restricted to specific W/Beijing sublineages (18, 20).

We have previously shown that strains of the W/Beijing lineage constitutively overexpress the DosR regulon and are natural mutants in one of the two sensor kinases (DosT) controlling activation of the DosR two-component regulatory system (13, 28). The present study was initiated as part of our ongoing effort to understand the regulatory control of the DosR regulon in this background. While generating a series of DosR disruption mutants, we discovered that members of the most recently evolved subgroups of the W/Beijing family possess two copies of the gene encoding DosR. Much to our astonishment, these two copies of *dosR* are actually part of a

* Corresponding author. Mailing address: RI-MUHC, Montreal General Hospital, Room Rs1-133, 1625 Pine Ave. West, Montreal, QC Canada H3G 1A4. Phone: (514) 934-1934. Fax: (514) 934-8261. E-mail: michael.reed@mcgill.ca.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 16 July 2010.

much larger chromosomal duplication event that encompasses more than 300 genes and spans 350 kb (kb). To the best of our knowledge, this is the first example of a duplication (large or small) that has been recorded for *M. tuberculosis* and serves to highlight the potential for the evolution of diversity in this global pathogen.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *M. tuberculosis* strains were grown in Middlebrook 7H9 broth (Difco) supplemented with 10% ADC (8.1 g/liter NaCl, 50 g/liter bovine serum albumin [BSA] fraction V [Calbiochem], 20 g/liter glucose), 0.2% glycerol, and 0.05% Tween 80 or on Middlebrook 7H11 agar (Difco) supplemented with 10% OADC (ADC plus 0.6 ml/liter oleic acid, 3.6 mM NaOH). Kanamycin (25 μg/ml), hygromycin (50 μg/ml), or 2% sucrose was added as needed. H37Rv (ATCC 27294) and *Mycobacterium bovis* (Ravenel) were purchased from the ATCC. The majority of the isolates examined in this study were collected between 2001 and 2007 from mostly foreign-born TB patients residing on the island of Montreal (29). HN878 and NHN5 were originally obtained from J. Musser (Methodist Hospital Research Institute, Houston, TX) (32); strains W4, W10, and W210 were obtained from B. Kreiswirth (Public Health Research Institute Center, New Jersey). CDC1551 was originally obtained from T. Shinnick (Centers for Disease Control and Prevention, Atlanta, GA) (36). Where necessary, our W/Beijing strain collection was supplemented with isolates obtained from San Francisco TB patients, kindly provided by S. Gagneux (National Institute for Medical Research, United Kingdom) and P. Small (Institute for Systems Biology, Seattle, WA) (14). *M. bovis* (BCG) and *Mycobacterium canetti* were obtained from the Pasteur Institute (Paris, France).

General nucleic acid techniques. Mycobacterial DNA was isolated according to the protocol of Pelicic et al. (27), and *M. tuberculosis* transformations were carried out as previously described (31). The oligonucleotide primers used for PCR and sequencing reactions are presented in Table S1 in the supplemental material. Where necessary, 5 to 10% DMSO was included in the PCRs. Southern hybridizations were carried out using the ECL direct nucleic acid labeling and detection system (GE Healthcare).

Construction of *dosR* disruption mutants. A 2.4-kb PCR fragment containing the *dosR* gene was generated using primers dosR-1 and dosS-1 and cloned into the XbaI site of pcDNA2.1 (Invitrogen). After sequencing, 386 bp of the *dosR* gene was replaced by inserting either the 1.5-kb hygromycin resistance cassette (*hyg*) from pHint (24) or the 1.2-kb kanamycin resistance (*kan*) cassette from pMV261 (33) into the BbsI and BlpI sites of *dosR*. Finally, the 3.5-kb *dosR*::*hyg*-containing fragment (or the 3.2-kb *dosR*::*kan* fragment) was excised and cloned into the mycobacterial shuttle vector, pPR23 (27). All selection and screening procedures for the isolation of homologous recombinants in H37Rv, HN878, and G4B1.2 were as previously described (11, 27). The G4B1.2-32 *dosR*::*hyg* mutant was used as the parental strain for generation of the double *dosR* knockouts (*dosR1*::*kan dosR2*::*hyg*) through the same approach, with only kanamycin being used in the initial selection process.

Microarray analysis. Whole-genome microarray analysis was carried out as described previously (10), with only minor modifications. Two micrograms of purified genomic DNA was labeled with either Cy3 or Cy5-NHS esters (GE Healthcare) via the aminoallyl indirect labeling method (39) using DNA polymerase I Klenow fragment and a mixture of random hexamers (Fermentas). The microarrays used in these experiments were provided by M. Behr (McGill University) and are composed of 70-bp oligonucleotides (TB Array-Ready Oligo Set; Operon) printed in duplicate. For genomic DNA comparisons, at least two independent arrays were analyzed for each pair of samples (Cy3/Cy5 and Cy5/Cy3).

Cloning and sequencing the beginning, end, and junction regions of the duplication. To obtain the sequences of these three regions, the following fragments were cloned: (i) a 9.6-kb HindIII-SspI fragment from G4B1.2-30 (*dosR1*::*hyg*) was used for isolating the sequence at the beginning of the duplication, (ii) a 6.2-kb AgeI fragment from the G4B1.2 wild type was used for obtaining the sequence at the end of the duplication, and (iii) the sequence of the junction region was obtained from a 13-kb NdeI-SspI fragment from the G4B1.2-32 (*dosR2*::*hyg*) strain. All DNA fragments were gel purified and ligated with pBluescript II KS+ (Stratagene). *Escherichia coli* NEB 10-beta (New England BioLabs) transformants were plated on LB agar plus hygromycin (200 μg/ml) or LB agar plus ampicillin (100 μg/ml) in the case of the unmarked fragment isolated from G4B1.2. For the latter, positive colonies were identified by screening for the presence of the *alr* gene by colony blotting using the ECL labeling and detection system. Sequence analysis was carried out at McGill University and the Genome Québec Innovation Centre.

qRT-PCR. Techniques used in the preparation of cDNA for quantitative real-time PCR (qRT-PCR) analysis of *dosR* expression are described elsewhere (13). RNA samples were prepared from at least two independent biological replicates for each strain under investigation. Primers dosR-F and dosR-R were used for quantification of *dosR* gene expression, while primers specific for the *sigA* housekeeping gene (sigA1-F and sigA1-R) were used to normalize the amount of cDNA template added to each sample. qRT-PCRs were carried out in quadruplicate using a model 7300 real-time PCR system and Power SYBR green PCR master mix (Applied Biosystems) according to the manufacturer's recommendations. The relative standard curve method was used for quantification as previously described (10, 13).

Microarray data accession number. Microarray data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE20917.

Nucleotide sequence accession numbers. Sequence data for the beginning, junction, and end regions of the 350-kb duplication have been deposited in GenBank under accession numbers HM053705 to HM053707.

RESULTS

Generation of *dosR*::*hyg* recombinant strains in W/Beijing isolates reveals the presence of two copies of *dosR*. In previous studies we have demonstrated that strains of the W/Beijing lineage overexpress the DosR "dormancy" regulon in a constitutive manner (13, 28). To begin to understand the relevance of this phenotype to the pathogenesis of W/Beijing strains, we set out to inactivate *dosR* in clinical isolates G4B1.2 and HN878 belonging to subgroups 4 and 5 of the W/Beijing lineage, respectively. As a control, *dosR* was also inactivated in the H37Rv laboratory strain (non-W/Beijing). Southern blot analysis of the *dosR*::*hyg* recombinants and wild-type parental strains using a *dosR*-containing probe revealed the following (Fig. 1A). (i) For H37Rv, the expected 11.4-kb fragment was detected in the wild type, whereas two fragments of 6.9 and 5.5 kb were observed in the *dosR*::*hyg* derivatives (H37Rv-19 and -24) due to introduction of a PstI site present in the *hyg* cassette (Fig. 1B). (ii) For HN878 and G4B1.2, we observed a single fragment of 11.4 kbp in the wild type and three fragments of 11.4, 6.9, and 5.5 kb in the *dosR*::*hyg* clones (HN878-25 and -28 and G4B1.2-30 and -32). The latter result was completely unexpected and suggested the presence of both intact and mutated copies of *dosR* in these strains. We speculated that either there were two copies of *dosR* in HN878 and G4B.1.2 or only a single crossover event had occurred. To rule out the second possibility, we hybridized the same membrane with the pPR23 vector backbone, which confirmed that no part of the vector remained integrated (see Fig. S1A in the supplemental material). We therefore hypothesized that there are two copies of *dosR* in the HN878 and G4B1.2 backgrounds.

To confirm that this surprising finding was not just an artifact of the recombination process, additional Southern analysis using a second set of enzymes and probes was carried out in an attempt to demonstrate the presence of two copies of *dosR* in the W/Beijing wild-type isolates (see Fig. S1B to D in the supplemental material). Using the same *dosR* probe as above, we again obtained the expected results with the H37Rv wild-type (7.3-kb) and *dosR*::*hyg* (8.3-kb) strains. However, on this occasion we could clearly observe two fragments of 8.6 and 18 kb for wild-type HN878 and G4B1.2, consistent with there being two copies of *dosR* (see Fig. S1B). Likewise, two fragments were revealed in the HN878-25 and -28 (8.6- and 9.6-

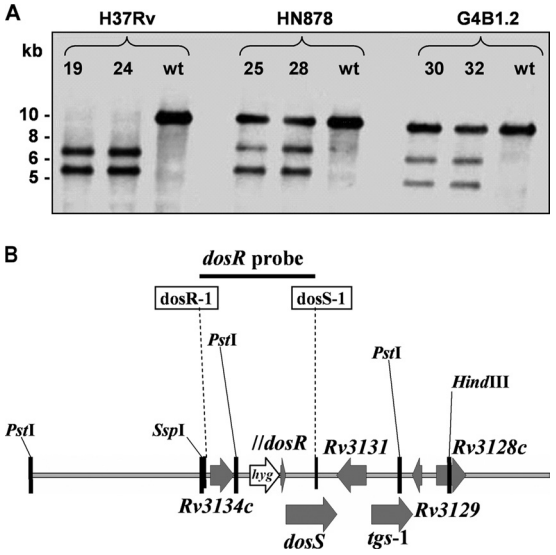

FIG. 1. Southern blotting reveals the presence of two copies of *dosR* in recombinant W/Beijing strains. Genomic DNA from wild-type and two independent *dosR*::*hyg* recombinants generated for H37Rv (clones 19 and 24) and the W/Beijing isolates HN878 (clones 25 and 28) and G4B1.2 (clones 30 and 32) was digested with PstI and transferred by Southern blotting. (A) The membrane was hybridized with a 2.4-kb PCR fragment (generated using dosR-1 and dosS-1) that includes Rv3134c, *dosR*, and 739 bp of *dosS*. Note the presence of an intact copy of *dosR* retained by the HN878 and G4B1.2 *dosR*::*hyg* strains. (B) Based on the published H37Rv sequence, a genetic map of the *dosR* region in the recombinant strains is shown. By replacing 486 bp of *dosR* with the *hyg* resistance cassette, an additional PstI site has been introduced.

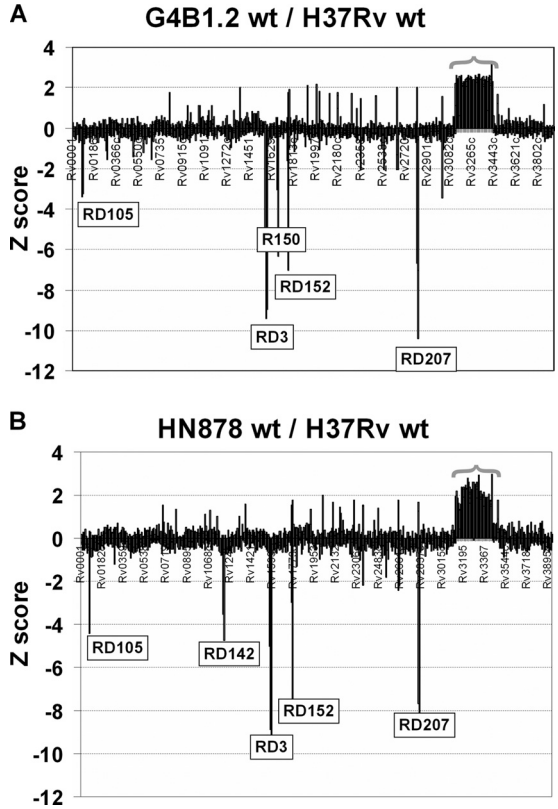

FIG. 2. Characterization of the W/Beijing duplication through whole-genome microarray. Comparative array hybridizations were carried out for genomic DNA prepared from wild-type W/Beijing isolates G4B1.2 (A) and HN878 (B) versus H37Rv. For each, the average Z scores from three arrays for each of the 3,924 genes present in H37Rv are shown. The RD105 and RD207 deletions unique to the W/Beijing lineage are indicated, as are the RD150 and RD142 deletions that are characteristic of group 4 and 5 W/Beijing isolates, respectively. The RD3 and RD152 deletions were also identified. The majority of the ~300 genes in the contiguous stretch from Rv3128c to Rv3427c showed a Z score of ≥2.0, consistent with the presence of a 350-kb duplicated region in G4B1.2 and HN878 relative to H37Rv.

kb), G4B1.2-30 (9.3- and 18-kb), and G4B1.2-32 (8.6- and 19-kb) *dosR*::*hyg* recombinants. The 8.3-kb (H37Rv-19 and -24), 9.6-kb (HN878-25 and -28), 9.3-kb (G4B1.2-30), and 19-kb (G4B1.2-32) fragments also cohybridize to a probe specific for the *hyg* gene, indicating that these copies of *dosR* have been inactivated through insertion of the *hyg* resistance cassette (see Fig. S1C). The same blot was reprobed with a *dosR*-specific probe that corresponds to the portion of *dosR* replaced by the *hyg* cassette during the generation of the *dosR*::*hyg* strains. For H37Rv, only a single fragment was detected in the wild type, with no signal detected for the *dosR* mutants. Two *dosR*-hybridizing fragments were again detected for wild-type HN878 and G4B1.2 (8.6 and 18 kb). For the W/Beijing recombinants, the alternate copy of *dosR* that was not detected by the *hyg* probe was recognized in this case (see Fig. S1D). In summary, the results confirm that the W/Beijing isolates HN878 and G4B1.2 are diploid for the *dosR* gene and that their *dosR*::*hyg* derivatives are haploid as they retain one functional copy. Interestingly, the two distinct copies of *dosR* have been inactivated independently in strains G4B1.2-30 (*dosR1*::*hyg*) and G4B1.2-32 (*dosR2*::*hyg*), respectively. Finally, it should be noted that there were some anomalies with the HN878 strain. First, the larger of the *dosR*-containing fragments (18 kb) identified in wild-type HN878 is much fainter than the smaller fragment. Second, a smaller product (9.6 kb) was observed in the *dosR* haploids in this background (*dosR2*::*hyg*; HN878-25 and -28) than for G4B1.2 (see Fig. S1). The explanation for these observations is given below.

The two copies of *dosR* are part of a much larger chromosomal duplication. In order to define the extent of the duplication event that gave rise to the two copies of *dosR* in the W/Beijing background, we carried out a whole-genome microarray analysis comparing the genome content of H37Rv with that of the wild-type G4B1.2 and HN878 isolates. Figure 2 shows the Z scores obtained for all 3924 H37Rv genes represented on the arrays (8) for the hybridizations comparing G4B1.2 with H37Rv (Fig. 2A) and HN878 with H37Rv (Fig. 2B). Z scores of ≥2.0 indicate that a particular gene is more abundant in the "test" strain (G4B1.2 or HN878) than in the reference strain (H37Rv). Conversely, Z scores of ≤2.0 indicate that a gene is more abundant in the reference strain (H37Rv) and are suggestive of a deletion in the test strains. Several deletions were detected; these included those characteristic of group 4 (RD105, RD207, and RD150) and group 5 (RD105, RD207, and RD142) W/Beijing isolates (14, 34) and the RD3 and RD152 deletions that have been described elsewhere (5, 35). The short RD181 deletion common to group 3 to 5 W/Beijing isolates was not detected due to the spacing of

TABLE 1. Functional categories of the 313 genes present in the 350-kb duplication

Functional category	No. of genes in duplication	% of total genes[a]
Virulence, detoxification, adaptation	15	0.38
Lipid metabolism	17	0.43
Information pathways	20	0.51
Cell wall and cell wall processes	35	0.89
Insertion sequences and phages	17	0.43
PE/PPE proteins	14	0.36
Intermediary metabolism and respiration	89	2.27
Regulatory proteins	21	0.54
Conserved hypothetical proteins	85	2.17
All genes in the duplication	313	7.98

[a] Relative to the complete H37Rv genome.

the oligonucleotides corresponding to the Rv2262c and Rv2263 genes affected by this deletion.

Strikingly, the microarray comparisons of G4B1.2 with H37Rv revealed a contiguous stretch of approximately 300 genes with a Z score of ≥2.0, suggesting that all these genes have been duplicated within the G4B1.2 isolate (Fig. 2). A very similar although slightly less-well-defined pattern was also seen for HN878 (discussed below). Hence, in these two isolates, there is a massive duplication that spans 350 kb of the chromosome and comprises all genes ranging from the Rv3128c to the Rv3427c genes. This equates to approximately 8% of the total H37Rv genome being present as two copies within these W/Beijing isolates. Table 1 shows the predicted functional categories of the 313 genes present in the 350-kb duplication as defined in the TubercuList (http://genolist.pasteur.fr/TubercuList/). Aside from the *dosR*/*dosS* regulatory system, the entire NADH dehydrogenase I (*nuoA* to *-N*) and succinate dehydrogenase (*sdhA* to *-D*) enzyme complexes are also present as two copies in the W/Beijing strains containing the duplication. Also noteworthy are the large number of genes encoding regulatory proteins (*mtrAB*, *whiB1* to *-B3*, and *whiB7*) and extracytoplasmic function sigma factors (σ^D, σ^F, σ^H, σ^J). Thus, any impact of the duplication could potentially extend well beyond the borders of this 350-kb region.

The twin copies of the duplicated sequence are tandemly arranged. We speculated that the two copies of the 350-kb region would most likely be arranged in a direct tandem duplication, the proposed structure of which is given in Fig. 3A. To confirm this particular arrangement, a series of Southern blot analyses were carried out for genomic DNAs prepared from the wild-type H37Rv, HN878, and G4B1.2 strains. When a probe specific for the Rv3127c gene was used, a single hybridization fragment was detected for all three strains, whereas two fragments were seen for HN878 and G4B1.2 when the same blot was rehybridized to a *tgs1* probe (Fig. 3B and C). This result confirms that the beginning of the duplication is located between Rv3127c and *tgs1*. On a separate blot, two fragments were detected for HN878 and G4B1.2 when an *alr* probe was used, whereas only one was detected when a *gadB*

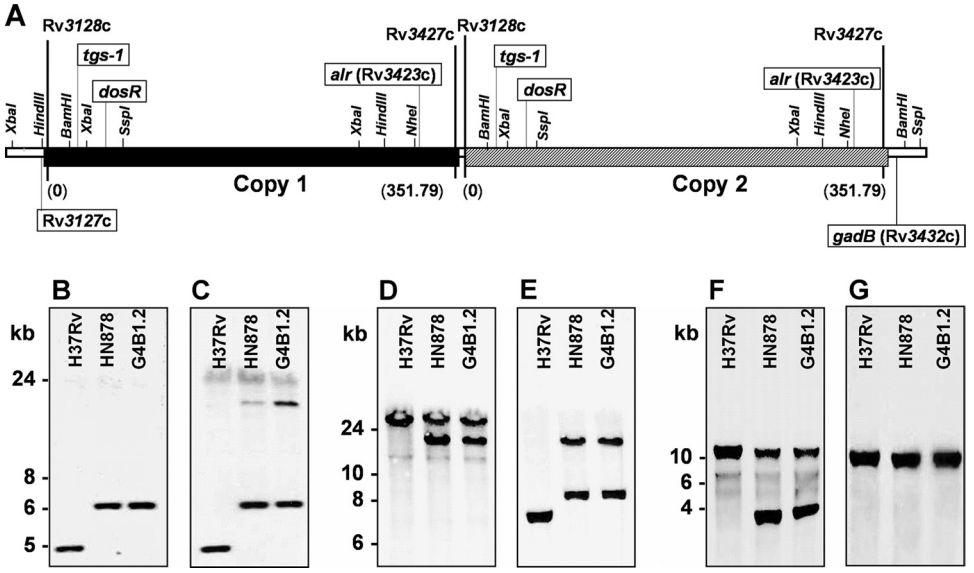

FIG. 3. The two copies of the 350-kb duplicated region are arranged in tandem. (A) A schematic representation of the duplicated region is shown. Genes Rv3128c and Rv3427c mark the approximate beginning and end of each duplicated segment, respectively. Genes Rv3127c and *gadB* flank the duplication externally. The length of the duplication is indicated in parentheses, and the two copies (Copy 1 and Copy 2) are arranged in a direct, tandem duplication. The arrangement of the genes indicated (boxed) was confirmed by Southern blotting of genomic DNA comparing wild-type H37Rv and the W/Beijing isolates HN878 and G4B1.2. The approximate locations of the restriction enzyme sites are based upon the published H37Rv sequence. (B and C) Southern blots confirming that the beginning of the duplication is located between Rv3127c (single copy) and *tgs1* (duplicated). DNAs were digested with XbaI and hybridized with a 620-bp probe derived from Rv3127c (B; primers Rv3127c-A and -B) or with a 920-bp *tgs1* probe (C; primers Rv3130-1 and Rv3130c-F). (D and E) Southern blots confirming the location of the junction of the duplication and that the two copies are arranged in a direct, tandem duplication. DNAs digested with HindIII and SspI were hybridized with a 540-bp *alr* probe (D; primers alr-C and alr-R) or with a 1-kb *dosR* probe [E; primers dosR-C and dosRrev(HindIII)]. (F and G) Southern blots confirming that the end of the duplication is located between Rv3427c (duplicated) and *gadB* (single copy). DNAs were digested with BamHI and NheI and hybridized with the 540-bp *alr* probe (F) or a 1.2-kb *gadB* probe (G; primers gadB-C and gadB-D).

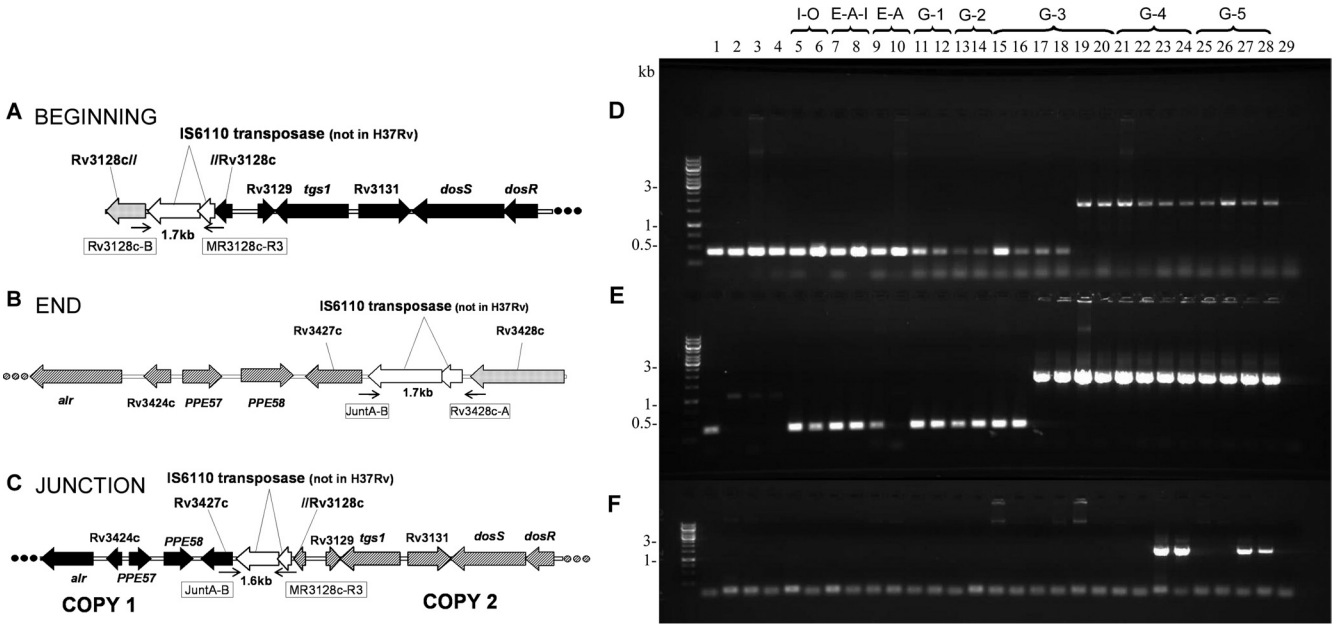

FIG. 4. Genetic map and PCR-based screening for identifying isolates bearing the 350-kb duplication. The genetic arrangement of the cloned fragments representing the beginning (A), end (B), and junction (C) regions of the duplication is shown. Genes present in "Copy 1" of the duplication are indicated in black, while those in "Copy 2" are in dark gray. The three sets of primers used to screen for the presence of the duplication in clinical isolates are highlighted (boxed). The screening method is based on detecting the IS*6110* transposase insertions (white) located in the beginning (A and D), end (B and E), and junction (C and F) regions. (D to F) Lanes: 1, *M. tuberculosis* West African-I; 2, *M. bovis* BCG; 3, *M. bovis*; 4, *M. canetti*; 5 and 6, *M. tuberculosis* Indo-Oceanic (I-O) lineage; 7 and 8, East African-Indian (E-A-I); 9 and 10, Euro-American (E-A); 11 and 12, W/Beijing group 1 (G-1); 13 and 14, W/Beijing group 2 (G-2); 15 to 20, W/Beijing group 3 (G-3); 21 to 24, W/Beijing group 4 (G-4); 25 to 28, W/Beijing group 5 (G-5); 29, negative control. In the example shown, the W/Beijing isolates in lanes 23 and 24 (group 4) and in lanes 27 and 28 (group 5) are positive in all three reactions, suggesting that they contain the full duplication.

probe was used (Fig. 3F and G). This confirms that the end of the duplication is located between *alr* and *gadB*. Finally, two distinct fragments are seen for HN878 and G4B1.2 when either an *alr* probe or a *dosR* probe is used to hybridize a third blot targeting the junction region of the duplication (Fig. 3D and E). Importantly, the 18-kb fragment cohybridizes with both probes, indicating that *alr* and *dosR* lie on the same HindIII-SspI fragment in this junction region, thereby confirming that copies 1 and 2 of the 350-kb sequence are arranged as a direct tandem duplication.

To define the beginning, end, and junction of the duplication more precisely, we cloned and sequenced each of these regions from isolate G4B1.2 and its *dosR*::*hyg* derivatives. The details of how these fragments were obtained are provided in Materials and Methods, and the resultant genetic maps are presented in Fig. 4A to C. Note that the sequences are almost identical to the published H37Rv sequence (8) except for a small number of 1.3-kb insertions comprising the genes associated with the IS*6110* insertion sequence element. At the beginning of the duplication, there is an IS*6110* inserted at position 744 of Rv3128c that results in this gene being truncated (Fig. 4A). Interestingly, this particular IS*6110* insertion has previously been reported as being characteristic of modern or "typical" W/Beijing isolates, with its location often referred to as the NTF region (3). At the end of the duplication, an IS*6110* has inserted itself between Rv3427c and Rv3428c (Fig. 4B). At the junction of the two copies of the duplication, an IS*6110* is inserted 307 bp downstream of Rv3427c and disrupts Rv3128c at position 740 (Fig. 4C).

Together, these sequence data confirm the direct, tandem arrangement of the HN878 and G4B1.2 duplication and support the microarray data indicating that the duplication encompasses genes Rv3128c to Rv3427c. The fact that copies of IS*6110* are found at the beginning, end, and junction of the duplication is highly suggestive that its mechanism of formation involved unequal homologous recombination mediated by complementary IS*6110* elements located on sister chromatids (2).

The 350-kb duplication is restricted to the most recently evolved W/Beijing isolates. The discovery of a large chromosomal duplication in two W/Beijing clinical isolates obviously raises a question regarding its frequency across distinct *M. tuberculosis* lineages. To address this question, we screened a panel of 126 isolates representing each of the major *M. tuberculosis* lineages (plus additional members of the *M. tuberculosis* complex [MTC], namely, *M. canetti*, *M. bovis*, *M. bovis* BCG, and *Mycobacterium caprae*) by PCR for the presence of the 350-kb duplication. For each isolate, three separate PCRs were carried out using primers anchored across the IS*6110* insertion sites located at the beginning (Fig. 4A), end (Fig. 4B), and junction (Fig. 4C) of the duplication. Only isolates positive in all three PCRs were considered to be potentially positive for the full-length duplication. A representative example of these results is shown in Fig. 4D to F. When isolates possess a copy of IS*6110* at the beginning of the duplication, a 1.7-kb PCR product is obtained (Fig. 4D, lanes 19 to 28), whereas a product of 421 bp indicates the absence of IS*6110* at this location (lanes 1 to 18). Similarly, the presence or absence of IS*6110* at

TABLE 2. Distribution of the 350-kb duplication in *M. tuberculosis* clinical isolates

Isolate group	n^a	No. of isolates with IS*6110* transposase at[b]:			No. of isolates with results confirmed by[d]:			
					Southern blotting		DNA microarray	
		Beginning	Junction	End	Dupl$^+$	Dupl$^-$	Dupl$^+$	Dupl$^-$
Non-W/Beijing	33	0	0	0[c]	ND	ND	ND	ND
Group 1	6	0	0	0	ND	ND	ND	ND
Group 2	2	0	0	0	ND	ND	ND	ND
Group 3	61	42	3	54	ND	ND	ND	ND
Group 4	14	14	6	13	6 (6)	8 (8)	3 (3)	1 (1)
Group 5	10	10	4	10	4 (4)	4 (4)	2 (2)	3 (3)
Total	126	66	13	77	10	12	5	4

[a] Number of isolates screened in each group.
[b] Isolates were screened by PCR for the presence of the IS*6110* transposase at each of the regions indicated. See Fig. 4 for further details.
[c] No "end" product was detected for 6 strains. This is due to the RD6 deletion that occurs in *M. bovis*, *M. canetti*, and some *M. tuberculosis* isolates (5).
[d] PCR screening results were confirmed by Southern blotting for 22 isolates and by DNA microarray for 9 isolates. The number of isolates in each group that were tested by the methods indicated is given in parentheses. Dupl$^+$, duplication positive; Dupl$^-$, duplication negative. ND, not done.

the end of the duplication is revealed by products of 1.7 kb (Fig. 4E, lanes 17 to 28) or 372 bp, respectively. Lastly, the appearance of the novel junction region is confirmed with a 1.6-kb PCR product (Fig. 4F, lanes 23, 24, 27, and 28). A negative result here is indicative of a lack of the junction region and of the 350-kb duplication. The results of these experiments are summarized in Table 2.

None of the non-W/Beijing isolates tested (9 Euro-American, 8 Indo-Oceanic, 8 East African-Indian, 4 West African, and 4 MTC) harbored copies of IS*6110* at the beginning or end of the duplication and, consequently, were also negative for the junction region (Table 2). The same result was obtained for the W/Beijing isolates belonging to groups 1 and 2. Among the 61 group 3 isolates that were tested, 54 contain a copy of IS*6110* at the end of the duplication and 42 of these were also positive for the IS*6110* at the beginning of the duplication. This result suggests that acquisition of the IS*6110* at the end of the duplication was an earlier evolutionary event than acquisition of IS*6110* at the beginning of the duplication. Three isolates of group 3 (5%) were also positive for the junction region, consistent with the original acquisition of the 350-kb duplication having taken place within this sublineage. All group 4 and 5 W/Beijing isolates that were tested have IS*6110* at the beginning of the duplication, and all but one of these was positive for IS*6110* at the end of the duplication. This peculiar isolate (group 4) was completely negative in the PCR screen for the end of the duplication and gave a smaller than expected hybridization product for a Southern blot probed with *alr*. Although not confirmed, it is likely that this strain contains the RD6 deletion that is variably deleted in a range of *M. tuberculosis* and MTC strains (5). Among the group 4 ($n = 14$) and group 5 ($n = 10$) W/Beijing isolates that we currently have available, 6 isolates belonging to group 4 (43%) and 4 isolates belonging to group 5 (40%) contain the junction region and were therefore considered positive for the 350-kb duplication.

Confirmation of the validity of the PCR screening results was achieved by Southern blotting of the group 4 and 5 isolates using an *alr* probe. Nine of these samples (including the HN878 and G4B1.2 isolates analyzed above) were further confirmed by whole-genome microarray (Table 2) (GEO accession number GSE20917). In order to select unique isolates for the array experiments, IS*6110* restriction fragment length polymorphism analysis (RFLP) was carried out, the results of which are shown in Fig. S2 in the supplemental material. In summary, for our strain set, we detected the duplication only in group 3 to 5 isolates, which is highly suggestive that its acquisition has been a relatively recent event along the evolutionary path of the W/Beijing lineage.

The 350-kb W/Beijing duplication is unstable *in vitro*. As indicated above, several anomalies were noted for our HN878 stock when analyzed by either Southern blot or microarray analysis (Fig. 2; see also Fig. S1 in the supplemental material). In addition, in a parallel project under way in our laboratory, we recently isolated a HN878 *tgs1*::*hyg* strain that was found not to contain the 350-kb duplication. Unlike the other clinical isolates analyzed in this study, HN878 is a strain that was originally isolated in the mid 1990s and has passed through multiple laboratories since that time (32). Together, these observations raise the possibility that the 350-kb duplication is in the process of being lost from *in vitro* stocks of HN878. We, and others, have previously noted a related example of *in vitro* genomic decay that is associated with the loss of phthiocerol dimycocerosate (PDIM) biosynthesis from H37Rv (10). In order to address this issue, an *in vitro* stock culture of HN878 was plated to obtain isolated single colonies, and 40 individual clones were screened for the presence of the 350-kb duplication by PCR as shown in Fig. 4. Only 7 clones appeared to contain an intact duplication, and 33 were found to have lost it either partially or completely. We then selected two of these laboratory-derived HN878 subclones for further analysis via genomic microarray. As demonstrated in Fig. S3 in the supplemental material, the HN878-wt-45 subclone (wt, wild type) contains the full-length 350-kb duplication, whereas HN878-wt-27 has lost it completely. We have also identified some interesting variations on this theme for a small number of our HN878 *dosR*::*hyg* recombinants. For example, additional array comparisons have revealed that strains HN878-25 and -28 (*dosR2*::*hyg*) (Fig. 5B; see also Fig. S1 in the supplemental material) retain a partial duplication of approximately 220 kb encompassing genes Rv3128c to Rv3324c (see Fig. S3). Similarly, clone HN878-58 (*dosR1*::*hyg*) (Fig. 5B) retains a duplication of approximately 295 kb that extends from Rv3180c to

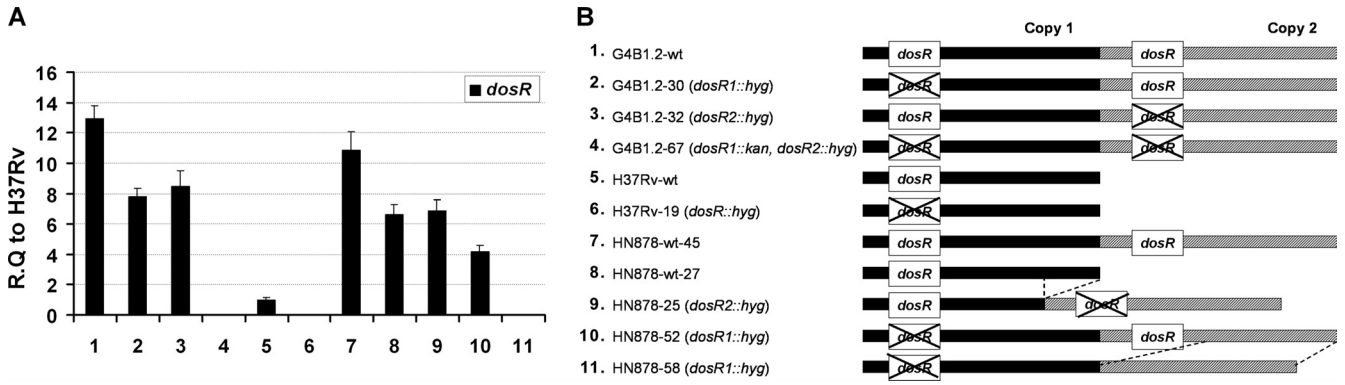

FIG. 5. Impact of the W/Beijing duplication on *dosR* expression. (A) qRT-PCR analysis of *dosR* expression in the wild type (wt) and the single or double *dosR* mutant strains indicated in panel B. The expression levels of *dosR* are normalized to the *sigA* housekeeping gene and are plotted relative to H37Rv. Each sample was assayed in quadruplicate, and at least two independent biological replicates were analyzed for each strain. Data from a single representative experiment are presented. Error bars represent standard deviation. (B) Schematic representation depicting the duplication and *dosR* genotypes for each of the strains analyzed in panel A. Further details for each strain can be found in the text.

Rv3427c (see Fig. S3) and therefore harbors only one copy of *dosR* that has been disrupted through insertion of the *hyg* cassette.

The duplication event is not responsible for the constitutive overexpression of *dosR* by W/Beijing isolates. To investigate the effect that two copies of *dosR* and/or the 350-kb duplication has on *dosR* expression levels, we quantified the expression of *dosR* by qRT-PCR in the strains indicated in Fig. 5. As seen previously, the *dosR* levels were 1 log higher in the wild-type G4B1.2 (Fig. 5A, lane 1) and HN878-45 (lane 7) isolates than in H37Rv (lane 5). Inactivation of either copy of the *dosR* gene leads to a reduction in *dosR* expression levels by 40 to 60% (lanes 2, 3, 9, and 10). A similar reduction was observed with the HN878-wt-27 strain that has lost the entire 350-kb duplication and contains only one copy of *dosR* (lane 8). As expected, no *dosR* expression was detected for any of the strains that lack a functional copy of *dosR*, including the double *dosR* knockout strain generated in the G4B1.2 background (G4B1.2-67 [*dosR1*::*kan dosR2*::*hyg*]; see Fig. S4 in the supplemental material) (Fig. 5A, lanes 4, 6, and 11). These results demonstrate that although there is an enhancement of *dosR* expression in strains bearing two copies of the *dosR* gene, the 350-kb duplication itself is not sufficient or responsible for the constitutive *dosR* phenotype displayed by W/Beijing isolates. This finding is consistent with a recent publication of ours demonstrating that the DosR regulon phenotype is associated with all group 2 to 5 W/Beijing isolates (13). In comparison, the duplication we describe herein is restricted to a subset of strains within the most recently evolved sublineages of the W/Beijing family.

DISCUSSION

To the best of our knowledge, the massive W/Beijing-specific duplication described herein is the first large-scale genomic remodeling event of its type ever to be recorded for *M. tuberculosis*. Moreover, what makes this duplication event particularly remarkable is that it has almost certainly evolved in response to some form of selective force exerted inside the host. While chromosomal duplications have been reported for other nonpathogenic mycobacteria, each of these has clearly been selected for through continuous *in vitro* passage. In *Mycobacterium smegmatis* mc^2155, a 56-kb duplication that is flanked by two copies of an IS*1096* element has been identified (37). The ATCC 607 progenitor does not contain the 56-kb duplication. Similarly, two independent tandem duplications have been described for the *M. bovis* BCG vaccine strain that was attenuated through continuous *in vitro* passage over a period of 13 years (4). DU1 is a 29-kb duplication that is present only in BCG Pasteur. DU2 initially arose as a 141-kb duplication and now exists in four different forms as a result of subsequent internal deletions.

Although it is difficult to completely exclude the possibility that the 350-kb W/Beijing duplication arose spontaneously *in vitro* during the process of bacterial isolation from patient samples, there are several lines of evidence that suggest that this is highly unlikely. First, the duplication is detected only in a select group of W/Beijing patient isolates, and aside from HN878, none of these has undergone multiple rounds of *in vitro* passage or extended periods of *in vitro* culture. Second, for the HN878 strain, which is the only one of these isolates that has spent any length of time in culture, we observe that the duplication appears to be highly unstable. This is not at all surprising given the presumed high energetic cost of maintaining duplicate copies of more than 300 genes in the absence of positive selective pressure. Third, if the acquisition of any part of the duplicated region enhanced *in vitro* "fitness" to any significant degree, we would expect to observe the entire 350 kb, or a subregion thereof, duplicated in many more isolates when cultivated *in vitro* than just the related group 3, 4, and 5 W/Beijing isolates described here. As highlighted above, no such duplication has been noted previously for *M. tuberculosis*.

Among 126 isolates tested, the 350-kb duplication was restricted to only 5% of group 3 isolates and approximately half of the group 4 and 5 isolates of the W/Beijing family, suggesting that the original duplication event has occurred relatively recently in the evolution of the *M. tuberculosis* genome. Our results indicate that the duplication initially arose in a group 3 isolate, the progeny of which independently gave rise to both the group 4 and 5 sublineages (note that group 5 is not a

sublineage of group 4) (14). Over time, a proportion of isolates appear to have resolved the duplication either because it has not been continuously selected or because a more stable mutational event (e.g., point mutation) has supplanted the need for the duplication. Indeed, gene duplications and amplifications appear relatively frequently within bacterial populations but typically disappear after just a few generations of growth in the absence of selection (2). They most often provide an adaptive response to the presence of toxic compounds or antibiotics or nutrient starvation or to compensate for a deleterious mutation (2). In addition, there are several examples, including *Vibrio cholerae* and *Haemophilus influenzae*, where duplications have been selected within the host in order to increase toxin or capsule production that results in enhanced virulence (6, 23).

The fact that such a large duplication is maintained with some frequency in at least two independent sublineages of the W/Beijing family would suggest that it is being maintained through selection in some specific host populations. Without selection, we suspect that this duplication would have been lost long ago, as appears to be occurring with our *in vitro* HN878 cultures. Exactly what that selective pressure is remains to be seen, although it is tempting to speculate that the duplication is associated with enhanced *in vivo* fitness (virulence) or antibiotic tolerance/resistance. In this regard, we note with great interest the two recent molecular epidemiological studies reporting a significant association between the more recently evolved W/Beijing sublineages and (i) an increased ability to transmit and cause disease in the Western Cape region of South Africa (18) and (ii) an increased ability to cause extrathoracic TB (20). In both of these studies, the W/Beijing strains referred to are the group 4 and 5 isolates, i.e., the same sublineages for which we observe the 350-kb duplication.

The global increase in BCG vaccination and the distribution of anti-TB drugs are two recent events that could potentially have contributed to selection of the W/Beijing duplication. The former possibility is particularly interesting in light of suggestions that the recent increase in mass immunization programs involving BCG may have inadvertently selected for the emergence of W/Beijing strains on a global basis (7, 21, 26). It is also tempting to speculate that the duplication may enhance the *in vivo* fitness of W/Beijing strains in the face of growing antibiotic use, given that the recent epidemic spread of W/Beijing strains is often associated with the development of drug resistance (3, 15, 25). To date, preliminary screens have failed to reveal any significant difference in the *in vitro* MICs when isolates harboring the 350-kb duplication (HN878-wt-45 and G4B1.2) are compared to those lacking the duplication (HN878-wt-27 and H37Rv) for any of the following compounds: isoniazid, rifampin, ethambutol, cycloserine, kanamycin, ofloxacin, or streptomycin. However, this certainly does not rule out the possibility of an effect of the duplication on drug sensitivity or tolerance *in vivo*. Clearly, additional studies are warranted in order to evaluate the potential implications of the massive W/Beijing duplication on the spread and development of TB disease.

ACKNOWLEDGMENTS

This work was supported by Canadian Institutes of Health Research (CIHR) grant MOP82931. M.B.R. is supported by a Peter Lougheed/CIHR New Investigator Award, while P.D. is supported by a Montreal General Hospital 175th Anniversary Fellowship.

We thank Christophe Guilhot for sharing the mycobacterial allelic exchange system and Peter Small for providing access to clinical isolates. We are also indebted to Marcel Behr and Fiona McIntosh for assistance with the microarray and RFLP experiments.

REFERENCES

1. **Alland, D., D. W. Lacher, M. H. Hazbon, A. S. Motiwala, W. Qi, R. D. Fleischmann, and T. S. Whittam.** 2007. Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of *Mycobacterium tuberculosis* and the utility of LSPs in phylogenetic analysis. J. Clin. Microbiol. **45:**39–46.
2. **Andersson, D. I., and D. Hughes.** 2009. Gene amplification and adaptive evolution in bacteria. Annu. Rev. Genet. **43:**167–195.
3. **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.
4. **Brosch, R., S. V. Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J. K. Inwald, P. Golby, J. N. Garcia, R. G. Hewinson, M. A. Behr, M. A. Quail, C. Churcher, B. G. Barrell, J. Parkhill, and S. T. Cole.** 2007. Genome plasticity of BCG and impact on vaccine efficacy. Proc. Natl. Acad. Sci. U. S. A. **104:**5596–5601.
5. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, 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. U. S. A. **99:**3684–3689.
6. **Cerquetti, M., R. Cardines, M. Giufre, A. Castella, M. Rebora, P. Mastrantonio, and M. L. Ciofi Degli Atti.** 2006. Detection of six copies of the capsulation b locus in a *Haemophilus influenzae* type b strain isolated from a splenectomized patient with fulminant septic shock. J. Clin. Microbiol. **44:** 640–642.
7. **Cohen, T., C. Colijn, and M. Murray.** 2008. Modeling the effects of strain diversity and mechanisms of strain competition on the potential performance of new tuberculosis vaccines. Proc. Natl. Acad. Sci. U. S. A. **105:**16302–16307.
8. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
9. **Cowley, D., D. Govender, B. February, M. Wolfe, L. Steyn, J. Evans, R. J. Wilkinson, and M. P. Nicol.** 2008. Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. Clin. Infect. Dis. **47:**1252–1259.
10. **Domenech, P., and M. B. Reed.** 2009. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown *in vitro*: implications for virulence studies. Microbiology **155:**3532–3543.
11. **Domenech, P., M. B. Reed, and C. E. Barry III.** 2005. Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect. Immun. **73:**3492–3501.
12. **European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis.** 2006. Beijing/W genotype *Mycobacterium tuberculosis* and drug resistance. Emerging Infect. Dis. **12:**736–743.
13. **Fallow, A., P. Domenech, and M. B. Reed.** 2010. Strains of the East Asian (W/Beijing) lineage of *Mycobacterium tuberculosis* are DosS/DosT-DosR two-component regulatory system natural mutants. J. Bacteriol. **192:**2228–2238.
14. **Gagneux, S., K. DeRiemer, T. Van, M. Kato-Maeda, B. C. de Jong, S. Narayanan, M. Nicol, S. Niemann, K. Kremer, M. C. Gutierrez, M. Hilty, P. C. Hopewell, and P. M. Small.** 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. **103:**2869–2873.
15. **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. Emerging Infect. Dis. **8:**843–849.
16. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. **193:**121–128.
17. **Hanekom, M., G. D. van der Spuy, N. C. Gey van Pittius, C. R. McEvoy, S. L. Ndabambi, T. C. Victor, E. G. Hoal, P. D. van Helden, and R. M. Warren.** 2007. Evidence that the spread of *Mycobacterium tuberculosis* strains with the Beijing genotype is human population dependent. J. Clin. Microbiol. **45:** 2263–2266.
18. **Hanekom, M., G. D. van der Spuy, E. Streicher, S. L. Ndabambi, C. R.**

McEvoy, M. Kidd, N. Beyers, T. C. Victor, P. D. van Helden, and R. M. Warren. 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. J. Clin. Microbiol. **45:**1483–1490.

19. **Hershberg, R., M. Lipatov, P. M. Small, H. Sheffer, S. Niemann, S. Homolka, J. C. Roach, K. Kremer, D. A. Petrov, M. W. Feldman, and S. Gagneux.** 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. **6:**e311.
20. **Kong, Y., M. D. Cave, L. Zhang, B. Foxman, C. F. Marrs, J. H. Bates, and Z. H. Yang.** 2007. Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: insights from epidemiologic and clinical characterization of the three principal genetic groups of *M. tuberculosis* clinical isolates. J. Clin. Microbiol. **45:**409–414.
21. **Kremer, K., M. J. van der Werf, B. K. Au, D. D. Anh, K. M. Kam, H. R. van Doorn, M. W. Borgdorff, and D. van Soolingen.** 2009. Vaccine-induced immunity circumvented by typical *Mycobacterium tuberculosis* Beijing strains. Emerging Infect. Dis. **15:**335–339.
22. **McEvoy, C. R., A. A. Falmer, N. C. Gey van Pittius, T. C. Victor, P. D. van Helden, and R. M. Warren.** 2007. The role of IS*6110* in the evolution of *Mycobacterium tuberculosis*. Tuberculosis (Edinb.) **87:**393–404.
23. **Mekalanos, J. J.** 1983. Duplication and amplification of toxin genes in *Vibrio cholerae*. Cell **35:**253–263.
24. **Ó Gaora, P., S. Barnini, C. Hayward, E. Filley, G. Rook, D. Young, and J. Tholea.** 1997. Mycobacteria as immunogens: development of expression vectors for use in multiple mycobacterial species. Med. Princ. Pract. **6:**91–96.
25. **Parwati, I., B. Alisjahbana, L. Apriani, R. D. Soetikno, T. H. Ottenhoff, A. G. van der Zanden, J. van der Meer, D. van Soolingen, and R. van Crevel.** 2010. *Mycobacterium tuberculosis* Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia. J. Infect. Dis. **201:**553–557.
26. **Parwati, I., R. van Crevel, and D. van Soolingen.** 2010. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. Lancet Infect. Dis. **10:**103–111.
27. **Pelicic, V., M. Jackson, J. M. Reyrat, W. R. Jacobs, Jr., B. Gicquel, and C. Guilhot.** 1997. Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U. S. A. **94:**10955–10960.
28. **Reed, M. B., S. Gagneux, K. Deriemer, P. M. Small, and C. E. Barry III.** 2007. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J. Bacteriol. **189:**2583–2589.
29. **Reed, M. B., V. K. Pichler, F. McIntosh, A. Mattia, A. Fallow, S. Masala, P. Domenech, A. Zwerling, L. Thibert, D. Menzies, K. Schwartzman, and M. A. Behr.** 2009. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. J. Clin. Microbiol. **47:**1119–1128.
30. **Smith, N. H., R. G. Hewinson, K. Kremer, R. Brosch, and S. V. Gordon.** 2009. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. Nat. Rev. Microbiol. **7:**537–544.
31. **Snapper, S. B., R. E. Melton, S. Mustafa, T. Kieser, and W. R. Jacobs, Jr.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol. Microbiol. **4:**1911–1919.
32. **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. U. S. A. **94:**9869–9874.
33. **Stover, C. K., V. F. de la Cruz, T. R. Fuerst, J. E. Burlein, L. A. Benson, L. T. Bennett, G. P. Bansal, J. F. Young, M. H. Lee, G. F. Hatfull, S. B. Snapper, R. G. Barletta, W. R. Jacobs, and B. R. Bloom.** 1991. New use of BCG for recombinant vaccines. Nature **351:**456–460.
34. **Tsolaki, A. G., S. Gagneux, A. S. Pym, Y. O. Goguet de la Salmoniere, B. N. Kreiswirth, D. Van Soolingen, and P. M. Small.** 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **43:**3185–3191.
35. **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. U. S. A. **101:**4865–4870.
36. **Valway, S. E., M. P. Sanchez, T. F. Shinnick, I. Orme, T. Agerton, D. Hoy, J. S. Jones, H. Westmoreland, and I. M. Onorato.** 1998. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N. Engl. J. Med. **338:**633–639.
37. **Wang, X. M., A. Galamba, D. F. Warner, K. Soetaert, J. S. Merkel, M. Kalai, P. Bifani, P. Lefevre, V. Mizrahi, and J. Content.** 2008. IS*1096*-mediated DNA rearrangements play a key role in genome evolution of *Mycobacterium smegmatis*. Tuberculosis (Edinb.) **88:**399–409.
38. **Wirth, T., F. Hildebrand, C. Allix-Beguec, F. Wolbeling, T. Kubica, K. Kremer, D. van Soolingen, S. Rusch-Gerdes, C. Locht, S. Brisse, A. Meyer, P. Supply, and S. Niemann.** 2008. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathog. **4:**e1000160.
39. **Yu, J., M. I. Othman, R. Farjo, S. Zareparsi, S. P. MacNee, S. Yoshida, and A. Swaroop.** 2002. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. Mol. Vis. **8:**130–137.