

# Simplified Model to Survey Tuberculosis Transmission in Countries without Systematic Molecular Epidemiology Programs

Juan Domínguez,<sup>1</sup> Fermín Acosta,<sup>1</sup> Laura Pérez-Lago, Dilcia Sambrano, Victoria Batista, Carolina De La Guardia, Estefanía Abascal, Álvaro Chiner-Oms, Iñaki Comas, Prudencio González, Jaime Bravo, Pedro Del Cid, Samantha Rosas, Patricia Muñoz, Amador Goodridge,<sup>2</sup> Darío García de Viedma<sup>2</sup>

Systematic molecular/genomic epidemiology studies for tuberculosis surveillance cannot be implemented in many countries. We selected Panama as a model for an alternative strategy. Mycobacterial interspersed repetitive unit–variable-number tandem-repeat (MIRU–VNTR) analysis revealed a high proportion (50%) of *Mycobacterium tuberculosis* isolates included in 6 clusters (A–F) in 2 provinces (Panama and Colon). Cluster A corresponded to the Beijing sublineage. Whole-genome sequencing (WGS) differentiated clusters due to active recent transmission, with low single-nucleotide polymorphism–based diversity (cluster C), from clusters involving long-term prevalent strains with higher diversity (clusters A, B). Prospective application in Panama of 3 tailored strain–specific PCRs targeting marker single-nucleotide polymorphisms identified from WGS data revealed that 31.4% of

incident cases involved strains A–C and that the Beijing strain was highly represented and restricted mainly to Colon. Rational integration of MIRU–VNTR, WGS, and tailored strain–specific PCRs could be a new model for tuberculosis surveillance in countries without molecular/genomic epidemiology programs.

Tuberculosis (TB) control depends on rapid diagnosis, efficient therapy, and control of transmission. New genotyping strategies in *Mycobacterium tuberculosis* have led to development of molecular epidemiology strategies that enable clarification of TB transmission dynamics (1–3). PCR-based molecular epidemiology approaches, namely mycobacterial interspersed repetitive unit–variable-number tandem-repeat (MIRU–VNTR), are being replaced by genomic epidemiology approaches. These approaches are based on the identification of single-nucleotide polymorphisms (SNPs) using whole-genome sequencing (WGS), which offers higher discriminatory power and precision in assigning transmission clusters (4–6).

The national TB-control programs of many resource-constrained countries with a high TB prevalence lack systematic molecular epidemiology analysis (7). These countries face insurmountable obstacles to availability of genomic epidemiology programs. Panama is an example of a country that, despite recently being classified in the upper-income category, has not implemented a systematic molecular/genomic epidemiology–based surveillance program (7). The lack of such a program, together with other limitations, means that Panama now has the second highest incidence of TB in Central America and the highest TB death rate (8). The recent expansion of the Panama Canal has attracted migrants into the country, which probably

Author affiliations: Instituto de Investigaciones Científicas y Servicios de Alta Tecnología, City of Knowledge, Panama (J. Domínguez, F. Acosta, D. Sambrano, V. Batista, C. De La Guardia, A. Goodridge); Instituto Conmemorativo Gorgas de Estudios de la Salud, Panama City, Panama (J. Domínguez, P. González, J. Bravo, P. Del Cid, S. Rosas); Hospital General Universitario Gregorio Marañón, Madrid, Spain (F. Acosta, L. Pérez-Lago, E. Abascal, P. Muñoz, D. García de Viedma); Instituto de Investigación Sanitaria Gregorio Marañón, Madrid (F. Acosta, L. Pérez-Lago, E. Abascal, P. Muñoz, D. García de Viedma); Centro Superior de Investigación en Salud Pública (FISABIO)–Universitat de València, Valencia, Spain (Á. Chiner-Oms); Instituto de Biomedicina de Valencia Consejo Superior de Investigaciones Científicas, Valencia (I. Comas); Centro de Investigación Biomédica en Red en Epidemiología y Salud Pública (CIBERESP), Madrid (I. Comas); Universidad Complutense de Madrid, Madrid (P. Muñoz); Centro de Investigación Biomédica en Red Enfermedades Respiratorias (CIBERES), Madrid (P. Muñoz, D. García de Viedma)

DOI: <https://doi.org/10.3201/eid2503.181593>

<sup>1</sup>These authors contributed equally to this article.

<sup>2</sup>These authors contributed equally to this article.

has affected host-country TB epidemiology, as reported in other settings (8–10).

The convergence of these circumstances made Panama ideal for evaluating an alternative simplified approach for optimized surveillance of TB transmission. We based our effort on a rational application of MIRU-VNTR, subsequent WGS analysis of clustered representatives, and final tailoring of allele-specific oligonucleotide PCR (ASO-PCR) for local prospective targeted surveillance of TB cases.

## Materials and Methods

### Study Sample

The retrospective study sample comprised 94 *M. tuberculosis* isolates from diagnostic specimens of patients with symptomatic respiratory disease (excluding those under treatment) from 19 health centers and hospitals in Panama and Colon provinces in 2015. This convenience sample comprised 80 isolates from Panama and 14 from Colon. These 2 provinces provided 660 TB cases in 2015; thus, the convenience sample represented 14.4% of TB cases from that year. Among the 91 isolates with susceptibility data, 84 (89.4%) were pansusceptible, 6 (6.4%) were monoresistant to isoniazid (2.1%) and rifampin (4.3%), and the remaining 1 (1.1%) was multidrug-resistant (MDR) (Table). We also included additional prospective isolates from Colon and Panama (January–August 2018), selected following the same selection criteria as the retrospective strain collection.

### Drug Susceptibility Testing

Drug susceptibility was determined at the National Tuberculosis Reference Laboratory in Gorgas Memorial Institute (Panama City, Panama) using GenoType MTB-DRplus (Hain Lifescience, <https://www.hain-lifescience.de/en/>). We confirmed drug susceptibility results by the proportion method on Löwenstein–Jensen medium against the first-line drugs isoniazid, rifampin, streptomycin, and ethambutol. Second-line resistance was not tested in this strain collection.

### DNA Extraction

DNA was extracted from Löwenstein–Jensen cultures using double-distilled water protocols from the GenoType Kit (Hain Lifescience) for first-line drug susceptibility testing. DNA for MIRU-VNTR and WGS was extracted using the QIAamp DNA Mini Kit (QIAGEN, <https://www.qiagen.com>) according to the manufacturer's protocol or the cetyl trimethylammonium bromide–based standard purification.

### MIRU-VNTR Analysis

We conducted genotyping using multiplex PCR based on 24-locus MIRU-VNTR (11). DNA extraction and MIRU-VNTR typing of the retrospective sample were performed

**Table.** Study population, drug resistance, and genotypes of *Mycobacterium tuberculosis*, Colon and Panama provinces, Panama, 2015

Characteristic	Result*
Province	
Panama, n = 80	80 (85)
Colon, n = 14	14 (15)
Antimicrobial drug susceptibility	
Susceptible	84 (89.4)
Monoresistant	6 (6.4)
Isoniazid	2 (2.1)
Rifampin	4 (4.3)
Multidrug-resistant	1 (1.1)
No data	3 (3.2)
Lineage/sublineage	
Lineage 1	3
East African Indian	3
Lineage 2	7
Beijing	7
Lineage 4	82
Latin-American-Mediterranean	31
Haarlem	28
H37Rv-like	18
Cameroon	3
X	2
Lineage 5	1
West_African_1	1
Lineage 6	1
West_African_2	1
Distribution of clustered isolates	
Cluster A (Beijing)	7 (15)
Cluster B (Haarlem)†	14 (30)
Cluster C Latin-American-Mediterranean†	9 (19)
Cluster D (H37Rv-like)	12 (26)
Cluster E Latin-American-Mediterranean†	3 (6)
Cluster F (H37Rv-like)	2 (4)

\*Values are no. (%) isolates or % isolates.

†One strain of this cluster was monoresistant to rifampin.

at Instituto de Investigaciones Científicas y Servicios de Alta Tecnología Laboratories (City of Knowledge, Panama). PCR products were sized using capillary electrophoresis in a 3130 Genetic Analyzer (Applied Biosystems, <https://www.thermofisher.com>). For the Beijing isolates, we added the 4 previously recommended additional hypervariable locus set (1982, 3232, 3820, and 4120) to the 24 loci MIRU-VNTR set (12). We assigned TB lineage and sublineages from MIRU-VNTR data using a lineage prediction tool (TBminer, <http://info-demo.lirimm.fr/TBminer>) (13).

### WGS

WGS was performed in Hospital Gregorio Marañón (Madrid, Spain) as detailed elsewhere (14) on 2–3 representative isolates of each cluster. We generated DNA libraries following the Nextera XT Illumina protocol (Nextera XT Library Prep kit [FC-131–1024]; Illumina, <https://www.illumina.com>). Libraries were run in a MiSeq device (Illumina) by applying a paired-end reading procedure. We mapped the reads for each strain by using the ancestral *M. tuberculosis* complex genome, which was identical to H37Rv in terms of structure

but with the ancestral alleles inferred by using a maximum-likelihood approach (15). SNP calls were made with SAMtools (<http://samtools.sourceforge.net/>) and VarScan (<http://varscan.sourceforge.net/>) (coverage of at least 20-fold, mean SNP mapping quality of 20) (14,16). From the variants detected, we kept only the homozygous calls (those present at a specific position in  $\geq 90\%$  of the reads). Moreover, to filter out potential false-positive SNPs attributable to mapping errors, we omitted the variants detected in repetitive regions, phages, and Pro-Glu (PE)- and Pro-Pro-Glu (PPE)-rich regions. In addition, we omitted SNPs close to insertion/deletions (10-bp window) and those in areas with an anomalous accumulation of variants ( $\geq 3$  SNPs in 10 bp). Alignments and SNP variants were visualized and checked in IGV version 2.3.59 (<http://software.broadinstitute.org/software/igv>). A cutoff of  $\leq 12$  SNPs was used to consider 2 isolates as clustered due to recent transmission, as defined in Walker et al. (6). We deposited the sequences obtained in EMBL-EBI (<http://www.ebi.ac.uk> [accession nos. PRJEB23681 and PRJEB29408] and <http://bioinfo.indicasat.org.pa>).

### Design of ASO-PCR

For MIRU-VNTR-defined clusters A, B, and C, we identified the strain-specific SNPs after comparing WGS data with a database containing 4,598 sequence genomes from strains circulating throughout the world (A. Chiner-Oms, unpub. data). We selected 3 strain-specific SNPs for cluster A and 4 for clusters B and C to be targeted by a multiplex ASO-PCR (Appendix Table 1, <https://wwwnc.cdc.gov/EID/article/25/3/18-1593-App1.pdf>). We designed these ASO-PCRs to ensure the presence of amplification patterns, regardless of the strain analyzed. For cluster A, we designed 1 selective primer (SNP2) to target the allele found in the cluster A strain, whereas the 2 remaining primers (SNP1 and SNP3) targeted the alternative alleles, as expected in any other non-A strain. For clusters B and C, 2 primers (SNP1/SNP4 and SNP3/SNP4, respectively) targeted the alleles found in strains B and C, whereas the remaining SNPs targeted the alternative alleles. The size of the amplicons was calculated to rule out overlap and led to different band patterns in each case (Appendix Table 1). We fixed the final PCR protocols (Appendix Tables 1, 2) after evaluating multiple experimental conditions until the expected patterns for the strains surveyed and a set of different control strains were confirmed.

The ASO-PCR design and optimization were conducted in Hospital General Universitario Gregorio Marañón (Madrid, Spain). The data were transferred to be applied at the National Tuberculosis Reference Laboratory in Gorgas Memorial Institute in Panama City.

## Results

### MIRU-VNTR-Based Cluster Analysis

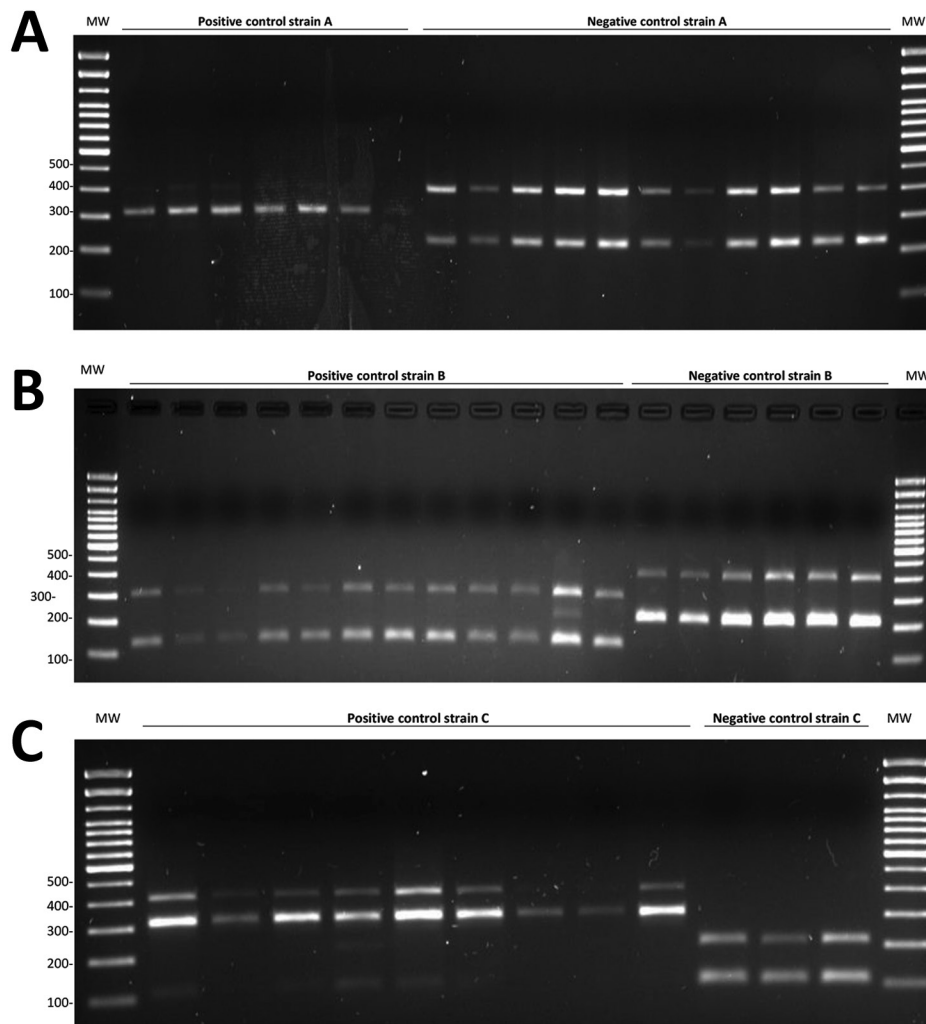
In MIRU-VNTR analysis, 47 (50.0%) of the *M. tuberculosis* retrospective isolates from Panama and Colon provinces grouped into 6 clusters, A–F. Four clusters (A–D) comprised 7–14 isolates each; the remaining 2 were smaller (E, 3 isolates, and F, 2 isolates) (Appendix Figure). Most (87%) isolates corresponded to lineage 4; 7.4% corresponded to lineage 2. Sublineages for the largest clusters corresponded to Beijing (cluster A) and Haarlem (cluster B), Latin-American-Mediterranean (cluster C), and H37Rv-like (cluster D) (Table). Clusters A and F consisted mostly of cases from Colon; the remaining 4 clusters consisted mostly of cases from Panama. Clusters A, D, and F corresponded exclusively to pansusceptible strains; clusters B, C, and E each included 1 isolate that was monoresistant to rifampin.

### WGS Analysis

Representative isolates (2 each from clusters A and B and 3 from cluster C) were available for a more in-depth analysis by WGS in Madrid, Spain. The sequences obtained offered a depth of coverage of  $20\times$ – $55\times$ , with a Phred score  $>20$ . The short pairwise distances (1–3 SNPs) for cluster C were consistent with recent transmission. However, 19 SNPs were found between the isolates in cluster A, and 20 SNPs were found between the isolates in cluster B, which was more consistent with the higher diversity expected for clusters involving long-term prevalent strains.

### ASO-PCRs for Targeted Surveillance of Selected Clusters

Once we identified the coexistence of 1 cluster due to active recent transmission with 2 clusters that were more likely to be prevalent, ASO-PCRs were designed in Madrid with 2 objectives: 1) prioritize the targeted surveillance of the actively transmitted cluster C; and 2) facilitate capture of new cases involved in clusters A and B with the aim of completing the phylogenetic reconstruction of long-term clusters. We designed 3 multiplex ASO-PCRs to target 3 SNPs for cluster A, 4 for cluster B, and 4 for cluster C (Appendix Tables 1, 2). We used all representatives for clusters A and C and 12 of 14 of those in cluster B as positive controls, and we detected the expected amplification patterns. Forty-nine Beijing isolates other than A strain, 46 Haarlem isolates other than B strain, and 36 Latin-American-Mediterranean isolates other than C strain were also used as negative controls of the strains targeted. In all cases, the expected patterns were obtained for the non-A–C control isolates (Figure 1).



**Figure 1.** PCR products for allelic-specific oligonucleotide PCRs for *Mycobacterium tuberculosis* cluster A, cluster B, and cluster C strains on a selection of representative strains A, B, and C and a selection of non-A, non-B, and non-C controls. Testing was for a simplified model to survey tuberculosis transmission using data from patients and controls in Panama and Colon provinces, Panama, 2015. A) 308-bp PCR product (single-nucleotide polymorphism [SNP] 2) for strain A and 400-bp and 228-bp products (SNP1/SNP3) for non-A strain. B) 127-bp and 297-bp PCR products (SNP1/SNP4) for strain B and 209-bp and 406-bp products (SNP2/SNP3) for non-B strain. C) 307-bp and 413-bp PCR products (SNP3/SNP4) for cluster C strain and 103-bp and 207-bp products (SNP1/SNP2) for non-C strain. MW, molecular weight (100-bp DNA ladder).

### Prospective Implementation of the ASO-PCR-Based Strategy

We used 51 prospective isolates from Colon and Panama collected during 2018 to evaluate the performance in situ of the 3 sets of cluster A–C–specific ASO-PCRs. The amplification patterns obtained in Madrid in the optimization step were now reproduced exactly in Panama. Cluster A PCR revealed 13 (25.5%) patients infected by Beijing strain A of the total sample, 11 from Colon, and 2 from Panama. Cluster B PCR identified 1 (1.9%) patient infected by strain B and 2 (3.9%) patients infected by strain C; all 3 of these patients were from Panama province.

All the isolates not labeled as corresponding to strains A, B, or C by the strain-specific PCRs were genotyped by VNTR. None showed a pattern corresponding to the strains targeted (they showed differences in >3 loci). This finding enabled us to rule out that we could be missing some strains related to the clusters targeted.

We used WGS to analyze all cases (except 2 for strain A) captured by the ASO-PCRs to determine the network of

relationships between the isolates (Figure 2). As expected, we observed lower SNP-based diversity acquired along a linear topology for cluster C, consistent with a sequential host-to-host recent-transmission nature for this cluster. In contrast, we found more SNPs, acquired along different branches, with several nonsampled nodes (median vectors, mv; Figure 2) for clusters A and B, more likely corresponding to prevalent strains that acquired higher diversity. In fact, for cluster A, this diversity was also detected by MI-RU-VNTR; 4 isolates captured by the ASO-PCR showed single-locus variations.

### Discussion

Molecular epidemiology strategies enable us to investigate TB transmission dynamics in a given population. However, to capture a true snapshot of TB transmission, these strategies require a universal, long-term fingerprinting scheme. In other words, it is necessary to ensure that transmission links are not missed because of incomplete sampling of all the TB cases in a complete population. Unfortunately, few

countries have implemented universal systematic genotyping. Many countries with more limited resources and a high TB prevalence can afford only fragmented molecular epidemiology studies based on convenience samples.

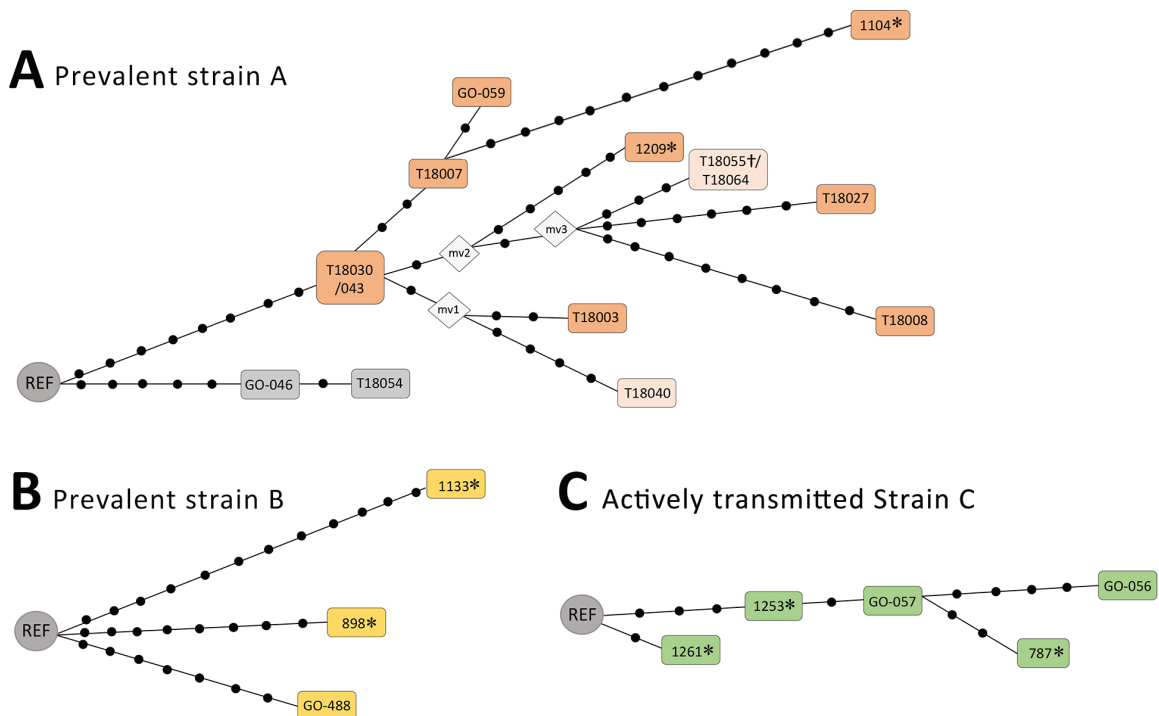
An additional concern is the recent process of substituting molecular epidemiology studies with more precise genomic epidemiology studies based on WGS analysis (4,6). If we failed to implement standard systematic molecular epidemiology in many high-prevalence areas, the gap between low- and high-income countries would widen in terms of genomic epidemiology. Therefore, we must define new strategies to offer an alternative to fill this gap.

In this study, we propose a model based on 4 stages: 1) updating the identification of strains responsible for the largest clusters in a population, 2) application of WGS to obtain more in-depth knowledge of clusters, 3) design of ASO-PCRs to identify specific strains, and 4) implementation of a simplified TB-control scheme based on targeted surveillance of predominant strains. To evaluate whether our approach could simplify detection of TB transmission in countries with no systematic molecular epidemiology programs, we applied it in Panama, a country with high TB-related mortality (8). No systematic universal genotyping is available in Panama, and the only data available are from studies of cases from a single clinic and of

MDR isolates (17,18). These studies were performed before the recent expansion of the Panama Canal, whose socioepidemiologic effect (19) can be considered similar to that of the construction of the interoceanic railroad and the Panama Canal in the mid-19th century, when large numbers of workers from Africa, India, and China arrived in Panama. High rates of malaria- and yellow fever-related deaths were associated with work on the Panama Canal (20). Regarding TB, the rates of latent infection were higher for migrant students than for US students living in the canal area (21).

The first step in our model revealed an unexpectedly high percentage of MIRU-VNTR-clustered cases. Half of the cases collected in Panama and Colon provinces grouped into 6 clusters. Fortunately, none of these clusters corresponded to resistant/MDR strains, despite the high clustering rates for MDR strains reported during 2002–2004 (17). Four of the 6 clusters identified in our study were responsible for nearly half of all cases analyzed.

Ours was a limited strain collection, and thus some transmission links might be missing because of the limited number of samples we were able to analyze, which covered only 1 year and 2 local settings, leading us to expect that the true figures for clustered cases are higher. In a previous study of a convenience pansusceptible *M. tuberculosis* strain collection



**Figure 2.** Networks of relationships based on whole-genome sequencing data of prevalent clusters (A, B) and active transmission cluster (C) in testing of a simplified model to survey tuberculosis transmission in Panama and Colon provinces, Panama, 2015. Each dot corresponds to a single-nucleotide polymorphism (SNP). In panel A, when 2 isolates are included in the same box, they showed no SNPs between them; the isolates within boxes with different colors show mycobacterial interspersed repetitive units–variable number of tandem repeats patterns with single-locus variations between them. mv, median vector corresponding to nonsampled nodes; REF, reference. \*Isolates from 2015 used to design the allele-specific oligonucleotide–PCRs; †strain identified in an additional analysis in Madrid (out of the prospective study applying the strain-specific PCRs in Panamá). A color version of this figure is available online (<http://wwwnc.cdc.gov/EID/article/25/3/18-1593-F2.htm>).

from 2005 (18), only 21% (13/62) were found to be clustered. Direct comparisons between the figures from that study and our study cannot be made because of marked differences in genotyping approaches applied. In any case, it seems that the clustering snapshot is now different, possibly because our sampling was restricted to a more precise geographic location.

Despite the short distance (74 km) between Panama and Colon provinces, which are connected by the best highway in Panama, clusters are rich in cases from one or the other province, thus alerting us to the epidemiologic singularities of each population, with some strains more restricted to specific geographic niches. This finding is especially relevant for the Beijing sublineage (associated with high transmissibility), which seems to be better represented in Colon province.

The second step in our model focused on applying WGS to the MIRU-VNTR clusters. This analysis revealed that 2 different phenomena were co-occurring in the MIRU-VNTR-defined large clusters A, B, and C, as follows: 1) recent transmission in cluster C, robustly determined by the low diversity identified by WGS; and 2) more SNPs in clusters A and B. The higher number of SNPs within clusters A and B (more than the 12 SNPs accepted as a threshold for inferring recent transmission from WGS data) suggests that these clusters involve prevalent strains that had accumulated higher diversity after circulating for longer periods. Since it was first determined (6), the SNP-based threshold for recent transmission has been proven correct in many studies (22–25). This threshold was robust even in circumstances that could lead to greater accumulation of diversity, such as prolonged host-to-host transmission and the simultaneous involvement of reactivation and recent transmission in the same event (16).

Robustly defined MIRU-VNTR clusters corresponded to different magnitudes of SNP-based diversity between the clustered isolates; that is, clusters caused by recent transmission coexisted with other clusters not associated exclusively with recent transmission but with prevalent strains that had circulated for long periods. This observation obliges us to reinterpret the load of true recent transmission in high-prevalence settings. According to our WGS findings, our initial estimation of 50.0% of cases from recent transmission, which was based on MIRU-VNTR analysis, must be reduced. This bias in the estimation of recent transmission in MIRU-VNTR-defined clusters has been discussed (26–28). We also demonstrated how MIRU-VNTR failed to differentiate between migrants who had acquired TB by recent transmission after arrival in Spain from other patients who had imported TB from their country of origin (28). Both shared identical MIRU-VNTR patterns, and only WGS succeeded in differentiating them based on the magnitude of SNP diversity (28). From our data, it can be deduced that WGS is necessary in high-prevalence settings to differentiate MIRU-VNTR-defined clusters that

genuinely result from recent transmission from those with the involvement of long-term prevalent strains.

Because the systematic application of WGS in low-resource settings is not realistic, our tailored ASO-PCR model for targeting marker SNPs for the strains previously identified to be responsible for a large percentage of TB cases remains an affordable strategy. Our proposal reconciles the high discriminatory power of WGS (targeted SNPs are obtained from the WGS analysis) with the low-cost and easy implementation of PCR-based tools. Based on the initial MIRU-VNTR analysis, application of these 3 multiplex ASO-PCRs would have enabled us to cover 32% of all retrospective TB cases in our sample. This coverage justified the prospective evaluation to assess their usefulness and precision, the last step in the implementation of our model. Because our multiplex ASO-PCRs were designed based on WGS data from a limited size convenience sample from Panama and Colon provinces, the PCRs should be used only for surveillance in these 2 major provinces.

The limited sample of isolates obtained in 2015, which were used to tailor the surveillance strain-specific PCRs, could make the PCRs useless when moving forward because the epidemiology of TB and the composition of circulating strains changes over time. However, the application in Panama of the set of 3 ASO-PCRs on the prospective samples from Panama City and Colon enabled us to determine the strain involved in 31.4% of the cases, a figure close to the 32% of the isolates belonging to the 3 surveyed strains in 2015. This finding suggested a rather stable composition of strains in this setting.

Initially, the quality of the DNA obtained from the isolates in 2015 for the fourth predominant cluster (cluster D) was not suitable for WGS. Just recently, the VNTR analysis from the 2018 isolates not belonging to the A/B/C clusters enabled us to identify new representative cases for this cluster, which will be analyzed by WGS to tailor a new PCR to complete and extend the coverage of the surveillance panel.

Once our strategy has proved to be useful, a country-wide genotyping effort will be necessary to determine highly transmitted strains nationwide and develop new ASO-PCR tools for surveillance of the entire country. The strategy will be evaluated periodically. It will mean updating the composition of circulating strains by MIRU-VNTR analysis as soon as we detect a reduction in the proportion of incident isolates that can be labeled by the strain-specific PCRs already implemented. From the VNTR analysis, we will select the new clusters to be targeted, according either to their magnitude or to the speed in which the involved strain is transmitted. Three representative isolates from each cluster selected will be analyzed by WGS to identify the strain-specific SNPs to tailor new PCRs to target them. Therefore, the panel of PCRs applied as a surveillance tool will be constantly updated according to the changing epidemiology of the setting.

The networks of relationships obtained from WGS data of the representatives of clusters A–C captured by the ASO-PCRs were consistent with the different nature (either prevalent strains A/B or recently transmitted strain C) of the clusters analyzed. A pattern of closely related isolates (below the 12-SNP cutoff determined as similarity threshold) (6) distributed along a simple linear topology was found for cluster C. However, more complex branched topologies and a higher diversity between the isolates were observed for the prevalent clusters A and B.

A high proportion (21.5%) of patients in Colon were infected by the Beijing strain (strain A). This finding constitutes a challenging alert for the national TB-control program because this lineage is involved in long-term outbreaks, which eventually become uncontrollable. In Gran Canaria Island, Spain, the importation of a Beijing strain in the 1990s caused a major outbreak that could not be controlled. Three years after importation, 27% of TB patients on the island were infected by this strain (29). Three decades later, the strain remains highly prevalent, not only on the island to which it was imported but also on neighboring islands (30). The marked presence for the Beijing lineage in Colon province seems to be an emergent issue because the lineage accounted for only 3.7% of the cases in 2005 (18). Those authors raised the alert about the effect of the expansion of the Panama Canal on the likely modification of the composition of circulating strains. Application of only 1 of our PCRs (the one targeting the Beijing strain) would have enabled us to characterize half of all current newly diagnosed TB cases in Colon; that is, persons infected by a highly transmissible strain. This PCR could constitute a key tool for optimizing the national TB-control programs to avoid future uncontrollable situations, such as the one reported for the Canary Islands.

In summary, our study evaluated the feasibility and usefulness of a new model for simplifying and optimizing surveillance of transmission of TB in countries with no systematic molecular or genomic epidemiology program. Preliminary MIRU-VNTR analysis alerted us to a high percentage (50%) of clustered TB cases and revealed a differential distribution of clustered cases in 2 provinces within Panama. Subsequent WGS analysis showed that some of the MIRU-VNTR clusters were the result of active recent transmission, whereas others more likely resulted from prevalent long-term strains. We applied the WGS data to tailor 3 ASO-PCRs, which covered most cases in the population. The prospective application of only 3 ASO-PCRs enabled us to identify the strains infecting 31.4% of incident cases. This strategy optimized tracking of actively transmitted *M. tuberculosis* strains and capture of new cases involving long-term prevalent strains, thus making it possible to reconstruct the global phylogeny of their clusters. Our strategy also made it possible to identify

the alarming presence of the Beijing sublineage in the incident cases in Colon province. A design based on a rational integration of MIRU-VNTR, WGS, and ASO-PCR could provide a new model for surveillance of TB transmission in countries without universal molecular/genomic epidemiology programs.

### Acknowledgments

We thank Thomas O'Boyle for proofreading the manuscript. We are grateful for the support of the healthcare workers who located TB patients to obtain *M. tuberculosis* isolates from Panama and Colon (Ana de Chavez, Laura Solis, Maritza Mayrena, Maybis Garay, William Teerán, Victoria Williams, Julio Jurado, Odemaris Luque, Isolina Martínez, Yaracelis Cuadra, and Luz Fruto). We thank the working group in the Clinical Microbiology section of the Gorgas Memorial Institute for its support in implementing the strain-specific PCRs in Panama.

This study received funding from: the Network of the European Union, Latin America and the Caribbean Countries on Joint Innovation and Research Activities ERANet-LAC (TRANS-TB-TRANS)(Refs AC16/00057; ELAC2015/T08-0664); Ministerio de Economía y Finanzas Panamá, Sistema Nacional de Investigadores de Panamá and Secretaría Nacional de Ciencia Tecnología e Innovación (grant no. 82-2015); ISCIII (grant no. AC16/00057, grant no.FIS15/01554 and Miguel Servet to L.P.-L.[contract no. CP15/00075]); IFARHU-SENACYT (grant no. 270-2016-293) to F.A.

### About the Author

Mr. Domínguez is responsible for the Laboratorio Central de Referencia en Salud Pública, Instituto Conmemorativo Gorgas de Estudios de la Salud, and is adjunct faculty at the Tuberculosis Biomarker Research Unit of Instituto de Investigaciones Científicas y Servicios de Alta Tecnología, City of Knowledge, Panama. His research focuses on diagnosis, identification, drug susceptibility testing, and genotyping of *Mycobacterium tuberculosis* complex, as well as on the development of strategies to improve TB transmission surveillance.

### References

1. Diel R, Rüsche-Gerdes S, Niemann S. Molecular epidemiology of tuberculosis among immigrants in Hamburg, Germany. *J Clin Microbiol.* 2004;42:2952–60. <http://dx.doi.org/10.1128/JCM.42.7.2952-2960.2004>
2. Allix-Béguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol.* 2008;46:1398–406. <http://dx.doi.org/10.1128/JCM.02089-07>
3. Lambregts-van Weezenbeek CS, Sebek MM, van Gerven PJ, de Vries G, Verver S, Kalisvaart NA, et al. Tuberculosis contact investigation and DNA fingerprint surveillance in The Netherlands: 6 years' experience with nation-wide cluster feedback and cluster monitoring. *Int J Tuberc Lung Dis.* 2003;7(Suppl 3):S463–70.

4. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, et al. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med*. 2013;10:e1001387. <http://dx.doi.org/10.1371/journal.pmed.1001387>
5. Homolka S, Projahn M, Feuerrriegel S, Ubben T, Diel R, Nübel U, et al. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One*. 2012;7:e39855. <http://dx.doi.org/10.1371/journal.pone.0039855>
6. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*. 2013;13:137–46. [http://dx.doi.org/10.1016/S1473-3099\(12\)70277-3](http://dx.doi.org/10.1016/S1473-3099(12)70277-3)
7. Plan Estratégico Nacional para el Control de la Tuberculosis en Panamá 2018–2022. Information at the Gaceta Oficial of Panama [cited 2018 Jun 1]. <https://www.gacetaoficial.gob.pa/pdfTemp/28549/67983.pdf>
8. Tarajia M, Goodridge A. Tuberculosis remains a challenge despite economic growth in Panama. *Int J Tuberc Lung Dis*. 2014;18:286–8. <http://dx.doi.org/10.5588/ijtld.13.0534>
9. Truzyan N, Crape B, Grigoryan R, Martirosyan H, Petrosyan V. Increased risk for multidrug-resistant tuberculosis in migratory workers, Armenia. *Emerg Infect Dis*. 2015;21:474–6. <http://dx.doi.org/10.3201/eid2103.140474>
10. Sandgren A, Schepisi MS, Sotgiu G, Huitric E, Migliori GB, Manissero D, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: a systematic review. *Eur Respir J*. 2014;43:1159–71. <http://dx.doi.org/10.1183/09031936.00117213>
11. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsche-Gerdes S, Willery E, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2006;44:4498–510. <http://dx.doi.org/10.1128/JCM.01392-06>
12. Allix-Béguec C, Wahl C, Hanekom M, Nikolayevskyy V, Drobniewski F, Maeda S, et al. Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of *Mycobacterium tuberculosis* Beijing isolates. *J Clin Microbiol*. 2014;52:164–72. <http://dx.doi.org/10.1128/JCM.02519-13>
13. Azé J, Sola C, Zhang J, Lafosse-Marin F, Yasmin M, Siddiqui R, et al. Genomics and machine learning for taxonomy consensus: the *Mycobacterium tuberculosis* complex paradigm. *PLoS One*. 2015;10:e0130912. <http://dx.doi.org/10.1371/journal.pone.0130912>
14. Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, et al. Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis*. 2014;209:98–108. <http://dx.doi.org/10.1093/infdis/jit439>
15. Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet*. 2010;42:498–503. <http://dx.doi.org/10.1038/ng.590>
16. Herranz M, Pole I, Ozere I, Chiner-Oms Á, Martínez-Lirola M, Pérez-García F, et al. *Mycobacterium tuberculosis* acquires limited genetic diversity in prolonged infections, reactivations and transmissions involving multiple hosts. *Front Microbiol*. 2018;8:2661. <http://dx.doi.org/10.3389/fmicb.2017.02661>
17. Rosas S, Bravo J, Gonzalez F, de Moreno N, Sanchez J, Gavilan RG, et al. High clustering rates of multidrug-resistant *Mycobacterium tuberculosis* genotypes in Panama. *BMC Infect Dis*. 2013;13:442. <http://dx.doi.org/10.1186/1471-2334-13-442>
18. Sambrano D, Correa R, Almengor P, Domínguez A, Vega S, Goodridge A. *Mycobacterium tuberculosis* isolates from single outpatient clinic in Panama City exhibit wide genetic diversity. *Am J Trop Med Hyg*. 2014;91:310–2. <http://dx.doi.org/10.4269/ajtmh.14-0134>
19. Hricko A. Progress and pollution: port cities prepare for the Panama Canal expansion. *Environ Health Perspect*. 2012;120:A470–3. <http://dx.doi.org/10.1289/ehp.120-a470>
20. Sanitation and the Panama Canal. *Hospital (Lond 1886)*. 1905;38:426.
21. Young RV, Pritham H. Prevalence of tuberculin sensitivity in the Panama Canal Zone. *Dis Chest*. 1963;43:172–5. <http://dx.doi.org/10.1378/chest.43.2.172>
22. Witney AA, Bateson AL, Jindani A, Phillips PP, Coleman D, Stoker NG, et al.; RIFAQUIN Study Team. Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Med*. 2017;15:71. <http://dx.doi.org/10.1186/s12916-017-0834-4>
23. Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. *PLoS Med*. 2016;13:e1002137. <http://dx.doi.org/10.1371/journal.pmed.1002137>
24. Tyler AD, Randell E, Baikie M, Antonation K, Janella D, Christianson S, et al. Application of whole genome sequence analysis to the study of *Mycobacterium tuberculosis* in Nunavut, Canada. *PLoS One*. 2017;12:e0185656. <http://dx.doi.org/10.1371/journal.pone.0185656>
25. Lalor MK, Casali N, Walker TM, Anderson LF, Davidson JA, Ratna N, et al. The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak. *Eur Respir J*. 2018;51:1702313. <http://dx.doi.org/10.1183/13993003.02313-2017>
26. Stucki D, Ballif M, Egger M, Furrer H, Altpeter E, Battegay M, et al. Standard genotyping overestimates transmission of *Mycobacterium tuberculosis* among immigrants in a low-incidence country. *J Clin Microbiol*. 2016;54:1862–70. <http://dx.doi.org/10.1128/JCM.00126-16>
27. Bjorn-Mortensen K, Soborg B, Koch A, Ladefoged K, Merker M, Lillebaek T, et al. Tracing *Mycobacterium tuberculosis* transmission by whole genome sequencing in a high incidence setting: a retrospective population-based study in East Greenland. *Sci Rep*. 2016;6:33180. <http://dx.doi.org/10.1038/srep33180>
28. Abascal E, Pérez-Lago L, Martínez-Lirola M, Chiner-Oms Á, Herranz M, Chaoui I, et al. Whole genome sequencing-based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. *Euro Surveill*. 2019;24:1800005. <http://dx.doi.org/10.2807/1560-7917.ES.2019.24.4.1800005>
29. Caminero JA, Pena MJ, Campos-Herrero MI, Rodríguez JC, García I, Cabrera P, et al. Epidemiological evidence of the spread of a *Mycobacterium tuberculosis* strain of the Beijing genotype on Gran Canaria Island. *Am J Respir Crit Care Med*. 2001;164:1165–70. <http://dx.doi.org/10.1164/ajrccm.164.7.2101031>
30. Pérez-Lago L, Campos-Herrera MI, Cañas F, Copado R, Sante L, Pino B, et al. A *Mycobacterium tuberculosis* Beijing strain persists at high rates and extends its geographic boundaries 20 years after importation. *Sci Rep*. In press 2019.

---

Address for correspondence: Dario Garcia de Viedma, Hospital Gregorio Marañon, Servicio de Microbiología, C/Dr. Esquerdo 46 (28007), Madrid, Spain; email: dgvedma2@gmail.com; Amador Goodridge, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología, Tuberculosis Biomarker Research Unit at Centro de Biología Molecular y Celular de Enfermedades, Bldg 208, City of Knowledge, Panama City 0843-01103, Panama; email: agoodridge@indicat.org.pa



# Simplified Model to Survey Tuberculosis Transmission in Countries Without Systematic Molecular Epidemiology Programs

## Appendix

**Appendix Table 1.** Specific SNP alleles targeted and PCR design for each *Mycobacterium tuberculosis* strain

Cluster	SNP per cluster	Position	SNP	Gen	Primer sequence (5' to 3')	Pool concentration, $\mu$ M	Target strain	PCR fragment size, bp
Cluster A, Beijing	SNP1	204386	G>A	<i>Rv0173</i>	Fw-Rv0173: GCTACCGAAAAATGCCACCG Rv-Rv0173: GAGCACCCGATTGAGGACT	2 2	Non-A strain	400
	SNP2	518748	G>A	<i>Rv0430</i>	Fw-Rv0430: CGATATGGACAGCGCCATGA Rv-Rv0430: TCACCTCGAAGCCAAGGC	6 6	Strain A	308
	SNP3	3479530	G>A	<i>Rv3111</i>	Fw-Rv3111: TGGAAATGGAAGCGTTGACC Rv-Rv3111: ACGAACGGCACCGAAGTAA	5 5	Non-A strain	228
Cluster B, Haarlem	SNP1	261991	G>A	<i>Rv0218</i>	Fw-Rv0218: CACCTGGAGCGGCGTA Rv-Rv0218: GCGTTGCAATGTCGAACATT	2 2	Strain B	297
	SNP2	825060	C>T	<i>Rv0732</i>	Fw-Rv0732: TGCCCTACATCACCGCC Rv-Rv0732: GATGTCCAGCGAGCAACC	2 2	Non-B strain	209
	SNP3	2425513	G>T	<i>RV2163c</i>	Fw-RV2163: CATCGAGCACGGGTTGAG Rv-RV2163: GGCCTTGATAATGCGTGGG	2 2	Non-B strain	406
	SNP4	568395	G>T	<i>Rv0479c</i>	Fw-Rv0479: GCATCCTTCGGTGTGGC Rv-Rv0479: GGATCGTGAGCTTTATCTGCATT	2 2	Strain B	127
Cluster C, LAM	SNP1	516278	G>T	<i>Rv0427c</i>	Fw-Rv0427: CGCACGTTTCCGAACCC Rv-Rv0427: CGATCTGGGCGTCCATCA	2 2	Non-C strain	207
	SNP2	781859	T>C	<i>Rv0682</i>	Fw-Rv0682: ACAACCTGCAGGAGCACTC Rv-Rv0682: CCCTGCGTATCCAGCGAA	6 6	Non-C strain	103
	SNP3	1802486	G>T	<i>Rv1601</i>	Fw-Rv1601: ACGCCAGCTTCGACCTC Rv-Rv1601: AACGATTGGAACACGTGCA	2 2	Strain C	307
	SNP4	4403893	T>C	<i>Rv3915</i>	Fw-Rv3915: AAGGACGGATGGCAGCC Rv-Rv3915: GAGAATGCCTTCGGCGATG	6 6	Strain C	413

\*LAM, Latin American-Mediterranean; SNP, single-nucleotide polymorphism.

**Appendix Table 2.** Reaction mixes for each PCR\*

Reagent	Volume per reaction, $\mu\text{L}$		
	Cluster A	Cluster B	Cluster C
HotStarTaq polymerase, 5 U/ $\mu\text{L}$ diluted 1:10	1.25	1.25	1.25
MgCl <sub>2</sub> , 25 mM	0.4	0.3	0.4
PCR buffer, 10 $\times$ with 15 mM MgCl <sub>2</sub>	2.5	2.5	2.5
Q solution, 5 $\times$	5.0	5.0	5.0
Primer pool, 10 $\times$ †	2.5	2.5	2.5
dNTPs, 25 mM	0.2	0.2	0.2
H <sub>2</sub> O	10.65	10.75	10.7
DNA target		2.5 $\mu\text{L}$ for each one	
Final reaction volume of 25 $\mu\text{L}$			

\*dNTPs, deoxynucleotides; MgCl<sub>2</sub>, magnesium chloride.

†Primer pool concentration described in Appendix Table 1. The PCR conditions for each strain was as follows: 95° C for 15 min, followed by 27 cycles of 95° C for 1 min, annealing (6° C for Custer A and 60° C both for Custer B and C for 1 min) and 72° C for 10 min.

