

---

# Discriminatory Ability of Hypervariable Variable Number Tandem Repeat Loci in Population-based Analysis of *Mycobacterium tuberculosis* Strains, London, UK

Preya Velji,<sup>1</sup> Vladyslav Nikolayevskyy,<sup>1</sup> Timothy Brown, and Francis Drobniowski

To address conflicting results about the stability of variable number tandem repeat (VNTR) loci and their value in prospective molecular epidemiology of *Mycobacterium tuberculosis*, we conducted a large prospective population-based analysis of all *M. tuberculosis* strains in a metropolitan setting. Optimal and reproducible conditions for reliable PCR and fragment analysis, comprising enzymes, denaturing conditions, and capillary temperature, were identified for a panel of hypervariable loci, including 3232, 2163a, 1982, and 4052. A total of 2,261 individual *M. tuberculosis* isolates and 265 sets of serial isolates were analyzed by using a standardized 15-loci VNTR panel, then an optimized hypervariable loci panel. The discriminative ability of loci varied substantially; locus VNTR 3232 varied the most, with 19 allelic variants and Hunter-Gaston index value of 0.909. Hypervariable loci should be included in standardized panels because they can provide consistent comparable results at multiple settings, provided the proposed conditions are adhered to.

Globally, tuberculosis (TB) accounts for almost 2 million deaths each year (1). Although TB notification rates in the United Kingdom (13.8/100,000 in 2007) remain low, rates differ substantially by region: London (43.2/100,000) accounts for ≈40% of all TB cases registered in the United Kingdom, and ≈75% of TB patients in London were born abroad (2). Rates of drug resistance also are higher in London than in the rest of the United Kingdom: 8.6% of isolates are isoniazid resistant, and 1.2% are

multidrug resistant (UK Health Protection Agency; www.hpa.org.uk).

In settings where incidence of TB is low or moderate, molecular genotyping is used to investigate suspected TB outbreaks, laboratory cross-contamination, and reactivation and (at a population level) to identify clustered cases that are not apparently linked; for the latter purpose, the highest possible level of discrimination is required (3). For these purposes, insertion sequence (IS) 6110 restriction fragment length polymorphism (RFLP) analysis—often supplemented with spoligotyping and, more recently, with variable number tandem repeat (VNTR) typing—is used routinely.

The highest levels of epidemiologic discrimination of strains of the *Mycobacterium tuberculosis* complex (MTBC) can be achieved by using multilocus VNTR typing, but these results depend on the number and loci used, particularly for homogenous strain groups such as the Beijing family (3–5). This approach overcomes technical difficulties associated with IS6110-RFLP and is amenable to automation that results in a high throughput (6–10). A standardized panel of 15 + 9 VNTR loci (24 loci) has been proposed (7,11), but it is unclear whether sufficient discrimination would be seen when the panel is used in populations with a substantial prevalence of homogenous MTBC families (4,5,12). In addition, the discriminative power of VNTR loci may vary markedly among genetic families (7,13). Recent studies evaluating the discriminative power of VNTR typing have produced conflicting results that were generated by using convenience samples (small populations with low diversity or populations confined to a single geographic setting). These studies highlighted a need for larger population-based studies to identify discrimina-

---

Author affiliations: Barts and The London School of Medicine, Queen Mary, University of London, London, UK (P. Velji, V. Nikolayevskyy); and Health Protection Agency National Mycobacterium Reference Laboratory, London (T. Brown, F. Drobniowski)

DOI: 10.3201/eid1510.090463

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

tive VNTR loci and ascertain their applicability for various genetic groups.

Concerns about the stability and reproducibility of particularly useful hypervariable loci, such as 3232, 2163a, 3336, and 1982 (3–5,14), have been raised (7,15). As a result, they have been excluded from the proposed international panels for VNTR typing. For these reasons, we conducted a study to examine the stability of hypervariable loci and the parameters associated with reproducibility, to select loci suitable for prospective molecular epidemiologic studies, and to evaluate the discriminatory power of these loci at a population level in a metropolitan setting.

## Materials and Methods

### Bacterial Isolates

A total of 2,261 individual MTBC isolates (1 per patient) were included in this prospectively designed population study. These isolates represented 95.7% of the bacteriologically confirmed TB cases reported from the 30 London hospitals in the 12 months from April 2005 through March 2006. These isolates had been characterized by using spoligotyping, and all but 4 were assigned to 1 of 36 spoligotype families (16,17). Multiple isolates were available from 265 patients (11.7%), resulting in serial isolate sets of 2–6 isolates, which had been sampled at intervals of 3 days to 11 months (N = 632).

### Multilocus VNTR Analysis

All extracts were typed by using 15 mycobacterial interspersed repetitive unit (MIRU)-VNTR loci as previously described (3). Isolates clustered when the 15 MIRU-Hunter-Gaston index value VNTR profiles we used were reanalyzed with an additional panel of VNTR loci 2163b, 2347, 3232, 2163a, 1982, 3336, and 4052 as previously described (3,5) after optimization of factors affecting reproducibility (see Hypervariable Loci Optimization). Variability or discrimination at a locus was assessed by using the Hunter-Gaston Discriminative Index (HGDI) (18). Loci with HGDI values <0.3, 0.3–0.6, and >0.6 were considered poorly, moderately, and highly discriminative, respectively (19).

### Hypervariable Loci Optimization

We selected 16 previously characterized MTBC isolates to cover the complete range of repeat sizes at control loci MIRU 26 and exact tandem repeat (ETR)-B and experimental hypervariable loci VNTRs 1982 and 3232 (except 0 repeats for the locus 3232). For each of the 16 extracts, four 10- $\mu$ L PCRs were conducted for each of the primer mixes in duplicate. Of these 4 reactions, the first was performed as described previously with BIOTAQ polymerase (Bioline, London, UK) (any enzyme in the given context means enzyme in conjunction with the buffer recommended and

supplied by a manufacturer). Three other sets of PCRs were conducted under different amplification conditions (1).

#### Method 1

Diamond DNA polymerase (Bioline) was used (9). The PCR amplification cycle was 3 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 60°C, and 2 min at 72°C, and 1 final cycle of 5 min at 72°C (2).

#### Method 2

HotStartTaq DNA polymerase (QIAGEN, Hilden Germany) was used. Each 10- $\mu$ L reaction contained 1 $\times$  PCR buffer (QIAGEN), 0.25 U/ $\mu$ L of the relevant polymerase, 0.2  $\mu$ mol/L dNTPs, 0.125  $\mu$ mol/L of relevant primer, and 5% dimethylsulfoxide. The DNA amplification cycle was 15 min at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, and 1 min at 72°C, and a final cycle of 10 min at 72°C (3).

#### Method 3

HotStartTaq Plus DNA polymerase (QIAGEN) was used. The PCR mixture was the same as in method 2, and the amplification cycle was the same, except that the initial 95°C activation time was reduced to 5 min.

We manually calculated the number of repeats within each PCR product by resolving 4  $\mu$ L of each product on a 1.2% (wt/vol) agarose gel (Agarose LE Analytical grade; Promega, Southampton, UK) against a 2,000-bp HyperLadder II standard (Bioline). The number of repeats at each locus also was calculated by sizing in a denaturing capillary electrophoresis system using a CEQ 8000 instrument with a DNA Size Standard 600 (Beckman Coulter, High Wycombe, UK) and MapMarker DI labeled 640–1000 (BioVentures, Inc., Murfreesboro, TN, USA) because fragments were expected to be >600 bp. Three parameter sets (Table 1) were used to analyze all fragments. The different parameters examined were capillary temperature (60°C for methods 1 and 2 and 50°C for method 3, respectively), denaturation time (120 s for method 1 and 180 s for methods 2 and 3, respectively) and separation time (60 min for methods 1 and 2 and 70 min for method 3, respectively). Fragment data traces were automatically analyzed by using the scheme shown in Table 1. For locus 3232, we accounted for offset values (i.e., difference among actual sizes of PCR fragments and apparent sizes indicated by electrophoresis) when calculating number of repeats in Table 1.

### Assessing Stability and Reproducibility of VNTR Loci

All isolates were grouped into 265 sets of serial isolates (2–6 isolates each) and typed at all 22 loci. Primer sequences for all loci were as described previously (3,9,20,21). PCR was set up by using BIOTAQ polymerase for amplifying 12 MIRU and 3 ETR loci and Diamond polymerase for the

Table 1. Expected molecular weights of *Mycobacterium tuberculosis* of fragments at each locus, with different numbers of copies, London, UK, 2005–2006\*

No. repeats	Length of expected fragments for each locus, bp			
	MIRU 26	ETR-B	VNTR 1982	VNTR 3232†
0	244	121	178	
1	295	174	256	242
2	344	227	334	286
3	393	280	412	330
4	442	333	490	372
5	491	386	568	415
6	540	439	646	458
7	589	492	727	501
8	638	545	802	546
9	687	598	880	587
10	736	651	958	630
11	785		1,038	673
12	834		1,116	716
13	883		1,194	759
14	932			802
15				845
16				888
17				931
18				974
19				1,017
20				1,060

\*MIRU, mycobacterial interspersed repetitive unit; ETR, exact tandem repeat; VNTR, variable number tandem repeats.

†No isolates had 0 repeats in locus 3232 in our population.

additional 7 VNTR loci. Capillary electrophoresis was performed by using the parameters described in method 1.

## Results

### Optimization of Hypervariable Loci

We evaluated factors that potentially affect the reproducibility of hypervariable VNTR loci by using various PCR and capillary and manual electrophoresis separation

conditions as described in the Materials and Methods. The ability to correctly amplify different VNTR loci depended on the enzyme used (Table 2); all polymerases efficiently amplified MIRU 26 and ETR-B, as indicated by the presence of PCR fragments on agarose gels and capillary electrophoresis peaks. However, locus VNTR 3232 was amplified effectively only with Bioline Diamond (15/16 strains, 93.8%). Although all polymerases except Bioline BIOTAQ were able to amplify DNA at locus VNTR 1982, longer fragments were amplified more efficiently by QIAGEN and Bioline Diamond polymerases. Therefore, Diamond polymerase was selected for the amplification of additional VNTR loci.

We assessed 3 methods for capillary electrophoresis. For each locus, apparent fragment sizes were plotted against expected fragment sizes for each method (Figure 1).

MIRU 26 fragments sizes were as expected for all allelic variants (except for the variant with 2 repeats) when BIOTAQ and Diamond polymerases were used, but sizes were larger than expected with QIAGEN polymerases. The smaller ETR-B fragments with 1 and 2 repeats all gave expected sizes with methods 1 and 2 but were less than expected with method 3 (where the capillary temperature was decreased). These results did not affect overall interpretation. For the higher number of repeats (4–6 repeats), all polymerases generated fragments that, when analyzed by using method 3, gave apparent sizes lower than expected. In some cases, this result affected the interpretation. The apparent sizes of VNTR 1982 fragments were all similar to the expected values independent of the polymerase used and the method used for capillary electrophoresis.

### Serial Isolates

Amplification was performed by using BIOTAQ polymerase for 12 MIRU and 3 ETR loci and Diamond

Table 2. Number of DNA extracts (from  $n = 16$ ) for which peaks were detected by different conditions for capillary electrophoresis of *Mycobacterium tuberculosis* after amplifying the loci with different polymerases, London, UK, 2005–2006\*

Locus	Method†	Bioline polymerases‡		QIAGEN polymerases‡	
		BIOTAQ	Diamond	HotStartTaq	HotStartTaq Plus
MIRU 26	1	16	16	16 (1)	16 (1)
	2	15	16	16 (1)	16 (1)
	3	16	16	16	16
ETR-B	1	16	16	16 (1)	16 (1)
	2	15	16	16 (1)	16 (1)
	3	16	16	16 (2)	16 (2)
VNTR 1982	1	8	13	14	14
	2	9	13	12	14
	3	6	11	12	14
VNTR 3232	1	11	15	13	14
	2	10	15	14	14
	3	11 (3)	15 (7)	13 (6)	13 (4)

\*MIRU, mycobacterial interspersed repetitive unit; ETR, exact tandem repeat; VNTR, variable number tandem repeats.

†Refer to Table 1.

‡Numbers in parentheses represent number of extracts whose calculated number of repeats were higher and lower than the expected value on the basis of that produced by the standard procedure (method 1).

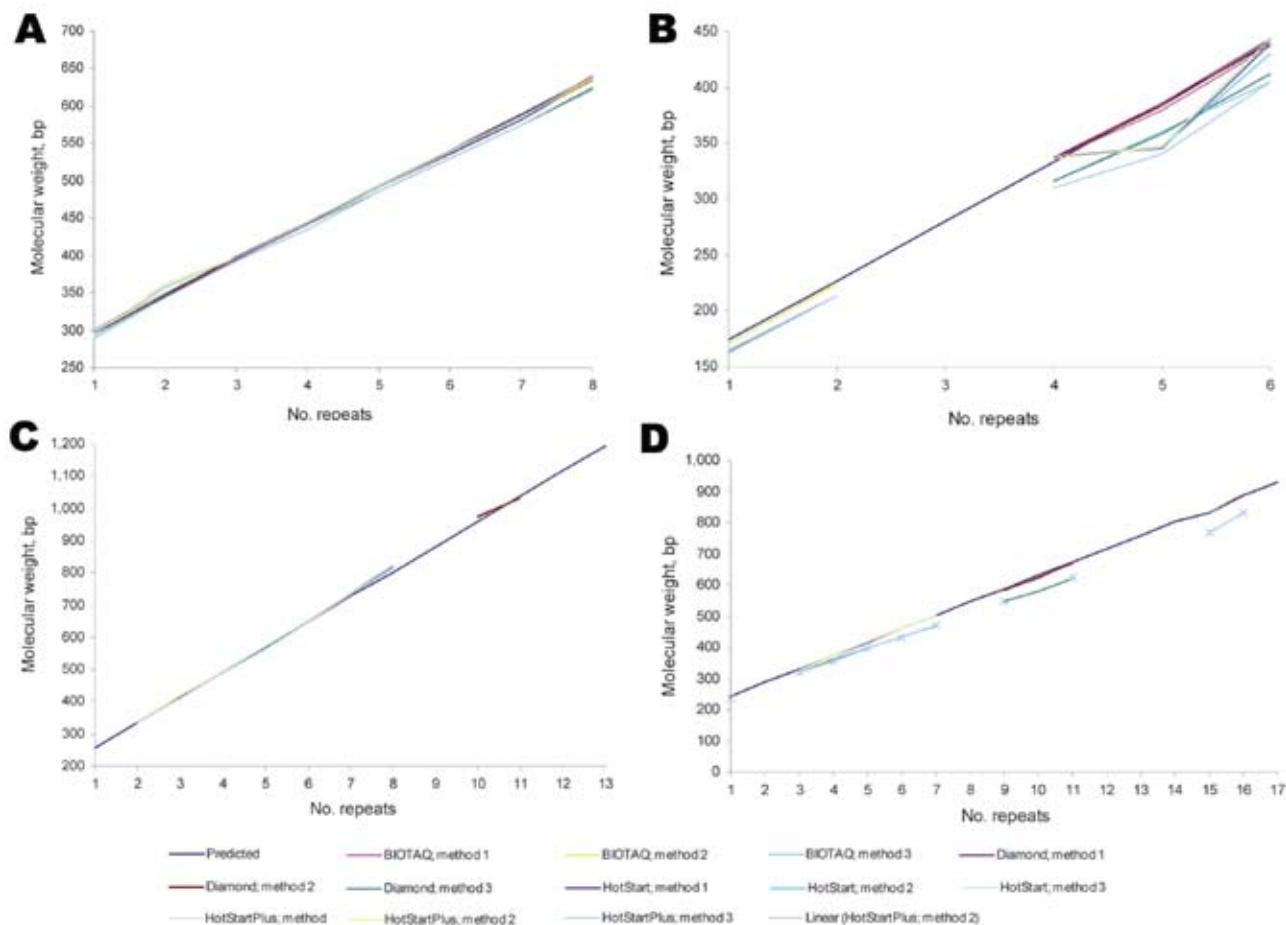


Figure 1. Effect of various enzymes and separation conditions on amplification and detectable molecular weights of PCR fragments for 4 variable number tandem repeat (VNTR) loci. A) Mycobacterial interspersed repetitive unit locus 26; B) locus exact tandem repeat; C) locus 1982; D) locus 3232.

polymerase for 7 VNTR loci with the optimized parameters in method 1. Analysis was blinded. No disagreements occurred in the interpretation of VNTR repeat numbers among isolates in a set. In a proportion of isolates ( $N = 124$ ), genotyping results were validated by using both capillary electrophoresis and manual electrophoresis for PCR fragment separation, and again, no discrepancies were found between VNTR loci copy numbers in strains isolated from the same patient at different time points (Figure 2).

#### Population Genotyping in Metropolitan Setting with 2 Panels of VNTR Loci

A total of 2,261 MTBC isolates circulating in London with known spoligotypes were genotyped by using a defined set of 15 loci (12 MIRU and 3 ETR); all known spoligotyping families were represented in the test population (online Technical Appendix, available from [www.cdc.gov/EID/content/15/10/1609-Techapp.pdf](http://www.cdc.gov/EID/content/15/10/1609-Techapp.pdf)). Complete 15-loci profiles were obtained for 2,046 strains (90.5% of all strains). Data for the remaining profiles were incomplete for  $\geq 1$  locus.

Overall PCR failure rate was 1.6%, with the highest number of failures ( $n = 72$ ) at locus ETR-A and the lowest number of failures ( $n = 4$ ) at locus ETR-C. When PCR failed, DNA was reextracted from original cultures, and genotyping was attempted again. If the second attempt was unsuccessful, the results for the locus were marked as missing.

Genotyping of MTBC isolates by using 15 MIRU-ETR loci yielded 1,036 unique profiles and 235 clusters containing 2–53 isolates (Table 3). Clustered profiles were shared by 1,225 isolates, giving a clustering rate of 54.2%.

Subsequently, 1,196 (97.6%) of 1,225 isolates (15 MIRU-ETR clustered isolates) were subjected to secondary typing by using VNTR loci 2163b, 2347, 3232, 2163a, 1982, 3336, and 4052. Resolution improved because strains that had been clustered initially were subdivided into new groups: 1,730 isolates now had unique genotyping patterns, and the remaining 502 isolates were grouped into 158 clusters, giving a new, substantially lower, clustering rate of 22.2% (Table 3).

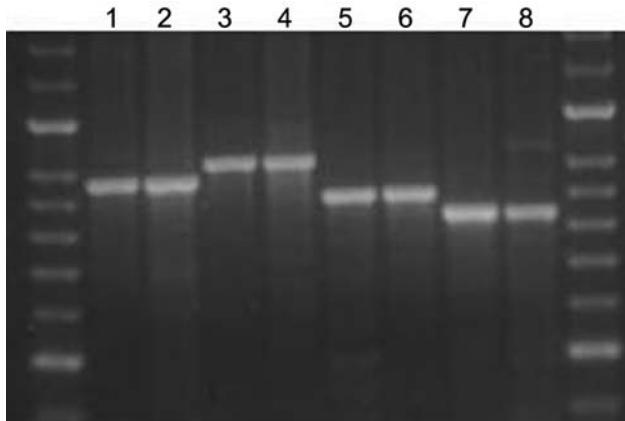


Figure 2. Agarose gel showing the stability of amplified fragments of variable number tandem repeat (VNTR) 3336 from 2 serial isolates isolated from 4 patients. Lane 1, patient A, isolate 1, isolated 2005 Jun 20, 8 copies; lane 2, patient A, isolate 2, isolated 2005 Jul 11, 8 copies; lane 3, patient B, isolate 1, isolated 2005 Jul 8, 9 copies; lane 4, patient B, isolate 2, isolated 2005 Aug 8, 9 copies; lane 5, patient C, isolate 1, isolated 2005 Nov 11, 7 copies; lane 6, patient C, isolate 2, isolated 2005 Nov 15, 7 copies; lane 7, patient D, isolate 1, isolated 2005 May 16, 6 copies; lane 8, patient D, isolate 2, isolated 2005 May 25, 6 copies.

#### Variability and Discriminative Power of VNTR Loci

The discriminative ability of VNTR loci varied markedly among the 22 VNTR loci and among spoligotyping families (online Technical Appendix) with locus VNTR 3232 showing the greatest variation (HGDI = 0.909 and 19 allelic variants) and loci MIRU 2 and 20, the least (HGDI = 0.134 and 0.196; number of allelic variants 4 and 3, respectively). Twelve loci each had  $\geq 10$  allelic variants. MIRU 4 showed moderate discriminative power, and MIRU 10, MIRU 16, MIRU 23, MIRU 26, MIRU 40, ETR-A, ETR-C, and VNTR 2163B, 2163A, 1982, 3232, 3336, and 4052 showed high discriminative power with HGDI values varying from 0.524 to 0.909. None of the 22 loci were monomorphic in the current study. With the exception of VNTR 2347, all loci included in the additional VNTR panel displayed higher variability than the primary panel of 15 MIRU-ETR loci used for UK national typing, which indicates their potential for increasing the power of prospective molecular genotyping.

The discriminative power of VNTR loci also varied among spoligotype families. The mean 15 MIRU-ETR

HGDI value for the Beijing family was low (0.163), which indicates that this family is relatively homogeneous, even within the diverse London population settings. Notably, mean 15 MIRU-ETR HGDI values for genetic families within the Euro-American lineage (T, Haarlem, S, X, Latin American–Mediterranean) were generally higher (0.307–0.378) than those for Beijing and Central Asian (CAS) (0.235). Within spoligotype families, the additional 7 VNTR increased variability in all cases, except for *M. bovis*. The highest HGDI were seen in the Latin American–Mediterranean family with locus 2163B; in Beijing, Haarlem, and *M. africanum* with VNTR 3232; in East African–Indian with VNTR 2163A; in X with VNTR 1982; in T with VNTR 3336; and in CAS with VNTR 4052. Within the East African–Indian family, the hypervariable loci VNTR 3232 varied little, with 93.7% isolates having a single copy. A small proportion of strains (Table 4) analyzed by using more discriminative loci, including VNTR 3232, 1982, 2163A, and 3336, generated PCR products that were too large for automated analysis but were resolved manually.

#### Discussion

Polymorphisms in rapidly evolving repetitive sequences, such as minisatellite VNTR, are a valuable tool for prospective epidemiologic analyses and provide a high degree of discrimination in situations in which few a priori epidemiologic data are available. In this population-based study, we genotyped 2,261 individual MTBC isolates obtained from patients residing in London by using 22 VNTR-MIRU loci.

Conflicting views on the use of hypervariable loci for typing have been reported, even when loci such as VNTR 3232 have been shown to have high discriminatory power (3,5,14). Some studies have demonstrated difficulty in amplification of multiple alleles, absence of PCR amplification products, varying data interpretation, and lack of reproducibility among laboratories (7). Similar problems were found with another potentially valuable hypervariable locus, VNTR 1982 (5,7). Therefore, we believed that by identifying the conditions that provided good, reproducible discrimination, we would be able to define the optimal conditions that would enable molecular epidemiologists to use VNTR 1982 and 3232. We addressed variability and reproducibility for these 2 loci using MIRU 26 and ETR-B as

Table 3. Discriminatory power of VNTR typing used in the study in establishing true minimum cluster size as marker of real transmission rate\*

Genotyping method	No. distinct profiles (variety of types)	No. clusters	Size of clusters, no. isolates	Clustering rate, % (n/N)	Recent transmission rate, % ((n - c)/N)	No. unique isolates
MIRU15 (n = 2261)	1,271	235	2–53	54.2	44.0	1,036
MIRU15 + Spoligotyping	1,619	196	2–48	37.1	29.0	1,423
MIRU15 + VNTR7	1,888	158	2–35	22.2	17.0	1,730

\*MIRU, mycobacterial interspersed repetitive unit; VNTR, variable number tandem repeats; n, no. clustered cases; N, total no. of strains; c, no. of clusters.

Table 4. Allelic variants of additional hypervariable VNTR loci that cannot be resolved with the CEQ automated sequencer\*†

Locus	Maximum no. repeats suitable for automated analysis	Fragment size, bp	Proportion of strains with allelic variants beyond the automated system resolution, %
3232	15	830	4.1
1982	9	880	11.9
2163A	11	876	10.8
3336	11	875	21.1

\*Beckman Coulter, Fullerton, CA, USA.

†VNTR, variable number tandem repeats.

controls that give stable comparable results in both agarose gel and capillary electrophoresis and have been used previously in a multilaboratory comparative study (7).

In all cases, identical data were produced for MIRU 26 and ETR-B irrespective of the DNA polymerase used. Amplification of VNTR 1982 and 3232 varied with different DNA polymerases, particularly when expected fragments were long.

The differing performances of polymerases for amplifying different loci can be explained by their varying properties. BIOTAQ polymerase is a basic Taq that can be used for a wide range of templates, whereas Diamond polymerase has been modified by a point mutation at the active site of the enzyme, enabling it to read through regions of secondary structure, microsatellites, and guanine cytosine-rich templates, such as those found in the *M. tuberculosis* genome. The QIAGEN polymerases are chemically modified polymerases with a high specificity similar to that of Diamond polymerase; thus they showed similar capabilities in amplifying VNTR 1982 and 3232. In addition, the buffer used with the QIAGEN polymerases is designed to increase the specificity of primer binding, making these polymerases suitable for dealing with complex genomic DNA.

Conditions that affect the denaturation of PCR products, and therefore their linearity before fragment sizing by electrophoresis, would be expected to influence apparent sizes of PCR fragments and copy number enumeration. We investigated the influence of DNA denaturation time and capillary separation temperature. As expected, we found that lowering the separation rate increased the discrimination of fragments >1,000 bp.

A marked difference was observed when the capillary temperature was decreased (method 3), which was independent of the polymerase used and locus investigated and demonstrated that separation conditions are critical for the correct interpretation of the VNTR typing results. In method 3, apparent fragment sizes were smaller and offset values were markedly larger, to the point that in some cases the calculated copy number was different from that expected.

Taking all the data together, we used BIOTAQ for amplifying MIRU and ETR loci, and Diamond polymerase for amplifying the extra 7 hypervariable VNTR loci, using the separation conditions detailed in method 1. We also demonstrated the reproducibility and stability of the extra

7 VNTR loci by comparing 22 MIRU-VNTR profiles from serial isolates. The resulting profiles of serial isolates from the same patients were identical, indicating that the conditions used for fragment amplification, detection, and analysis were ideal for typing of these loci and that these loci could be used for routine genotyping.

Clustering rates seen by using 15 MIRU-ETR loci far exceeded those previously reported when *IS6110* RFLP was used in a London population study (22,23). We concluded that 15-MIRU-ETR genotyping was insufficiently discriminative and was producing so-called false clustering. This view was supported by the spoligotyping results in which 38 (16%) of 235 isolates of 15 MIRU-ETR clusters contained isolates that belonged to  $\geq 2$  spoligo families (Table 3).

Applying all 22 loci gave the lowest clustering rate (22.2%) in MTBC strains obtained over 1 year from a single metropolitan setting (London), a rate almost identical to the proportion established in previous studies conducted in London in 1993 and 1995–1997 (22,23) and similar to previously reported rates in population-based studies in low- to-middle TB incidence settings where RFLP and PCR-based genotyping methods were used (11,24–26). These findings suggest, from the public health viewpoint, that TB transmission in London has remained stable over the past decade. Our study provides strong evidence that PCR-based methods, especially VNTR-MIRU, can replace *IS6110* RFLP typing for prospective analysis and that 12 MIRU (27), and 15 MIRU-ETR loci panels alone are insufficiently discriminating for evaluation of TB transmission.

The recently proposed VNTR panel (3,5,7,11) provides similar degrees of discrimination (comparable to that achieved by *IS6110* RFLP), although discrimination of individual VNTR loci is not equal for different MTBC genetic families (13). Inclusion of highly polymorphic VNTR loci effectively differentiates strains within highly conserved groups and is vital for prospective genotyping. Our study demonstrated that even in settings of low TB incidence and relatively low TB transmission rates, TB families, such as Beijing and CAS, remain more conserved than others, and hypervariable loci (e.g., VNTR 3232, 2163A, 4052) provide much higher discrimination than MIRU and ETR loci either alone or in combination.

Our current results agree with the preliminary results of our earlier studies about the applicability of hypervariable VNTR loci (VNTR 3232, VNTR 3336; VNTR 2163a,

and VNTR1982, in particular) and recent reports (28–30) demonstrating their effectiveness for discrimination among Beijing strains. This agreement suggests that these loci are discriminating and reproducible, especially where Beijing strains are dominant (e.g., China, Russia, Baltic countries) (28) and should be included in standardized VNTR panels. They can be used successfully at multiple laboratories with consistent results, provided the conditions for proposed reaction and PCR fragment separation are adhered to and specific DNA polymerases are used.

### Acknowledgments

We thank the reference staff at the Health Protection Agency National Mycobacterium Reference Laboratory for providing the DNA extracts used in this study and the research staff for their assistance with the VNTR typing of all of the isolates.

This research was funded through the UK Department of Health grant “Genotyping of *Mycobacterium tuberculosis* in London.”

Ms Velji is a PhD student at the UK Health Protection Agency Mycobacterium Reference Laboratory, Clinical TB and HIV Group, Barts and The London School of Medicine, Queen Mary, University of London, UK. Her research interests are molecular microbiology and respiratory infections, especially TB.

### References

- World Health Organization (WHO). Global tuberculosis control: surveillance, planning, financing. WHO report 2007. Geneva: The Organization; 2007 [cited 2009 Sep 7]. Available from [www.who.int/tb/publications/global\\_report/2007/en](http://www.who.int/tb/publications/global_report/2007/en)
- Anderson SR, Maguire H, Carless J. Tuberculosis in London: a decade and a half of no decline [corrected]. *Thorax*. 2007;62:162–7. DOI: 10.1136/thx.2006.058313
- Gopaul KK, Brown TJ, Gibson AL, Yates MD, Drobniewski FA. Progression toward an improved DNA amplification-based typing technique in the study of *Mycobacterium tuberculosis* epidemiology. *J Clin Microbiol*. 2006;44:2492–8. DOI: 10.1128/JCM.01428-05
- Kam KM, Yip CW, Tse LW, Leung KL, Wong KL, Ko WM, et al. Optimization of variable number tandem repeat typing set for differentiating *Mycobacterium tuberculosis* strains in the Beijing family. *FEMS Microbiol Lett*. 2006;256:258–65. DOI: 10.1111/j.1574-6968.2006.00126.x
- Nikolayevskyy V, Gopaul K, Balabanova Y, Brown T, Fedorin I, Drobniewski F. Differentiation of tuberculosis strains in a population with mainly Beijing-family strains. *Emerg Infect Dis*. 2006;12:1406–13.
- Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Loch C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol*. 2001;39:3563–71. DOI: 10.1128/JCM.39.10.3563-3571.2001
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-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. DOI: 10.1128/JCM.01392-06
- Roring S, Scott AN, Glyn Hewinson R, Neill SD, Skuce RA. Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. *Vet Microbiol*. 2004;101:65–73. DOI: 10.1016/j.vetmic.2004.02.013
- Kwara A, Schiro R, Cowan LS, Hyslop NE, Wiser MF, Roahen Harrison S, et al. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. *J Clin Microbiol*. 2003;41:2683–5. DOI: 10.1128/JCM.41.6.2683-2685.2003
- van Deutekom H, Supply P, de Haas PE, Willery E, Hoijing SP, Loch C, et al. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit–variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J Clin Microbiol*. 2005;43:4473–9. DOI: 10.1128/JCM.43.9.4473-4479.2005
- Oelemann MC, Diel R, Vatin V, Haas W, Rusch-Gerdes S, Loch C, et al. Assessment of an optimized mycobacterial interspersed repetitive-unit–variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *J Clin Microbiol*. 2007;45:691–7. DOI: 10.1128/JCM.01393-06
- Yokoyama E, Kishida K, Uchimura M, Ichinohe S. Improved differentiation of *Mycobacterium tuberculosis* strains, including many Beijing genotype strains, using a new combination of variable number of tandem repeats loci. *Infect Genet Evol*. 2007;7:499–508. DOI: 10.1016/j.meegid.2007.02.006
- Arnold C, Thorne N, Underwood A, Baster K, Gharbia S. Evolution of short sequence repeats in *Mycobacterium tuberculosis*. *FEMS Microbiol Lett*. 2006;256:340–6. DOI: 10.1111/j.1574-6968.2006.00142.x
- Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S, et al. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. *J Clin Microbiol*. 2002;40:2126–33. DOI: 10.1128/JCM.40.6.2126-2133.2002
- Kremer K, Arnold C, Cataldi A, Gutierrez MC, Haas WH, Panaiotov S, et al. Discriminatory power and reproducibility of novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. *J Clin Microbiol*. 2005;43:5628–38. DOI: 10.1128/JCM.43.11.5628-5638.2005
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol*. 2006;6:23. DOI: 10.1186/1471-2180-6-23
- Vitol I, Driscoll J, Kreiswirth B, Kurepina N, Bennett KP. Identifying *Mycobacterium tuberculosis* complex strain families using spoligotypes. *Infect Genet Evol*. 2006;6:491–504. DOI: 10.1016/j.meegid.2006.03.003
- Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson’s index of diversity. *J Clin Microbiol*. 1988;26:2465–6.
- Sola C, Filliol I, Legrand E, Lesjean S, Loch C, Supply P, et al. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. *Infect Genet Evol*. 2003;3:125–33. DOI: 10.1016/S1567-1348(03)00011-X
- Warren RM, Victor TC, Streicher EM, Richardson M, van der Spuy GD, Johnson R, et al. Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS6110 copy numbers. *J Clin Microbiol*. 2004;42:5774–82. DOI: 10.1128/JCM.42.12.5774-5782.2004
- Frothingham R, Meeker-O’Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology*. 1998;144:1189–96. DOI: 10.1099/00221287-144-5-1189

22. Maguire H, Dale JW, McHugh TD, Butcher PD, Gillespie SH, Costetsos A, et al. Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission. *Thorax*. 2002;57:617–22. DOI: 10.1136/thorax.57.7.617
23. Hayward AC, Goss S, Drobniewski F, Saunders N, Shaw RJ, Goyal M, et al. The molecular epidemiology of tuberculosis in inner London. *Epidemiol Infect*. 2002;128:175–84. DOI: 10.1017/S0950268801006690
24. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med*. 1994;330:1703–9. DOI: 10.1056/NEJM199406163302402
25. Allix-Beguec 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. DOI: 10.1128/JCM.02089-07
26. Durmaz R, Zozio T, Gunal S, Allix C, Fauville-Dufaux M, Rastogi N. Population-based molecular epidemiological study of tuberculosis in Malatya, Turkey. *J Clin Microbiol*. 2007;45:4027–35. DOI: 10.1128/JCM.01308-07
27. Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Loch C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol*. 2001;39:3563–71. DOI: 10.1128/JCM.39.10.3563-3571.2001
28. Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetsov S, Zakharova S, et al. Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA*. 2005;293:2726–31. DOI: 10.1001/jama.293.22.2726
29. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan. *J Med Microbiol*. 2007;56:1052–7. DOI: 10.1099/jmm.0.46990-0
30. Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, et al. Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. *FEMS Microbiol Lett*. 2007;270:67–74. DOI: 10.1111/j.1574-6968.2007.00658.x

Address for correspondence: Francis Drobniewski, UK Health Protection Agency Mycobacterium Reference Unit, Clinical TB and HIV Research Group, Queen Mary, University of London, 2 Newark St, E1 2AT, London, UK; email: f.drobniewski@qmul.ac.uk

# EMERGING INFECTIOUS DISEASES

A Peer-Reviewed Journal Tracking and Analyzing Disease Trends

Vol.12, No.1, January 2006



Search  
past issues

**EID**  
online  
www.cdc.gov/eid

# Discriminatory Ability of Hypervariable Variable Number Tandem Repeat Loci in Population-based Analysis of *Mycobacterium tuberculosis* Strains, London, UK

## Technical Appendix

Allelic diversity and discriminative values of VNTR loci

### A. MIRU-ETR loci

Families	HGDI value															
	2	4	10	16	20	23	24	26	27	31	39	40	A	B	C	Mean for 15 loci
EAI (N = 395)	0.01	0.577	0.299	0.363	0.097	0.411	0.115	0.089	0.259	0.637	0.583	0.667	0.788	0.788	0.402	0.406
Beijing (N = 131)	0	0.031	0.296	0.131	0.015	0.018	0.015	0.540	0.354	0.075	0.272	0.377	0.234	0	0.089	0.163
CAS (N = 555)	0.057	0.015	0.701	0.342	0.046	0.053	0.015	0.702	0.044	0.451	0.236	0.49	0.315	0.04	0.014	0.235
T (N = 371)	0.119	0.247	0.693	0.630	0.177	0.393	0.049	0.665	0.197	0.352	0.109	0.751	0.585	0.210	0.497	0.378
Haarlem (N = 207)	0.085	0.131	0.678	0.372	0.325	0.557	0.04	0.413	0.145	0.209	0.076	0.495	0.302	0.315	0.463	0.307
S (N = 20)	0.468	0.485	0.298	0.432	0.426	0.511	0	0.679	0.105	0.433	0.1	0.684	0.526	0.205	0.279	0.375
X (N = 108)	0.262	0.319	0.525	0.073	0.230	0.160	0	0.493	0.091	0.435	0.073	0.731	0.353	0.232	0.704	0.312
LAM (N = 350)	0.173	0.045	0.570	0.600	0.457	0.628	0.011	0.411	0.401	0.211	0.04	0.603	0.567	0.405	0.29	0.361
<i>M. bovis</i> (N = 75)	0	<b>0.529</b>	0.054	0.240	0.027	0.053	0.054	0.178	0.207	0.079	0	0.027	0.248	0.335	0.085	0.141
<i>M. afr</i> (N = 46)	0	0.406	0.686	0.605	0.087	0.044	0.390	0.380	0.465	0.641	0	0.489	0.591	0.509	0.555	0.390
<i>All families</i>	0.134	0.524	0.781	0.65	0.196	0.615	0.454	0.797	0.303	0.72	0.559	0.727	0.813	0.551	0.667	-
<i>No of allelic variants</i>	4	11	11	9	3	11	5	11	6	8	4	11	14	9	8	-

B. Additional panel of 7 VNTR loci

Families	HGDI value							Mean for 7 VNTR loci
	2163b	2347	3232	2163a	1982	3336	4052	
EAI (N = 109)	0.804	0.14	0.123	<b>0.821</b>	0.796	0.762	0.797	0.606
Beijing (N = 96)	0.737	0.081	<b>0.830</b>	0.615	0.399	0.351	0.531	0.506
CAS (N = 332)	0.104	0.042	0.846	0.331	0.635	0.696	<b>0.855</b>	0.501
T (N = 208)	0.821	0.055	0.804	0.779	0.651	<b>0.832</b>	0.806	0.678
Haarlem (N = 128)	0.826	0.434	<b>0.935</b>	0.684	0.760	0.812	0.709	0.737
S (N = 5)	0.4	0	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>	<b>0.7</b>	0.557
X (N = 56)	0.616	0.036	0.591	0.478	<b>0.820</b>	0.714	0.787	0.577
LAM (N = 252)	<b>0.880</b>	0.192	0.8	0.589	0.584	0.826	0.757	0.661
<i>M. bovis</i> (N = 13)	0	0	0	0	0	0.167	0	0.024
<i>M. afr</i> (N = 6)	0.7	0.143	<b>0.824</b>	0.681	0.495	0.495	0.813	0.593
All families	0.790	0.382	<b>0.909</b>	0.882	0.843	0.887	0.827	-
No. allelic variants	13	6	19	18	14	15	10	-