



## **HPA *Mycobacterium tuberculosis* Strain Typing: A guide to data production and distribution.**

**August 2012**

**This guide was prepared on behalf of the HPA TB Delivery Board by Dr Tim Brown (HPA NMRL), Dr Jason Evans (HPA Birmingham RCM) and Dr Andy Sails (HPA Newcastle RCM), with contributions from the TBDB, Dr Laura Anderson (HPA HPS TB Section), Dr Madeline Stone (HPA NMRL), Dr Heather Jebbari (HPA NMRL) and Dr Vlad Nikolayevskyy (HPA NMRL).**

### **Table of Contents**

1	Overview. ....	2
2	Background. ....	2
3	Purpose. ....	4
4	Method of delivery. ....	5
5	Investigation.....	10
6	Portability of data. ....	10
7	Ongoing developments.....	13
8	References. ....	13

## 1 Overview.

This guide is intended to complement the “TB strain typing handbook for Health Protection Units” by giving the users of typing data an insight into how and why the HPA is using 24VNTR data and how it is being generated, how the laboratory data is analysed to how and to whom results are reported.

## 2 Background.

DNA typing has made a major contribution to the understanding of both *Mycobacterium tuberculosis* complex (MTBC) epidemiology and evolution. High levels of strain discrimination are required in clinical settings where questions concerning transmission of MTBC between patients, reactivation versus re-infection in an individual or laboratory cross contamination are asked. Lower levels of strain discrimination are required when studying MTBC evolution where robust markers identifying significant genetic events are required for construction of phylogenetic lineages.

In 1991, the typing of TB strains using IS6110 restriction fragment length polymorphism analysis (RFLP) was described and quickly became the standard for epidemiological investigations [1]. However, technical difficulties in data production and comparison of output between laboratories restricted its use to specialist centres and the format used had limited value for investigating the phylogeny of MTBC.

The use of ‘spacer oligonucleotide’ or ‘spoligotyping’ as described in 1997 [2] for the typing of MTBC, is a PCR based method which circumvents the technical difficulties associated with IS6110 RFLP typing, producing digital data allowing simple data analysis and data sharing between laboratories. Spoligotype data has been found to be useful in distinguishing between the members of the MTBC and defining *M. tuberculosis* clades and families in phylogenetic studies [3]. Spoligotyping has been shown to be of limited value for epidemiological transmission studies due to its low discriminative power when used alone [4]. However the combination of IS6110 RFLP analysis and spoligotyping provided a useful epidemiological tool.

The analysis of repeated DNA motifs often referred to as minisatellites or variable number tandem repeats (VNTR) for human fingerprinting was first described in 1985 [5]. Similar features were identified in *M. tuberculosis* [6] and initially six VNTR loci, designated exact tandem repeats (ETR) A to E, were shown to be simple to amplify by PCR and polymorphic but their level of discrimination was not sufficient to be useful for MTBC epidemiological investigations [4]. Subsequently with the publication of the *M. tuberculosis* H37Rv genome a total of 41 VNTR loci were identified [7] and designated mycobacterial interspersed repetitive units (MIRU) 1 to 41; 12 of which were shown to be polymorphic in *M. tuberculosis* strains. This panel of 12 MIRU/VNTR loci have been repeatedly shown to be useful in epidemiological investigations [8, 9]. The discriminative power of a series of VNTR locus panels consisting of various numbers and combinations of the 59 polymorphic VNTR loci that have been described to date [6, 7, 10-15] have been evaluated against both test collections of MTBC strains [10, 12, 13] [16] and in population studies [17]. These studies have shown that levels of discrimination can be achieved which are in excess of those seen with the original 12 MIRU/VNTR panel. In 2004 the HPA standardised its typing service across its constituent TB laboratories in London, Birmingham and Newcastle by combining the ETR and MIRU loci into a panel of 15 loci. In 2006 a standard panel of 24VNTR was proposed[18] which attempted to optimise both reproducibility and discrimination. This panel has rapidly become the global standard method for typing MTBC with thirty seven laboratories from across the world participating in an ECDC multi centre proficiency study for 24-locus VNTR typing [19] in 2009. The 24 VNTR panel consists of the 2004 HPA 15VNTR panel plus an additional 9 VNTRs giving increased discrimination.

In addition to its value as an epidemiological tool VNTR data has been shown to be a useful tool in phylogeny studies [20] and the 24 VNTR panel has been shown as useful for the typing of non *M.tuberculosis* members of the MTBC [21].

The use of 24 VNTR analysis has made routine universal typing programmes a practical proposition due to the high levels of strain discrimination achieved, the methodology being amenable to automation, the ease of analysis and the portability of the typing data.

### **3 Purpose.**

Before the development of 24 VNTR loci typing molecular strain typing in the UK and elsewhere was largely restricted in routine settings to retrospective investigations confirming or refuting suspected relationships within outbreaks identified by traditional epidemiology, confirming or refuting cases of reactivation, investigating suspected cross contamination events and research studies.

When VNTR data is routinely generated for all MTBC isolates at the time of identification these questions can be answered at the time they are posed rather than days or weeks later in the case of a retrospective approach. Suspected transmission can be refuted or confirmed before public health measures and outbreak investigations are initiated assisting in the directing and focusing of public health resources. Timely determination of reactivation can effect patient management by informing questions around drug resistance and patient compliance. Finally identification of specimen cross contamination can prevent unnecessary drug therapy.

In addition to expediting these answers prospective typing can perform functions that retrospective typing cannot. Groups of patients can be identified that are likely to be epidemiologically linked where there is no pre-existing suspicion and likewise patients with no obvious traditional epidemiological links can be identified as being part of a known outbreak. The prospective approach also allows the identification of potential cross contamination events where there is no pre-existing suspicion. Clustering of strain typing fingerprints has been shown to be indicative of recent transmission [22] therefore the clustering rate seen in a population is indicative of the level of transmission in that population. Prospective typing means that a calculated clustering rate can be used as surveillance tool to indicate changes in transmission rates in the population and hence the efficacy of TB controls measures.

The HPA commenced a prospective strain typing programme of 24VNTR typing of all MTBC isolates in January 2010. This is a multidisciplinary programme involving a

team of laboratory, epidemiology, health protection and contact tracing/investigation personnel integrating all the processes from identification of MTBC in cultures through to the initiation of public health action.

#### **4 Method of delivery.**

a. Patient isolates.

Where MTBC infection is suspected in clinic, material is submitted to a local microbiology laboratory for culture. Once growth of acid fast bacteria is detected in a liquid or solid culture it is sent to a Regional Centre for Mycobacteriology (RCM) for identification and if identified as MTBC drug susceptibility testing will be set up. At this point a single, usually the first culture received at reference laboratories from each new patient is selected for typing. Subsequent cultures are typed when dated > 2 months after the first. DNA extracts are prepared from these cultures for further analysis.

b. VNTR.

A set of 24 PCR DNA amplifications are performed on these DNA extracts one for each VNTR locus to be analysed. The resulting PCR products consist of the VNTRs and flanking regions of known length containing the PCR primer targets. VNTRs are enumerated by determining the size of the PCR product, each VNTR number giving a signature product size. Three technologies are used for product sizing all rely on UV detection of PCR product.

- i. Automated Capillary electrophoresis. PCR products generated using primers labelled with fluorescent dyes are drawn by electrophoresis through an acrylamide matrix housed in a fine capillary, the time of travel being related to the size of the fragment. The passage of PCR product through the capillary is detected by means of laser excitation of the fluorescent dye incorporated into the product and a UV detector. This analysis is performed on commercially available automated systems (Beckman Coulter or ABI).

- ii. dHPLC. Non-dHPLC employs reverse-phase ion pair chromatography in which DNA binds to the column, mediated by the ion-pairing reagent triethylammonium acetate; subsequently, this DNA bonding to the column is disrupted by a linear gradient of acetonitrile, which results in the movement of the DNA from the stationary phase into the mobile phase and progress through the column to the UV detector[23]. This analysis is performed on commercially available automated systems (Transgenomic WAVE Inc.).
- iii. Slab gel electrophoresis. PCR products are drawn by electrophoresis through an agarose matrix cast as a gel slab, the time of travel being related to the size of the fragment. Gels are stained using fluorescent dye and PCR products are visualised by UV excitation of the dye incorporated into the product.

The choice of technology employed is driven by the throughput required by each laboratory and the expected fragment sizes, the automated systems having an effective sizing limit of c1000bp.

c. Analysis.

The size of a PCR product is determined by comparing the migration or elution time against that seen with standards sets run in the same system. In the case of the automated systems software incorporated into each system is used to assign sizes and determine copy number at each VNTR locus. Where slab gels are used sizes are estimated and VNTR copy number assigned by manual inspection. Outputs are exported in standard spreadsheet format for downstream processing.

d. Quality control.

In order to maximize the value of a data set generated at multiple laboratories across the UK and have confidence in the portability of data internationally, three levels of laboratory quality control have been built into the HPA strain typing programme:

- i. IQC. The analysis of a minimum of 1% of all cultures processed at each laboratory is repeated and consistency monitored.
  - ii. EQA. All the laboratories participate in the NEQAS mycobacteriology EQA. This scheme distributes 3 sets of 4 samples per year. Where samples contain MTBC typing data is submitted and the results are scored.
  - iii. HPA. In order to rapidly identify and resolve inconsistencies between data produced across the laboratory network contributing to this programme the NMRL produce and distribute a panel of 8 DNA samples three times per year. These are interspersed with the NEQAS distributions meaning a panel is analysed at each lab at 2 month intervals. Results are returned to NMRL where they are collated and once the set is complete distributed to all participants. A teleconference including all participating labs is used to resolve discrepancies and identify corrective actions within the laboratory network.
- e. Reporting. The individual 24VNTR types are reported at two levels.
  - i. To the laboratory submitting the positive culture on the standard lab report containing identification and susceptibility test data.
  - ii. To the Local Health Protection Unit within whose geographical footprint the patient is being treated.
- f. Surveillance. 24VNTR types are loaded onto the Enhanced Tuberculosis surveillance (ETS) system, this being an HPA on-line public health surveillance tool that combines patient, laboratory and epidemiological data.
- g. Cluster identification. Individual VNTR 24 digit codes standing alone have little value. Their power emerges when combined in sets of multiple loci to identify groups of indistinguishable isolates which in turn identify the patients from which they were cultured as potentially epidemiologically linked. To establish whether recent transmission has occurred within these

groups of patients traditional epidemiological data is required. It is convenient to use a short-hand “cluster name” to identify a molecular cluster rather than the full 24 character code which for most is difficult to commit to memory. To enable the investigation of possible national transmission networks a universal nomenclature within the UK and beyond is required.

The following method is used to designate cluster numbers;

- i. A cluster comprises  $\geq 2$  isolates with indistinguishable 24VNTR codes.
- ii. For an isolate to be included in a cluster  $\geq 22$  digits must be present in the 24VNTR code.
- iii. An unequivocal cluster number is given when  $\geq 1$  members of the cluster have 24 digits.
- iv. Isolates are not included in a cluster if the permitted  $\leq 3$  missing loci do not at least correspond to all those missing in the most complete member of a cluster.
- v. Where missing loci fail to place an isolate unequivocally within a single cluster, typing is repeated.
- vi. Prefix letters indicating the phylogenetic lineage are determined using table1.

MtbC Lineage	Locus - repeat number	Prefix
I (Beijing)	39-3, A-4, C-4, 24-1	B
II (Euro-american)	16-1,2 or 3, 39-2, B-1 or 2	E
III (CAS)	23-5, C-2	C
IV (EAI)	24-2, 26-2	A
<i>M. bovis</i>	10-2, 40-2, C-5	V
<i>M. africanum</i>	10 $\geq$ 4, 23-4, 26-3,4 or 5, A $\geq$ 4	F
<i>M. microti</i>	23-4, 40-2, A $\geq$ 8	M
None of the above		U
Multiple of the above		X

Table 1. VNTR signatures that identify global MTBC lineages as described elsewhere [20, 24].

- vii. The cluster designation is completed by adding a unique 4 digit number in the form 0001.
- h. Reporting of clusters. Reports are created within xls sheets on a monthly basis and distributed using encrypted email, these contain 3 types of entry.
  - i. New clusters with indistinguishable types isolated within an HPU area.
  - ii. Previously reported clusters within an HPU area with new members.
  - iii. New clusters with indistinguishable types isolated within an HPU area that are indistinguishable from larger clusters whose members originate from multiple HPU areas.

Reports are sent to Regional HPA offices where clusters contain isolates originating from >1 HPU area within a region and the TB section HPS where

clusters that contain isolates originating from >1 region are investigated. All clusters reports are copied to the HPA TB cluster investigators.

- i. An on-line tool <http://www.hpa-bioinformatics.org.uk/TBCluster/tbhome.php> using the rules shown in table 1 has been developed by the HPA as a repository for cluster names enabling all the laboratories in the network to use the same cluster designations.

## 5 Investigation.

- a. When cluster reports are received by HPUs, regional and National HPA offices their significance is assessed as outlined in the *TB Strain Typing Cluster Investigation Handbook for Health Protection Units*.  
[http://www.hpa.org.uk/webc/HPAwebFile/HPAweb\\_C/1317131018354](http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317131018354)
- b. Data from ETS and the London Tuberculosis register (LTBR) is used alongside cluster reports to gauge appropriate action. The National TB office (TB Section Colindale) collates cluster reports from the RCM laboratories producing a complete picture of national clusters (clusters spanning >1 region).

## 6 Portability of data.

The transfer of typing data between the constituent parts of the HPA system has been simplified by the adoption of a set of conventions covering repeat enumeration and the standardization of the report data string.

- a. HPA 24VNTR calling nomenclature.  
VNTRs at a number of loci within the *M.tuberculosis* genome are flanked by partial repeats. A variety of analytical tools have been developed to

enumerate VNTR at these loci some of which have included these repeats in their calling, leading to systematic differences in called number depending on which system is used. In order to compare data generated at different centres a standard enumeration system and calling nomenclature must be used.

- i. 24VNTR typing data is reported as a 24 digit string in the following order. 5xETR A to E followed by 10xMIRU  
2,10,16,20,23,24,26,27,39,40 followed by 9xVNTR  
424,1955,2163b,2347,2401,3171,3690,4052,4156
- ii. The enumeration at each genetic locus should be as used at Rijksinstituut voor Volksgezondheid en Milieu (RIVM)/ECDC for calling their annual QC panels which can be used to index HPA VNTR calling.
- iii. Where a locus contains >9 repeats capital letters are used to indicate copy number A for 10 and so on in order to preserve the 24 character string.
- iv. Where a partial deletion of a repeat is present in ETR-D all repeats are counted including the partial repeat and the presence of the partial repeat in the count is indicated by a '.
- v. H37Rv obtained from Pasteur Institute Paris, this being the stock from which the published H37Rv genome was derived, is held at the NMRL. Using the agreed calling protocol identified in i to iv above the following VNTR code should be obtained from this material.  
3323'32322613321225223552

b. Transposition from non HPA centres.

A variety of reporting systems and locus names are used for 24VNTR typing both in the literature and by centres across the world. In order to make comparisons between VNTR types it is often necessary to transpose the data to a common form. Table 2 shows the most common aliases used for the 24

loci and should be used in conjunction with locus lists obtained from the centres producing the types for comparison.

H37Rv genome position	VNTR name	Aliases					
		HPA	ETR [6]	MIRU [7]	Mtub[10]	QUB [12, 13]	RIVM
2165223	2165	ETR-A	ETR-A				
2461279	2461	ETR-B	ETR-B				VNTR48
577172	577	ETR-C	ETR-C				VNTR43
580578	580	ETR-D	ETR-D	MIRU4			
3192202	3192	ETR-E	ETR-E	MIRU31			
154073	154	MIRU2		MIRU2			
960173	960	MIRU10		MIRU10		QUB-0960C	
1644261	1644	MIRU16		MIRU16			
2059441	2059	MIRU20		MIRU20			
2531898	2531	MIRU23		MIRU23		QUB-2531C	
2687128	2687	MIRU24		MIRU24			
2996003	2996	MIRU26		MIRU26			
3007063	3007	MIRU27		MIRU27		QUB-5	
4348721	4348	MIRU39		MIRU39			
802429	802	MIRU40		MIRU40			
424010	424	VNTR 424			Mtub4		VNTR42
1955580	1955	VNTR 1955			Mtub21		
2163741	2163b	VNTR 2163b				QUB-11b	
2347393	2347	VNTR 2347			Mtub29		VNTR46
2401815	2401	VNTR 2401			Mtub30	QUB-2401	VNTR47
3171468	3171	VNTR 3171			Mtub34		VNTR49
3690947	3690	VNTR 3690			Mtub39		VNTR52
4052971	4052	VNTR 4052				QUB-26	
4156797	4156	VNTR 4156				QUB-4156c	VNTR53

Table 2. Different names given to the same VNTR loci by different centres and publications. The descending order shown in this table is the order used in the standard HPA string.

## 7 Ongoing developments.

- a. Increase discrimination of molecular tools
  - vi. Additional VNTR. The 24VNTR panel gives a high level of discrimination but epidemiological investigation of some clusters has been aided by the use of additional experimental VNTR panels [17].
  - vii. Whole genome data. Whole genome data has been shown to be useful in understanding the relationships between MTBC strains [25, 26]. The utility of this data within this programme is being investigated.
- b. Modelling of transmission/risk of transmission.

## 8 References.

1. van Soolingen, D., et al., *Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis*. J. Clin. Microbiol., 1991. **29**(11): p. 2578-2586.
2. Kamerbeek, J., et al., *Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology*. J Clin Microbiol, 1997. **35**(4): p. 907-14.
3. Brudey, K., et al., *Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology*. BMC Microbiology, 2006. **6**(1): p. 23.
4. Kremer, K., et al., *Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility*. J Clin Microbiol, 1999. **37**(8): p. 2607-18.
5. Jeffreys, A.J., V. Wilson, and S.L. Thein, *Individual-specific /'fingerprints/' of human DNA*. 1985. **316**(6023): p. 76-79.
6. Frothingham, R. and W. Meeker-O'Connell, *Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats*. Microbiology, 1998. **144**(5): p. 1189-1196.
7. Supply, P., et al., *Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome*. Mol Microbiol, 2000. **36**(3): p. 762-71.
8. Mazars, E., et al., *High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology*. Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1901-6.
9. Hawkey, P.M., et al., *Mycobacterial Interspersed Repetitive Unit Typing of Mycobacterium tuberculosis Compared to IS6110-Based Restriction Fragment Length Polymorphism Analysis*

- for Investigation of Apparently Clustered Cases of Tuberculosis*. J. Clin. Microbiol., 2003. **41**(8): p. 3514-3520.
10. Le Fleche, P., et al., *High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing*. BMC Microbiology, 2002. **2**(1): p. 37.
  11. Namwat, W., P. Luangsuk, and P. Palittapongarnpim, *The genetic diversity of Mycobacterium tuberculosis strains in Thailand studied by amplification of DNA segments containing direct repetitive sequences*. International journal of tuberculosis and lung disease, 1998. **2**(2): p. 153-159.
  12. Roring, 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**(6): p. 2126-2133.
  13. Skuce, R.A., et al., *Discrimination of Mycobacterium tuberculosis complex bacteria using novel VNTR-PCR targets*. Microbiology, 2002. **148**(2): p. 519-528.
  14. Smittipat, N., et al., *Polymorphism of Variable-Number Tandem Repeats at Multiple Loci in Mycobacterium tuberculosis*. J. Clin. Microbiol., 2005. **43**(10): p. 5034-5043.
  15. Spurgiesz, R.S., et al., *Molecular Typing of Mycobacterium tuberculosis by Using Nine Novel Variable-Number Tandem Repeats across the Beijing Family and Low-Copy-Number IS6110 Isolates*. J. Clin. Microbiol., 2003. **41**(9): p. 4224-4230.
  16. van Deutekom, H., 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**(9): p. 4473-4479.
  17. Gopaul, K.K., et al., *Progression Toward an Improved DNA Amplification-Based Typing Technique in the Study of Mycobacterium tuberculosis Epidemiology*. J. Clin. Microbiol., 2006. **44**(7): p. 2492-2498.
  18. Supply, P., et al., *Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis*. J. Clin. Microbiol., 2006. **44**(12): p. 4498-4510.
  19. de Beer, J.L., et al., *First Worldwide Proficiency Study on Variable-Number Tandem-Repeat Typing of Mycobacterium tuberculosis Complex Strains*. J. Clin. Microbiol. 2011. **50**(3): p. 662-669.
  20. Gibson, A., et al., *Can 15-Locus Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Analysis Provide Insight into the Evolution of Mycobacterium tuberculosis?* Appl. Environ. Microbiol., 2005. **71**(12): p. 8207-8213.
  21. Stone, M.J., T.J. Brown, and F.A. Drobniowski, *Human Mycobacterium bovis Infections in London and Southeast England*. J. Clin. Microbiol. 2012 **50**(1): p. 164-165.
  22. Small, P.M., et al., *The Epidemiology of Tuberculosis in San Francisco -- A Population-Based Study Using Conventional and Molecular Methods*. N Engl J Med, 1994. **330**(24): p. 1703-1709.
  23. Evans, J.T., et al., *Automated High-Throughput Mycobacterial Interspersed Repetitive Unit Typing of Mycobacterium tuberculosis Strains by a Combination of PCR and Nondenaturing High-Performance Liquid Chromatography*. J. Clin. Microbiol., 2004. **42**(9): p. 4175-4180.
  24. Brown, T.J., et al., *Associations between Mycobacterium tuberculosis strains and phenotypes*. Emerg Infect Dis. , 2010 **16**(2): p. 8.
  25. Casali, N., et al., *Microevolution of extensively drug-resistant tuberculosis in Russia*. Genome Research. 2012. **22**(4): p. 735-45.
  26. Gardy, J.L., et al., *Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak*. New England Journal of Medicine. 2011. **364**(8): p. 730-739.