

# Tuberculosis Genotyping in British Columbia

10-year Retrospective Study Report

Prepared by: Jennifer L. Guthrie, PhD Candidate

Email: [jennifer.guthrie@alumni.ubc.ca](mailto:jennifer.guthrie@alumni.ubc.ca)

School of Population and Public Health

University of British Columbia

June 2018



# Tuberculosis Genotyping in British Columbia

10-year Retrospective Study Report

## Summary

In 2012, a project was initiated by the British Columbia Centre for Disease Control (BCCDC) to retrospectively genotype the first *Mycobacterium tuberculosis* (*Mtb*) isolate from each patient with a culture confirmed diagnosis of tuberculosis (TB) using 24-locus Mycobacterial Interspersed Repetitive Unit - Variable Number Tandem Repeat (MIRU-VNTR). This report describes the resulting cluster analyses and includes the geographical and temporal distribution of large ( $\geq 10$  persons) genotype clusters for *Mtb* isolated from specimens received at the BCCDC Public Health Laboratory (BCCDC PHL) from 2005 through 2014.

Overall, MIRU-VNTR genotyping grouped 2,290 isolates into 189 clusters (2–70 isolates/cluster) with an overall clustering rate of 42.4% and an estimated endemic transmission rate of 34.1% (“*n*-1” method).<sup>1</sup> Large clusters ( $\geq 10$  persons) occurred more frequently within the *Mtb* Euro-American lineage and included mainly Canadian-born persons (87.1%–100.0%). For full details of the *Mtb* molecular epidemiology in British Columbia see Guthrie et al. (2017).<sup>2</sup>

## Key Facts



No. Isolates Genotyped  
2,290

No. Distinct Genotypes  
1,508

Percentage Clustered  
42%

No. Clusters  
189

Cluster Size Range  
2–70

Canadian-born Clustered  
77%

Foreign-born Clustered  
30%

Urban Clustered  
39%

Rural Clustered  
74%



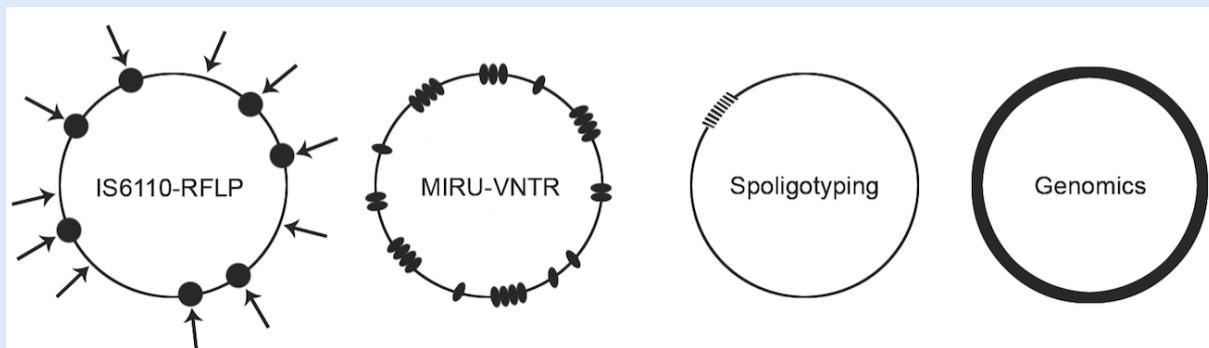
## Table of Contents

Summary.....	1
Introduction to Genotyping .....	3
What is 24-locus MIRU-VNTR? .....	4
How is Genotyping Used? .....	4
Limitations.....	5
Data and Analysis.....	6
TB Genotyping in British Columbia Infographic.....	9
MIRU-VNTR Cluster Summaries.....	10
References .....	27
Appendix I: 24-LOCUS MIRU-VNTR PATTERNS OF LARGE CLUSTERS .....	29
Appendix II: MIRU-VNTR ALIASES .....	30

## Introduction to Genotyping

*Mycobacterium tuberculosis* (*Mtb*) genotyping uses DNA based techniques to target specific segments of the genome allowing for the differentiation of *Mtb* strains. Genotyping has a number of public health and research applications, which will be discussed in a later section.

24-locus Mycobacterial Interspersed Repetitive Unit - Variable Number Tandem Repeat (MIRU-VNTR) genotyping has become the standard tool for molecular typing of *Mtb* for many TB programs world-wide. As a rapid technique resulting in a portable digital signature, MIRU-VNTR has replaced genotyping by restriction fragment length polymorphism (RFLP) in most laboratories. A similarly rapid method known as spoligotyping is often used in molecular studies; however, its low resolution makes it unsuitable for inferring transmission. Genomics, which utilizes the entire genome sequence, is the most recent method available and has the highest discriminatory power; however, sequencing technology and analyses have not been fully standardized for routine use and at this time is most often used for research purposes or specific outbreak investigations.

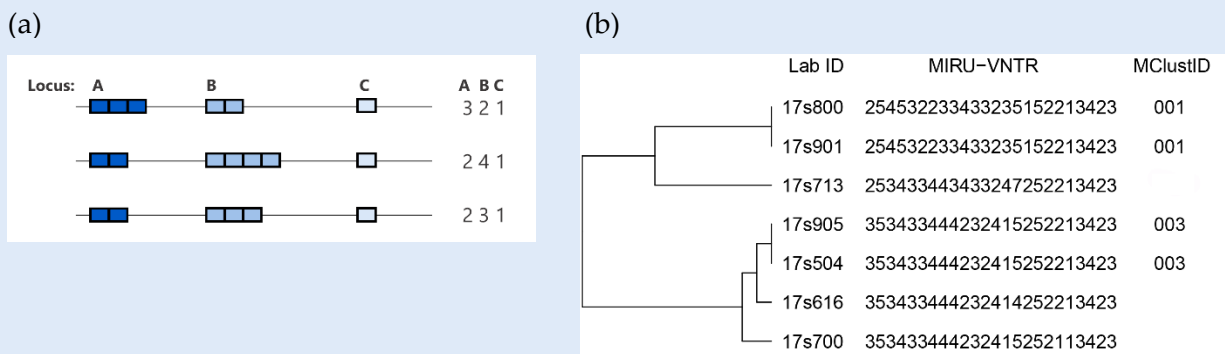


**Figure 1. Common Molecular Methods for Genotyping *Mycobacterium tuberculosis*.** This simple schematic, not to scale, compares four common methods used in TB molecular epidemiology, and the markings provide an appreciation of the targeted regions for analysis. In contrast, genomics which uses whole genome sequencing interrogates the entire genome, with single nucleotide polymorphisms revealing the relationship between isolates.



## What is 24-locus MIRU-VNTR?

The BCCDC PHL uses a standard method<sup>3</sup> of 24-locus MIRU-VNTR for routine genotyping. MIRUs (Mycobacterium Interspersed Repetitive Units) represent repeated DNA sequences 40 to 110 base pairs long which are found in a number of locations around the *Mtb* genome.<sup>4</sup> MIRU-VNTR genotyping is performed by PCR amplification of each MIRU locus using primers specific for the flanking region. Following capillary electrophoresis, the size of each amplicon is determined, and calculations are performed based on the known length of the repeat unit at each locus. The number of repeats at each of the 24 loci are combined to generate a digital signature that can be used to determine the phylogenetic structure and epidemiological links between strains.



**Figure 2. 24-locus MIRU-VNTR genotyping.** (a) Schematic illustrating the principle of MIRU-VNTR. Each rectangle represents a repeated DNA sequence which are counted to determine the number of repeats at each locus. (b) Example dendrogram of 24-locus MIRU-VNTR profiles for 7 patients identified by a unique Lab ID. Twenty-four different loci are analyzed for each isolate and the resulting digital code is compared between patient isolates. A unique cluster identifier (MClustID) is assigned for patients with identical MIRU-VNTR profiles.

## How is Genotyping Used?

As previously stated, genotyping data has numerous public health and research applications. When combined with epidemiological information *Mtb* genotyping can be a very useful tool. Genotyping results have been used to detect specimen mix-up/laboratory cross-contamination events, identify outbreaks, confirm/refute suspected transmission and differentiate between reinfection and reactivation of tuberculosis. Studies have shown that the routine use of genotyping data enhances contact investigations, leads to more effective use of resources, and can uncover previously unrecognized sources and sites of transmission.<sup>5,6</sup> Furthermore, genotyping data can be used to monitor clusters over time, evaluate program performance, and understand *Mtb* population dynamics in a particular region or setting.



### Limitations

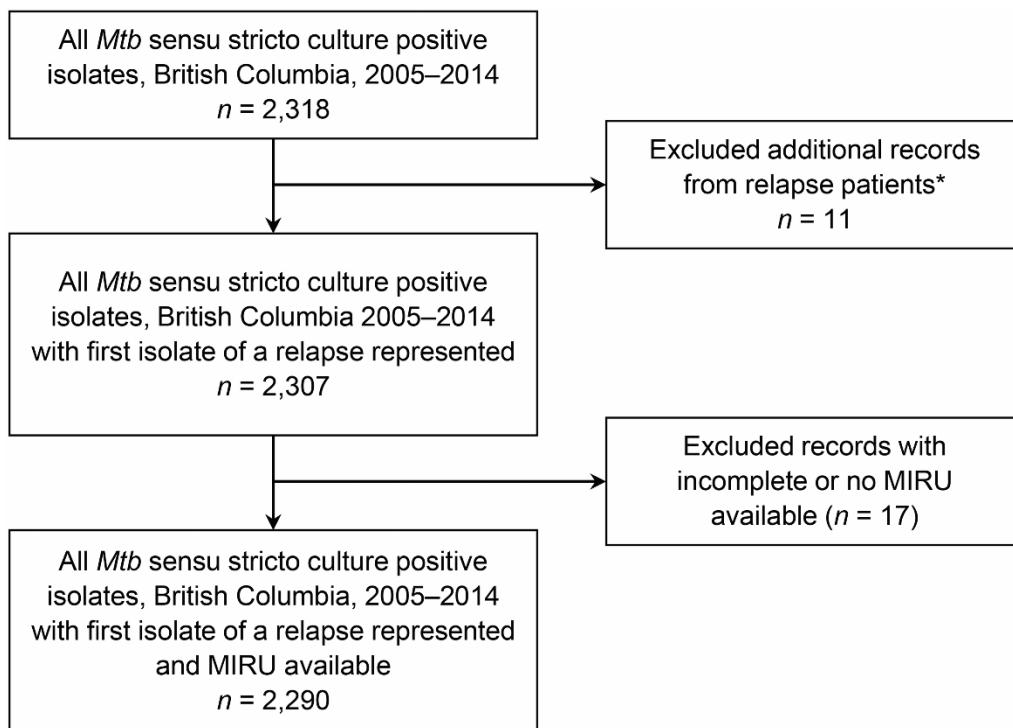
It should be noted that as with any biological test there are limitations to its interpretation. In the case of *Mtb* genotyping, the first limitation is a technical one. Bacterial isolation is required for DNA extraction and genotyping. Consequently, clinically diagnosed cases without culture confirmation (~20% of TB diagnoses in BC) are not able to be genotyped, and therefore their strain type cannot be matched to other cases and will not contribute to the genotyping database. The second issue involves the testing methodology. Standard PCR primers have been designed based on the most common DNA sequences found across *Mtb* strains used during method development but cannot capture all possible sequences that may exist globally and mismatched sequences may result in poor or failed amplification. Moreover, some strains may have a large number of repeats for a particular locus (e.g. MIRU-4052) causing the amplicon size to exceed the upper limit of the genetic analyzer instrument. Genomic rearrangements due to mobile genetic elements may also prevent amplification of some loci (e.g. MIRU-2163 and MIRU-2165). The result of these technical issues is an incomplete MIRU-VNTR pattern which impedes interpretation and, in most cases, does not allow for cluster assignment.

Further limitations involve epidemiological interpretation of genotype information. First, directionality of transmission cannot be determined by genotype data alone, and clustering only indicates that patient strains are genotypically related. Secondly, genotypic clustering does not necessarily mean transmission has occurred between the patients in question, especially in particular patient groups. Whole genome sequencing (WGS) of large MIRU-VNTR clusters elsewhere in Canada, belonging to the Indo-Oceanic lineage revealed that these clusters were not representative of local transmission but rather a common country or region of origin.<sup>7</sup> In contrast, WGS of large MIRU-VNTR clusters comprised largely of Canadian-born individuals in both BC and Ontario have demonstrated that these clusters represent ongoing local transmission of TB.<sup>8-10</sup> Ultimately, genotypic clustering should always be interpreted within the context of epidemiological information.

## Data and Analysis

The study population included all persons with culture-confirmed TB residing in BC whose first specimen with *Mtb* isolated was received by the BCCDC PHL from 2005 through 2014 ( $n = 2,318$ ). *Mycobacterium africanum*, *Mycobacterium bovis*, and *Mycobacterium bovis* bacilli Calmette-Guérin (BCG) were excluded. For individuals with a recurrence during the study period, data from their first episode only was used if isolates from their first and second episode had matching MIRU-VNTR patterns ( $n = 11$ ), and data from both episodes where MIRU-VNTR indicated reinfection ( $n=2$ ).

Isolates lacking an amplicon peak at any locus were repeated with newly extracted DNA, and where there remained no peak at a single locus – excluding MIRU-VNTR loci 2163 and 2165, which are treated as absent when there is no amplification<sup>11</sup> – the locus was coded as missing data and included in the analyses ( $n = 93$ ). Of the 2,307 culture-positive isolates meeting study criteria (Figure 3), 17 isolates had incomplete MIRU-VNTR patterns or were unavailable for genotyping – leaving a total of 2,290 (99.2%) isolates which were successfully genotyped by 24-locus MIRU-VNTR using standard methods.<sup>3</sup>



\*First episode for a relapse patient was maintained in the study; relapse was defined as a subsequent episode with a genotype  $\leq 1$  MIRU loci different to the initial episode.

**Figure 3. Analytic Sample.** Selection of the analytic sample to examine the molecular epidemiology of tuberculosis in British Columbia, 2005–2014.



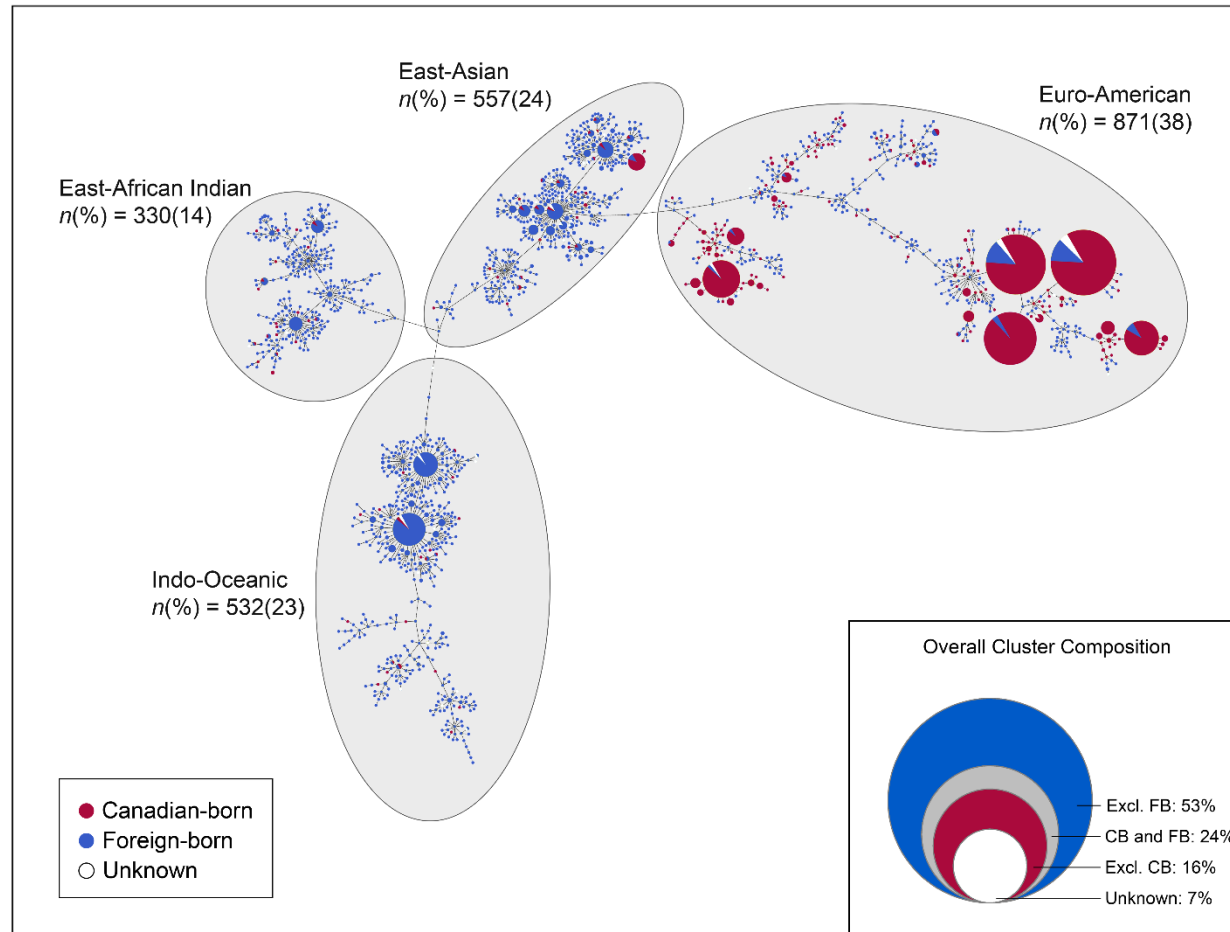
Isolates with an identical 24-locus MIRU-VNTR pattern were assigned an “MClust” number (a unique cluster ID) representing a unique genotypic cluster ( $\geq 2$  individuals). Large clusters ( $\geq 10$  cases) were analyzed to determine the predominant birthplace (Canada or Outside Canada), and were assigned as Canada where  $>50\%$  of persons in the cluster were born in Canada, otherwise the predominant birthplace was classified as Outside Canada. Cluster composition for each cluster in the study was categorized as: (i) exclusively Canadian-born, (ii) exclusively foreign-born, (iii) mixed Canadian- and foreign-born, or (iv) unknown. “Unknown” was defined as a cluster in which 1 or more individuals' birthplace was not known and the remaining clustered individuals were uniformly born in Canada or Outside Canada.

Individual-level clinical and demographic data were extracted from BCCDC's Integrated Provincial Health Information System (iPHIS). Community type was determined using the population density of the geographic service area in which each patient resided – urban ( $>40,000$ ), or rural ( $\leq 40,000$ ).

*Mtb* can be classified into seven major phylogeographic lineages reflective of the coevolution of tuberculosis and humans, and linked to ancient human migration patterns.<sup>12,13</sup> As a result, lineage information provides additional epidemiological information which contributes to the overall understanding of the *Mtb* population dynamics in a setting, and may contribute to case investigations – acting as an alert where lineage does not match what is expected based on a patient's demographics and travel history. Here, major lineage was predicted for each isolate based on MIRU-VNTR using TB-Insight's CBN method.<sup>14</sup> Phylogenetic relationships within each major lineage were visualized using a minimum-spanning tree (MST) in PHYLOViZ 2.0<sup>15</sup> and were coloured by birthplace (Figure 4).



## Tuberculosis Genotyping in British Columbia



**Figure 4.** Minimum spanning tree analysis of 24-locus MIRU-VNTR genotyping for *Mycobacterium tuberculosis* isolates, British Columbia (2005–2014). The size of each circle is proportional to the number of isolates. Classification of strains by birthplace is visualized by color coding. The inset demonstrates overall cluster composition with respect to birthplace; relative frequency of clusters that were exclusively Canadian-born (Excl.CB), exclusively foreign-born (Excl. FB), Canadian- and foreign-born (CB and FB), or where there were cases in a cluster with only CB or FB identified in addition to  $\geq 1$  case of unknown birthplace. \*Percentages have been rounded and may not total to 100%.



## TUBERCULOSIS GENOTYPING IN BRITISH COLUMBIA, 2005-2014

2012 **GOAL**

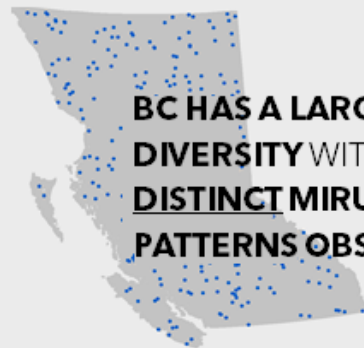


TB INCIDENCE  
IN **10-YEARS**

GENOTYPING SUPPORTS THIS GOAL BY **INFORMING** CONTACT INVESTIGATIONS OF TB STRAIN **MATCHES**

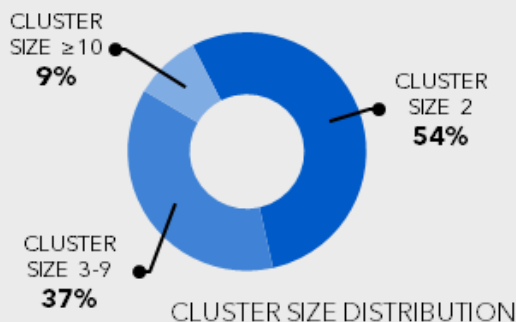


OF **2,290** ISOLATES  
GENOTYPED **42%**  
CLUSTERED, FOR A  
RECENT TRANSMISSION  
ESTIMATE OF **34%**



BC HAS A LARGE STRAIN  
DIVERSITY WITH **1,508**  
**DISTINCT MIRU-VNTR**  
PATTERNS OBSERVED

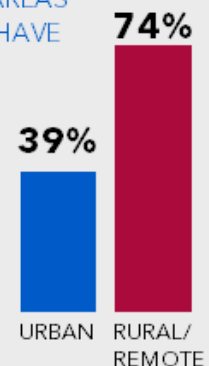
BC'S LARGEST  
CLUSTER HAS **70**  
PERSONS



PATIENTS RESIDING IN  
**RURAL** and **REMOTE** AREAS  
ARE MORE LIKELY TO HAVE

**CLUSTERED**

STRAINS





---

# Large MIRU-VNTR Cluster Summaries

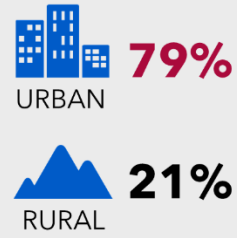
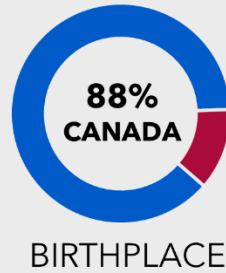
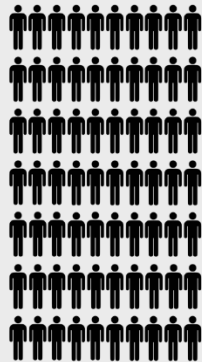
---



# MIRU-VNTR CLUSTER SUMMARY

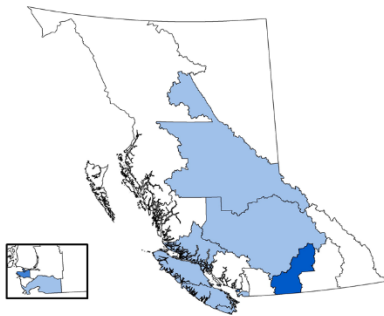
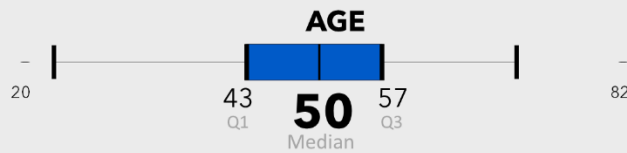
## MCLUST-002

CLUSTER  
OF **70**  
PERSONS



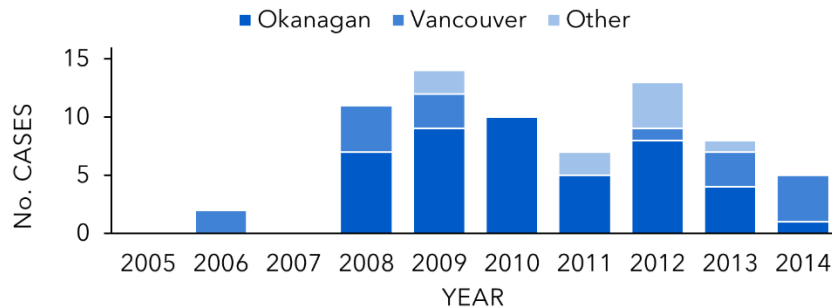
FEMALE

MALE



**EAm** **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**10** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, SK, MB, ON**

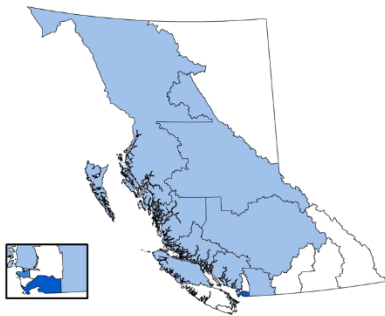
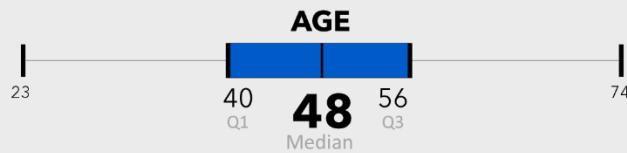
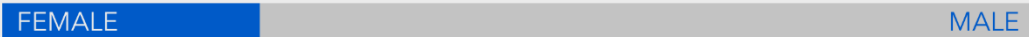
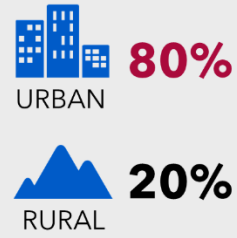
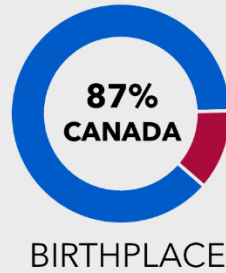
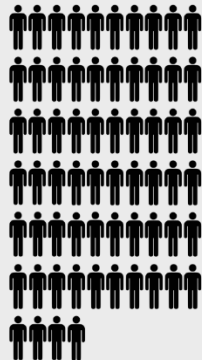




# MIRU-VNTR CLUSTER SUMMARY

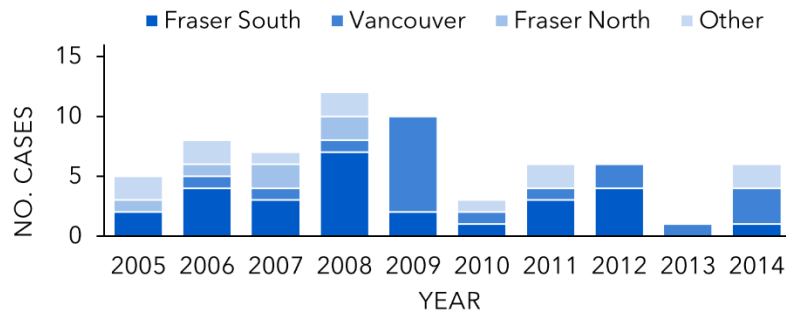
## MCLUST-012

CLUSTER  
OF **64**  
PERSONS



**EAm** **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **33** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, SK, MB, ON, NWT, YT**

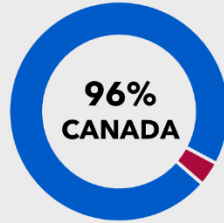
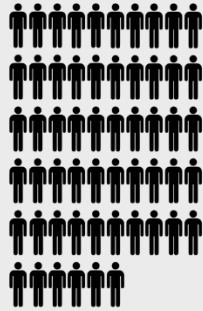




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-001

CLUSTER  
OF **56**  
PERSONS



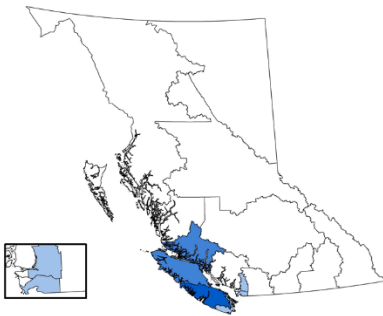
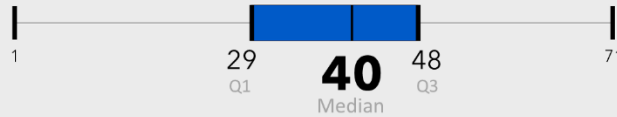
BIRTHPLACE



FEMALE

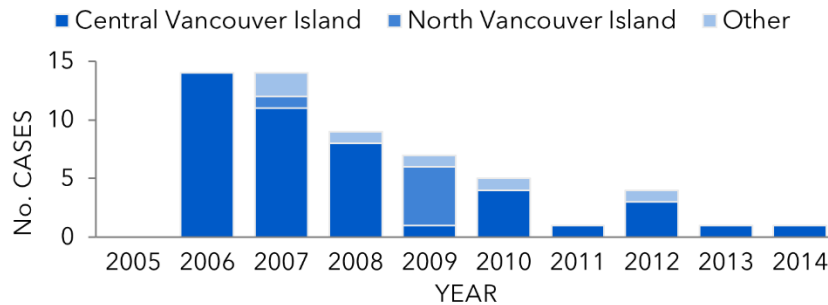
MALE

AGE



**EAm** EURO-AMERICAN LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **1** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB**

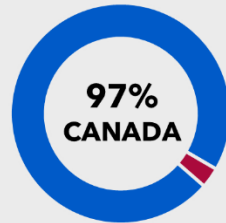
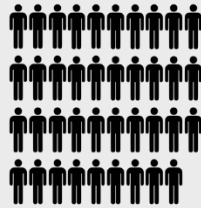




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-003

CLUSTER  
OF **39**  
PERSONS

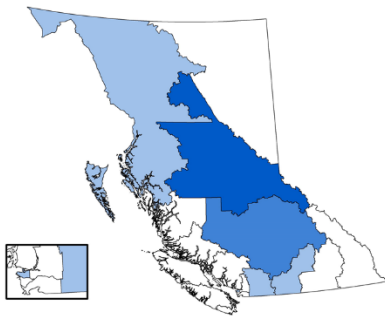
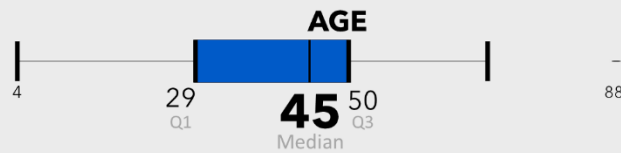


BIRTHPLACE



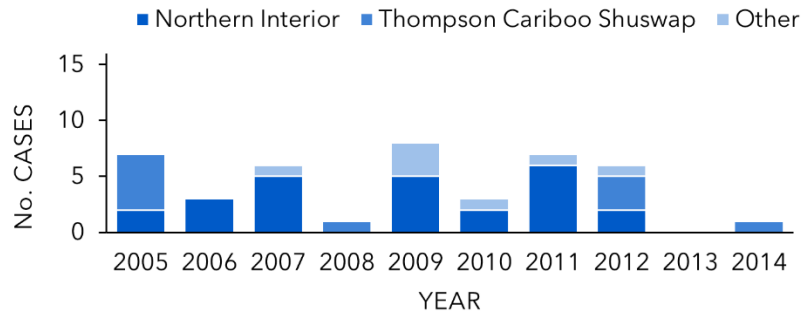
FEMALE

MALE



**EAm** **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **4 TIMES** IN THESE PROVINCES/TERRITORIES:  
**AB, MB**

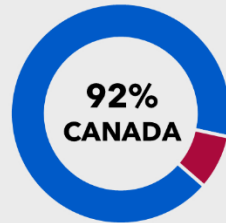




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-008

CLUSTER  
OF **36**  
PERSONS



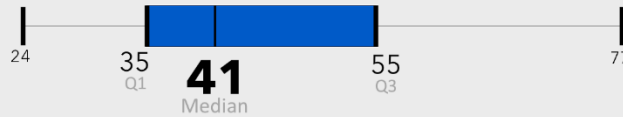
BIRTHPLACE



FEMALE

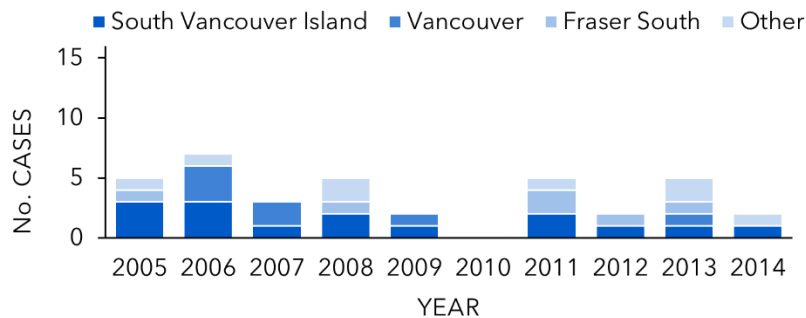
MALE

AGE



**EAm** **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **13** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, MB, YT**



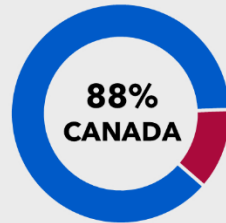




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-035

CLUSTER  
OF **17**  
PERSONS



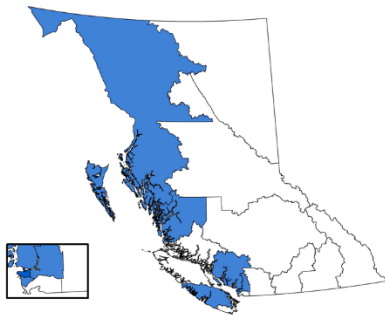
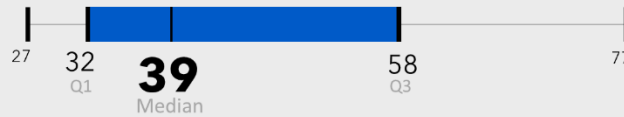
BIRTHPLACE



FEMALE

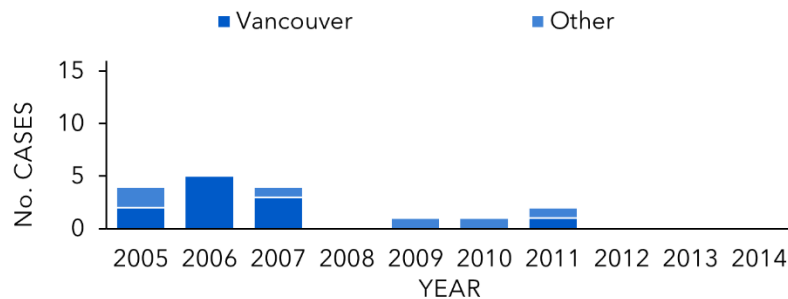
MALE

AGE



EA **EAST-ASIAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**2 TIMES** IN THESE PROVINCES/TERRITORIES:  
**AB, ON**

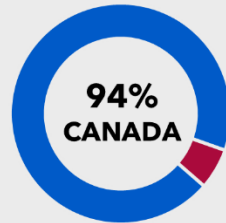




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-052

CLUSTER  
OF **17**  
PERSONS

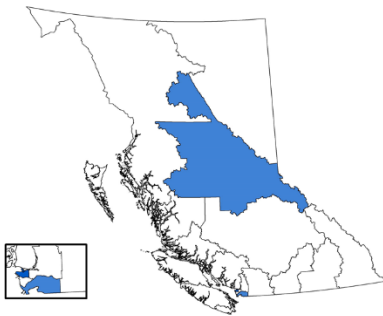
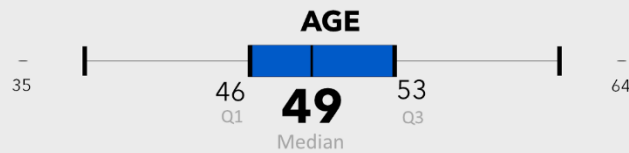


BIRTHPLACE



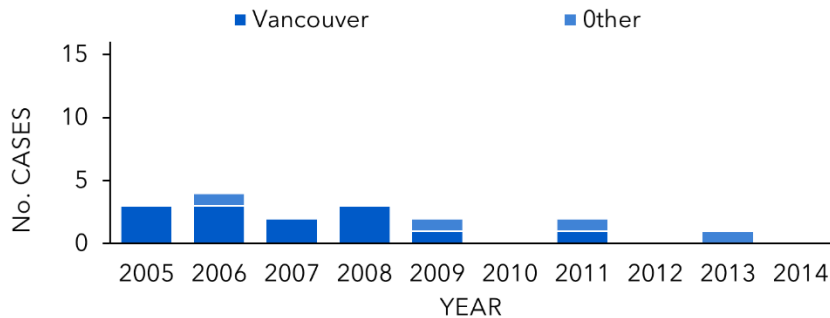
FEMALE

MALE



**EAm** EURO-AMERICAN LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**8** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, SK, MB, ON, YT**





# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-134

CLUSTER  
OF **13**  
PERSONS

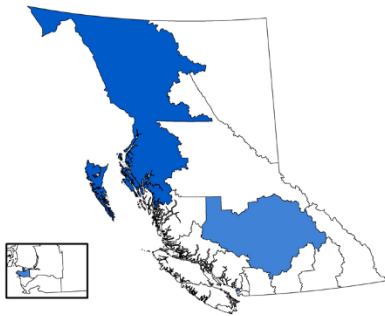
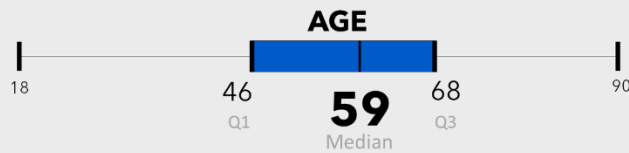


BIRTHPLACE



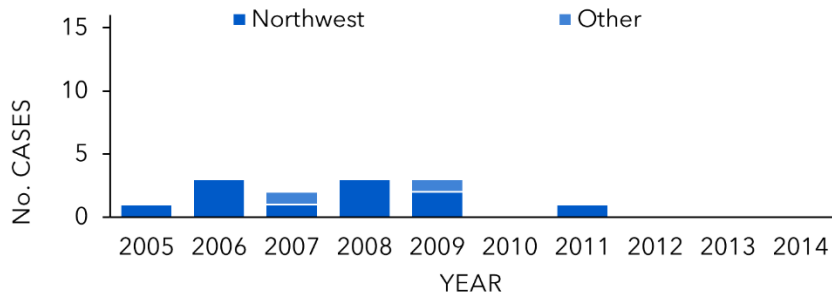
FEMALE

MALE



EAm **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**1** TIME IN THESE PROVINCES/TERRITORIES:  
**AB**





# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-055

CLUSTER  
OF **10**  
PERSONS



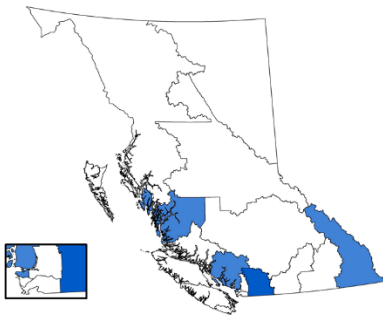
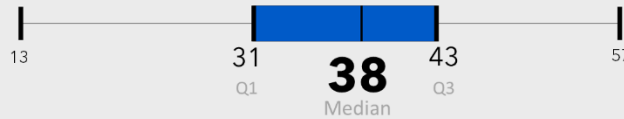
BIRTHPLACE



FEMALE

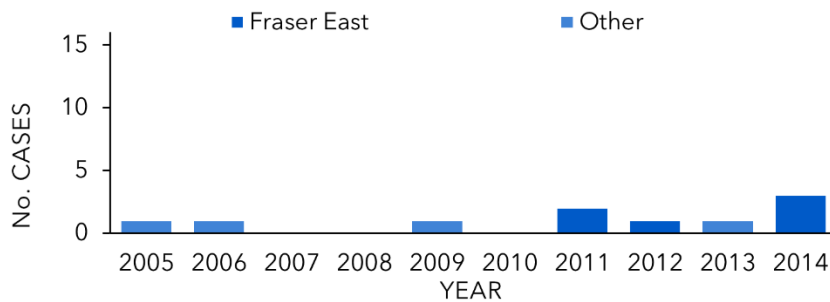
MALE

AGE



**EAm** **EURO-AMERICAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**4 TIMES** IN THESE PROVINCES/TERRITORIES:  
**AB, QC**





# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-011

CLUSTER  
OF **34**  
PERSONS

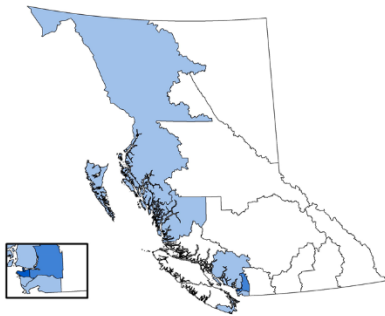
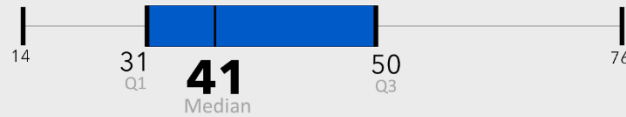


BIRTHPLACE

FEMALE

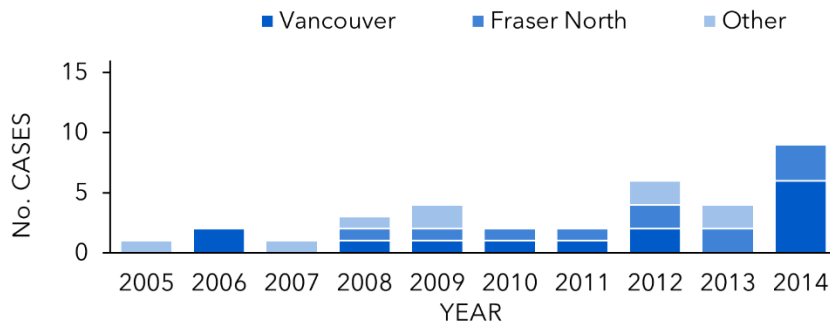
MALE

AGE



**IO** **INDO-OCEANIC** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**146** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, SK, MB, ON, QC**

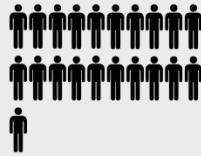




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-021

CLUSTER  
OF **21**  
PERSONS

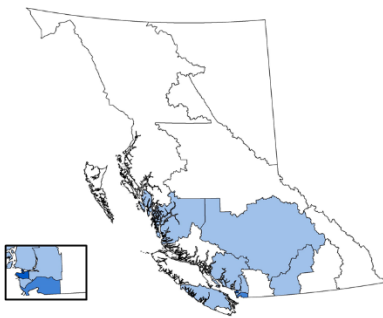
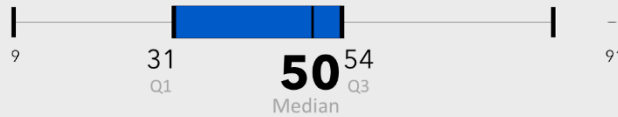


BIRTHPLACE

FEMALE

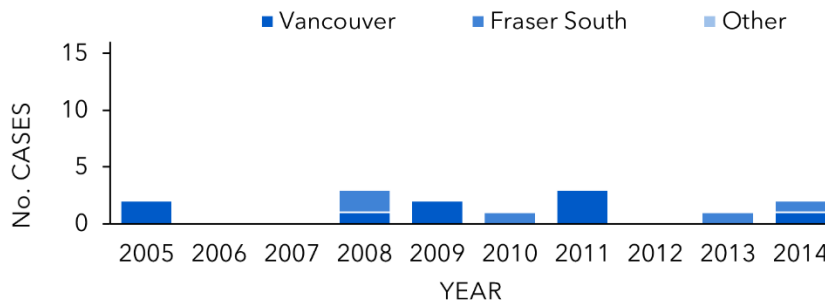
MALE

AGE



**IO** **INDO-OCEANIC** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**119** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, SK, MB, ON, QC**

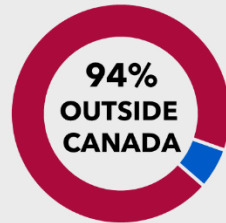




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-038

CLUSTER  
OF **16**  
PERSONS



BIRTHPLACE



**100%**

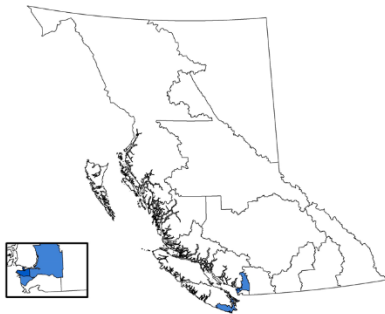
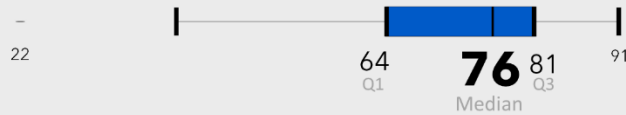


**0%**

FEMALE

MALE

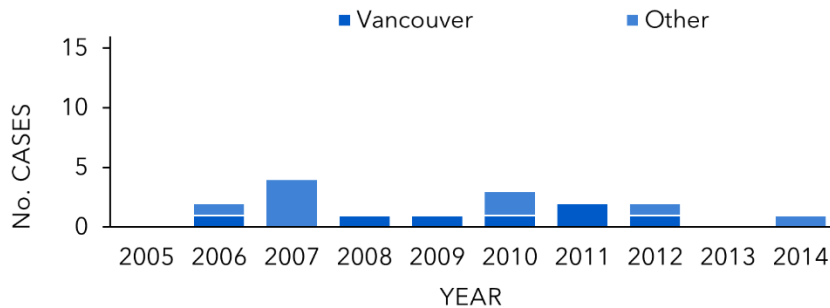
AGE



**EAST-ASIAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **48** TIMES IN THESE PROVINCES/TERRITORIES:

**AB, MB, ON, QC, NS**

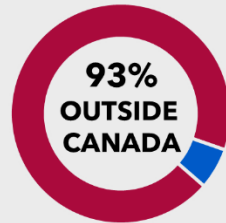




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-187

CLUSTER  
OF **16**  
PERSONS



BIRTHPLACE



**100%**

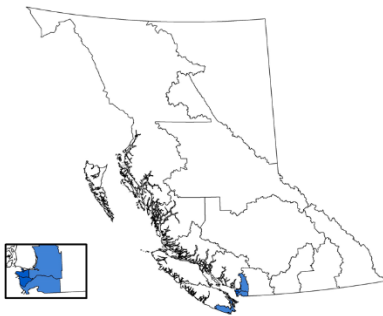


**0%**

FEMALE

MALE

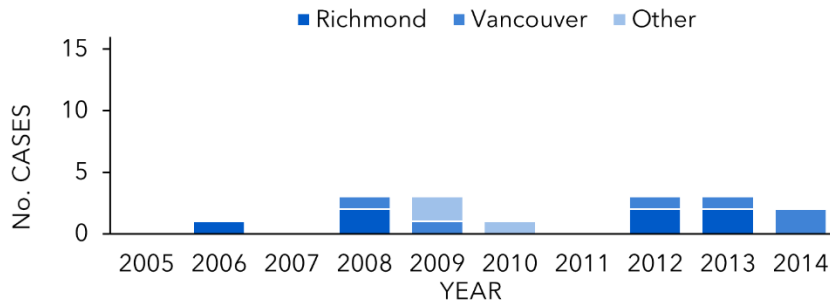
AGE



**EA** EAST-ASIAN LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST **65 TIMES** IN THESE PROVINCES/TERRITORIES:

**AB, SK, MB, ON, QC**







# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-149

CLUSTER  
OF **13**  
PERSONS



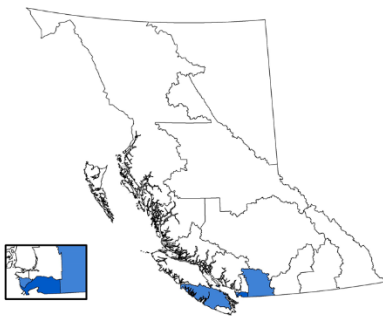
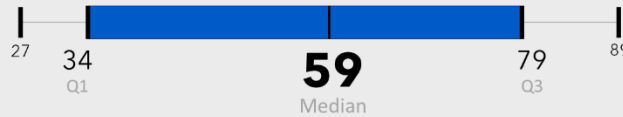
BIRTHPLACE



FEMALE

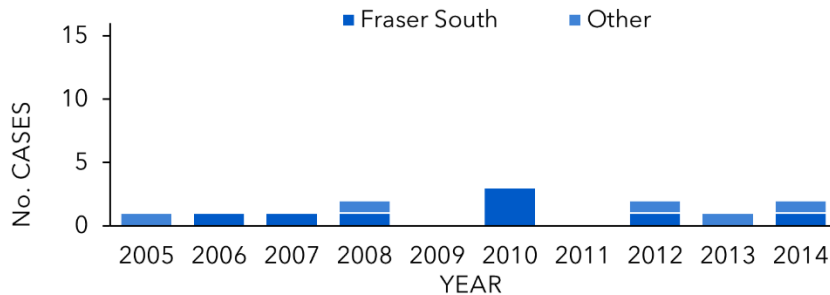
MALE

AGE



**EAI** EAST-AFRICAN INDIAN LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**22** TIMES IN THESE PROVINCES/TERRITORIES:  
**AB, ON, QC**

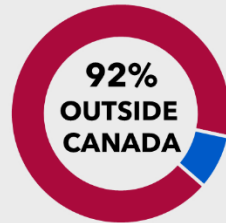




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-046

CLUSTER  
OF **12**  
PERSONS



BIRTHPLACE



URBAN

92%



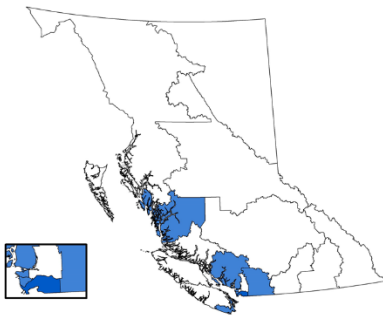
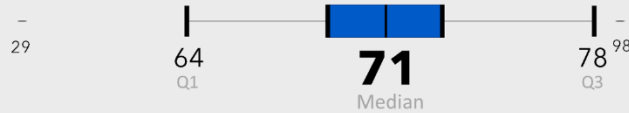
RURAL

8%

FEMALE

MALE

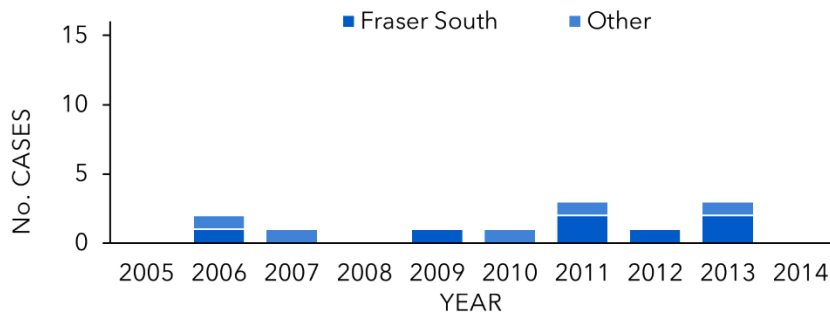
AGE



EAI EAST-AFRICAN INDIAN LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**37** TIMES IN THESE PROVINCES/TERRITORIES:

**AB, MB, ON, QC**

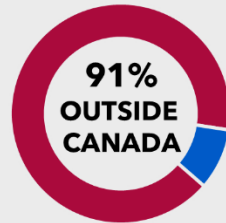




# MIRU-VNTR CLUSTER SUMMARY

## MCLUST-032

CLUSTER  
OF **11**  
PERSONS

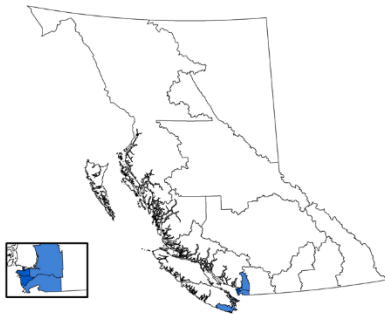
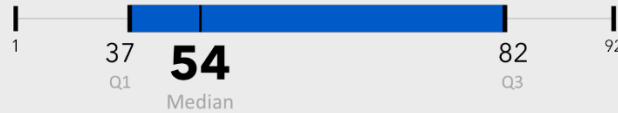


BIRTHPLACE

FEMALE

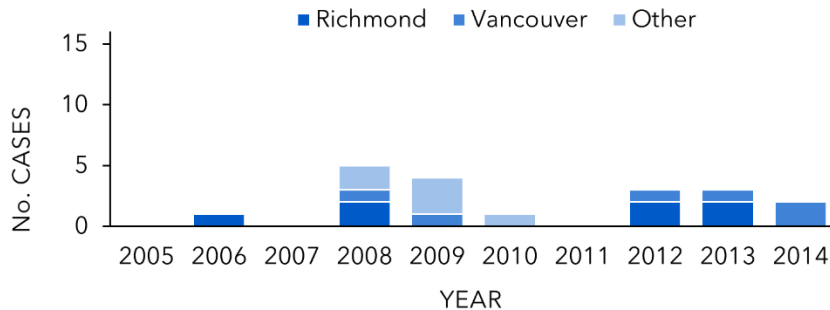
MALE

AGE



EA **EAST-ASIAN** LINEAGE

THIS GENOTYPE HAS BEEN SEEN AT LEAST  
**20 TIMES** IN THESE PROVINCES/TERRITORIES:  
**AB, ON, QC**





## References

1. 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 Jun 16;330(24):1703–9.
2. Guthrie JL, Kong C, Roth D, Jorgensen D, Rodrigues M, Hoang L, et al. Molecular Epidemiology of Tuberculosis in British Columbia, Canada: A 10-Year Retrospective Study. *Clin Infect Dis*. 2017 Oct 21;66(6):849–856.
3. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-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 Dec 1;44(12):4498–510.
4. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Loch C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol*. 2000 May 1;36(3):762–71.
5. Clark CM, Driver CR, Munsiff SS, Driscoll JR, Kreiswirth BN, Zhao B, et al. Universal Genotyping in Tuberculosis Control Program, New York City, 2001–2003. *Emerg Infect Dis*. 2006 May;12(5):719–24.
6. Lambregts-van Weezenbeek CSB, Sebek MMGG, van Gerven PJHJ, 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 Dec 1;7(12):S463–70.
7. Jamieson FB, Teatero S, Guthrie JL, Neemuchwala A, Fittipaldi N, Mehaffy C. Whole-genome sequencing of the *Mycobacterium tuberculosis* Manila sublineage results in less clustering and better resolution than mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping. *J Clin Microbiol*. 2014 Oct;52(10):3795–8.
8. Cheng JM, Hiscoe L, Pollock SL, Hasselback P, Gardy JL, Parker R. xA clonal outbreak of tuberculosis in a homeless population in the interior of British Columbia, Canada, 2008–2015. *Epidemiol Infect*. 2015 Nov;143(15):3220–3226.
9. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med*. 2011 Feb 24;364(8):730–9.
10. Mehaffy C, Guthrie JL, Alexander DC, Stuart R, Rea E, Jamieson FB. Marked microevolution of a unique *Mycobacterium tuberculosis* strain in 17 years of ongoing transmission in a high risk population. *PloS One*. 2014;9(11):e112928.



11. Menéndez MC, Buxton RS, Evans JT, Gascoyne-Binzi D, Barlow REL, Hinds J, et al. Genome analysis shows a common evolutionary origin for the dominant strains of *Mycobacterium tuberculosis* in a UK South Asian community. *Tuberculosis*. 2007 Sep 1;87(5):426–36.
12. Firdessa R, Berg S, Hailu E, Schelling E, Gumi B, Erenso G, et al. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg Infect Dis*. 2013 Mar;19(3):460–3.
13. Gagneux S. Host–pathogen coevolution in human tuberculosis. *Philos Trans R Soc Lond B Biol Sci*. 2012 Mar 19;367(1590):850–9.
14. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL, Yener B, et al. TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infect Genet Evol*. 2012 Jun;12(4):789–97.
15. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*. 2012;13:87.



Appendix I: 24-LOCUS MIRU-VNTR PATTERNS OF LARGE CLUSTERS

MClustID	MIRU 02	MIRU 04	MIRU 10	MIRU 16	MIRU 20	MIRU 23	MIRU 24	MIRU 26	MIRU 27	MIRU 31	MIRU 39	MIRU 40	424	577	1955	2163	2165	2347	2401	2461	3171	3690	4052	4156	MIRU (InternationalOrder)*
<b>MClust-002</b>	2	2	4	3	2	5	1	5	3	3	2	3	4	4	4	2	3	4	4	2	3	3	7	3	253433443433247252213423
<b>MClust-012</b>	2	2	5	3	2	5	1	5	3	3	2	3	2	3	3	3	3	4	4	2	3	3	5	3	253533233433335252213423
<b>MClust-001</b>	1	2	5	3	2	5	1	5	3	2	2	4	2	3	3	2	3	4	4	2	3	3	5	3	254532233433235152213423
<b>MClust-003</b>	2	3	4	3	2	5	1	5	3	3	2	3	4	4	1	4	4	4	2	2	3	3	5	2	353433444232415252213423
<b>MClust-008</b>	2	2	8	2	2	5	1	1	3	2	2	2	3	4	3	2	4	4	4	3	3	4	8	3	212822344443238252213433
<b>MClust-035</b>	2	2	3	3	2	5	1	6	3	5	3	3	2	4	5	5	4	4	4	2	3	3	8	2	263335244432558253213423
<b>MClust-052</b>	2	3	2	3	2	5	1	5	3	3	2	4	3	4	1	4	4	4	2	2	3	3	9	2	354233344232419252213423
<b>MClust-134</b>	2	2	8	2	2	5	1	1	3	2	2	1	3	4	3	2	4	4	4	2	3	3	8	3	211822344433238252213423
<b>MClust-055</b>	2	2	5	3	1	3	1	5	3	3	2	1	2	3	3	6	3	2	4	2	3	3	7	3	251533233433637232113223
<b>MClust-038</b>	2	2	3	3	2	5	1	7	3	5	3	3	4	4	5	5	4	4	4	2	3	3	8	2	273335444432558253213423
<b>MClust-187</b>	2	2	3	3	2	5	1	7	3	5	3	3	4	4	5	6	4	4	4	2	3	3	8	2	273335444432658253213423
<b>MClust-046</b>	2	2	6	4	2	5	1	7	3	4	2	3	4	2	4	2	4	4	4	2	3	4	7	4	273644424444247252213423
<b>MClust-149</b>	2	2	5	4	2	5	1	7	3	5	3	3	5	2	4	2	4	4	2	2	3	3	8	4	273545524234248253213423
<b>MClust-011</b>	2	5	4	3	2	6	2	2	3	4	3	2	1	4	10	8	4	3	2	6	3	2	7	1	5224341442218A7263223363
<b>MClust-021</b>	2	5	4	3	2	6	2	2	3	4	3	2	1	4	10	9	4	3	2	6	3	2	7	1	5224341442219A7263223363
<b>MClust-032</b>	2	2	2	3	2	5	1	7	3	5	4	3	4	4	5	6	4	4	4	2	3	3	8	2	273235444432658254213423

\*International order of loci: MIRU 04, MIRU 26, MIRU 40, MIRU 10, MIRU 16, MIRU 31, 424, 577, 2165, 2401, 3690, 4156, 2163, 1955, 4052, MIRU 02, MIRU 23, MIRU 39, MIRU 20, MIRU 24, MIRU 27, 2347, 2461, 3171



## Appendix II: MIRU-VNTR ALIASES

Locus	Alias1	Alias2	12-locus	15-locus	24-locus
<b>154</b>	MIRU 02		x		x
<b>424</b>	Mtub04			x	x
<b>577</b>	ETRC			x	x
<b>580</b>	MIRU 04	ETRD	x	x	x
<b>802</b>	MIRU 40		x	x	x
<b>960</b>	MIRU 10		x	x	x
<b>1644</b>	MIRU 16		x	x	x
<b>1955</b>	Mtub21			x	x
<b>2059</b>	MIRU 20		x		x
<b>2163b</b>	QUB11b			x	x
<b>2165</b>	ETRA			x	x
<b>2347</b>	Mtub29				x
<b>2401</b>	Mtub30			x	x
<b>2461</b>	ETRB				x
<b>2531</b>	MIRU 23		x		x
<b>2687</b>	MIRU 24		x		x
<b>2996</b>	MIRU 26		x	x	x
<b>3007</b>	MIRU 27	QUB5	x		x
<b>3171</b>	Mtub34				x
<b>3192</b>	MIRU 31	ETRE	x	x	x
<b>3690</b>	Mtub39			x	x
<b>4052</b>	QUB26			x	x
<b>4156</b>	QUB4156			x	x
<b>4348</b>	MIRU 39		x		x