



UNSW
THE UNIVERSITY OF NEW SOUTH WALES

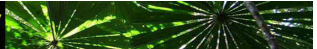
Estimates for the mutation rates of spoligotypes and VNTR types of *Mycobacterium tuberculosis*

Josephine F. Reyes and Mark M. Tanaka

November 30, 2009

School of
Biotechnology and Biomolecular Sciences

Evolution & Ecology
Research Centre



Understanding diversity in bacterial pathogens

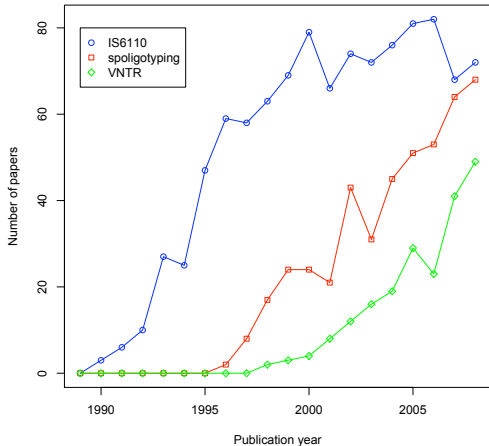
- How did transmission (infection) occur among individuals?
- What are the relationships among strains infecting different individuals?
- How do molecular markers change?
- At what rate do these changes occur?

Mycobacterium tuberculosis

- *M. tuberculosis*, causative bacterium of TB
- Tuberculosis (TB) disease: kills a person every 15 seconds
- Growing number of studies on its molecular epidemiology

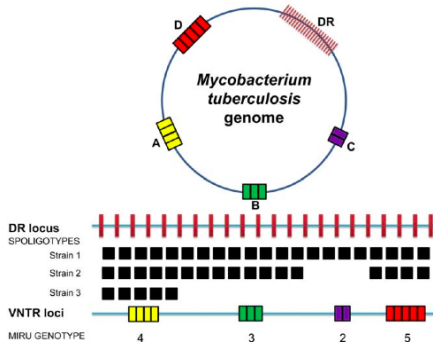


Studies on molecular epidemiology of TB (PubMed search results)

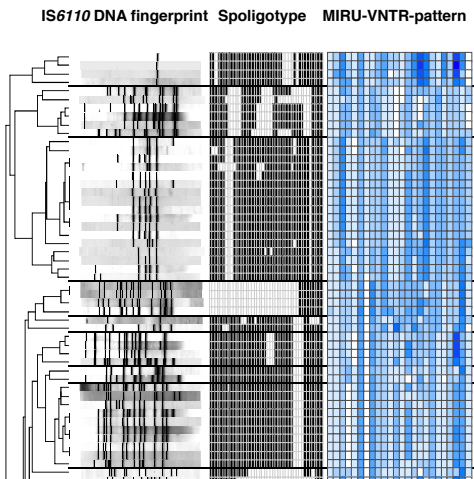


MTB genome

VNTR - variable numbers of tandem repeats
spoligotypes - spacer oligonucleotide typing

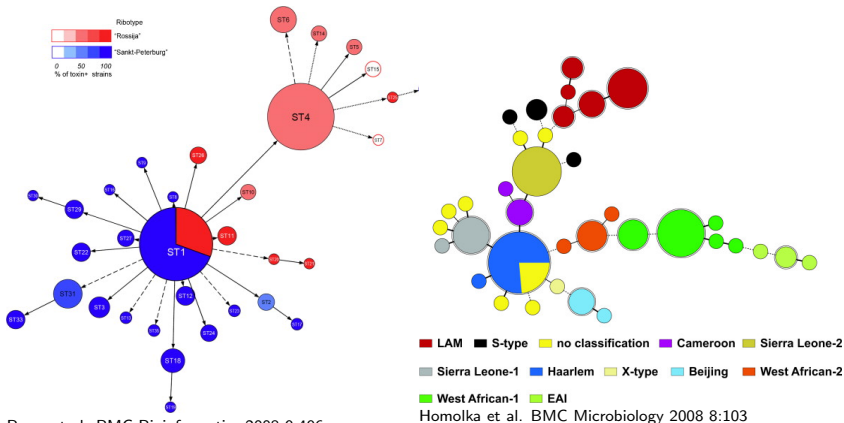


Molecular markers for TB: VNTRs and spoligotypes

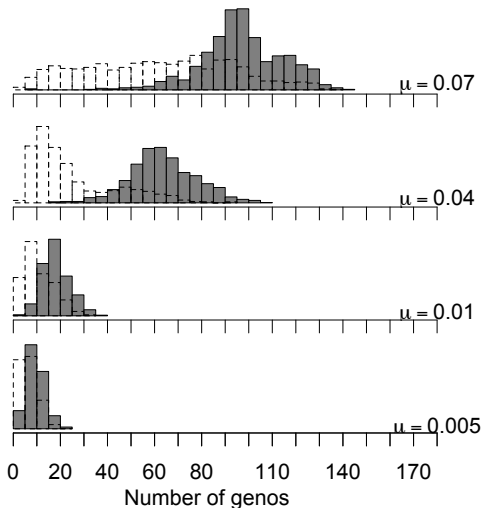


Homolka et al, BMC Microbiology 2008 8:103 doi:10.1186/1471-2180-8-103

Spoligoforests and minimum spanning trees



Application: Analysis of homoplasy



Mutation rates

- Mutation rate is the rate at which mutations appear within the *M. tuberculosis* population in patients and reach fixation (events per case per year)
- Mutation plays a large role in diversity of molecular markers
- Utility of molecular markers in epidemiology are enhanced by knowledge of mutation rates (and factors affecting these rates), mutation mechanisms

Table: Mutation rate of a VNTR locus per year

Organism	Reference	Point estimate	No. of loci used
<i>M. tuberculosis</i>	Grant et al 2008	10^{-5}	12
<i>M. tuberculosis</i>	Wirth et al 2008	$10^{-3.9}$	24 (5)
<i>Y. pestis</i>	Vogler et al 2007	0.098	14
<i>E. coli</i>	Vogler et al 2006	0.626	25

Table: Mutation rate of a spoligotype per year

Reference	Point estimate	(S.E.)
(spol-L09-C) Luciani et al 2009 (Cuba)	0.0133	0.009000
(spol-L09-V) Luciani et al 2009 (Venez)	0.0029	0.001975

Methods

Diversity parameter involving μ

$$\theta = 2N_e\mu$$

Point estimate for the mutation rate of a molecular marker (MM)

$$\begin{aligned}\hat{\mu}_{\text{MM}} &= \frac{\hat{\theta}_{\text{MM}}}{\hat{\theta}_{\text{reference}}} \times \hat{\mu}_{\text{reference}} \\ &= \frac{\hat{\mu}_{\text{MM}}}{\hat{\mu}_{\text{reference}}} \times \hat{\mu}_{\text{reference}}\end{aligned}$$

Reference estimates for the mutation rate
of *IS6110* types (used as $\hat{\mu}_{\text{reference}}$)

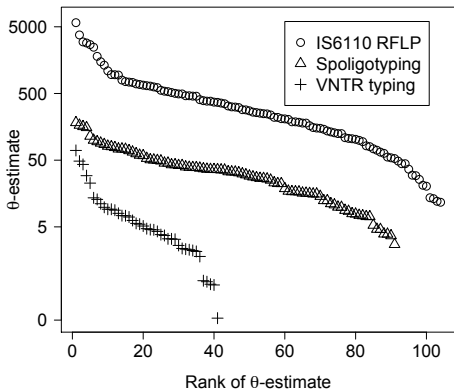
Reference	Point estimate	S.E.
(IS-TR01) Tanaka & Rosenberg 2001	0.1390	0.02308
(IS-R03) Rosenberg et al 2003	0.2870	0.10300
(IS-dB99) de Boer et al 1999	0.2166	0.04786
(IS-W02) Warren et al 2002	0.0793	0.00649
(IS-L09-C) Luciani et al 2009 (Cuba)	0.0685	(0.03545)
(IS-L09-E) Luciani et al 2009 (Estonia)	0.0882	(0.04240)

Methods (continued)

- Collect data sets with isolates typed using either:
 - IS6110 and VNTR typing
 - spoligotyping and VNTR typing
 - IS6110 and spoligotyping
- Estimate θ for each data set of sample size n and number of molecular marker types (genotypes) g using

$$g = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i}$$

Values of $\hat{\theta}$ from available data sets



Among these data sets, 25 are typed by both IS6110 and spoligotyping, 19 by both VNTR and IS6110, and 27 by both VNTR and spoligotyping.

Methods (continued)

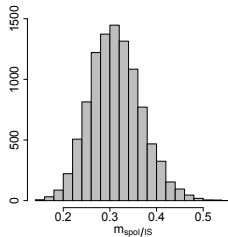
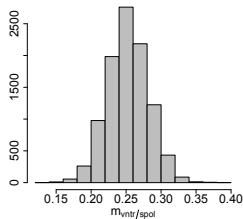
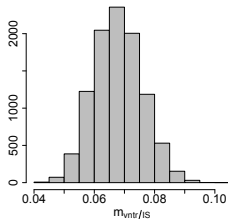
$$\hat{\mu}_{MM} = \frac{\hat{\theta}_{MM}}{\hat{\theta}_{\text{reference}}} \times \hat{\mu}_{\text{reference}}$$

- Compute $m = \frac{\hat{\theta}_{MM}}{\hat{\theta}_{\text{reference}}}$
- Estimate standard error for m (by bootstrapping)
- Find a confidence interval around $\hat{\mu}_{MM}$ using

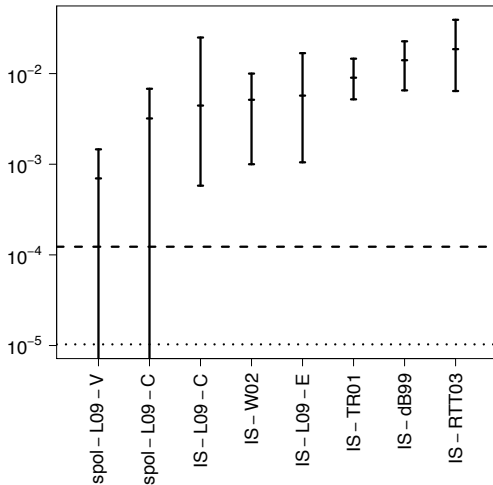
$$p(u) = \int_{\Omega} g(z)h(u/z)\frac{1}{z}dz$$

where $z = m$, $u/z = \hat{\mu}_{MM}/m$ with densities g and h , and Ω is the support for z

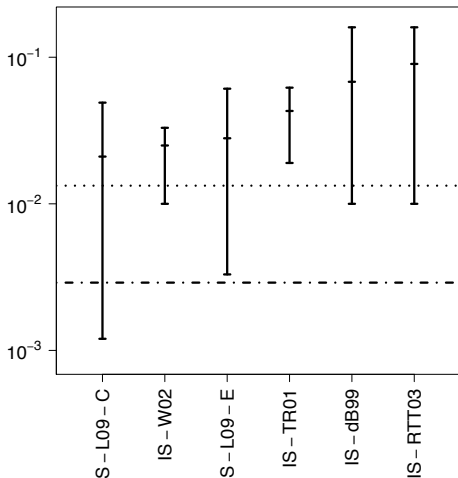
Results: relative mutation rates



Results: new estimates for the mutation rate of a VNTR locus (per year)



Results: new estimates for the mutation rate of a spoligotype (per year)



Summary

- The following issues should be addressed when studying the mutation rate of molecular markers for pathogenic bacteria
 - epidemiologically-linked isolates
 - accounting for the abundance/frequency of isolates
- We found evidence for
 - higher mutation rate estimates of a VNTR locus (at least 10-fold higher)
 - Mutation rate estimates for spoligotypes overlap with previously reported rates

Future directions

- Use approximate Bayesian computation to find the mutation rate of a VNTR locus
- Compute mutation rates of other genetic markers in other bacterial organisms

Some thank yous

- Richard Zach Aandahl (UNSW School of Maths and Stats)
- Ruiting Lan (UNSW BABS)
- Andrew Francis (University of Western Sydney, School of Computing and Maths)
- Fabio Luciani (UNSW School of Medical Sciences)
- UNSW University International Postgraduate Award (UIPA)
-  GRADUATE RESEARCH SCHOOL
Postgraduate Research Student Support (PRSS) Scheme 2009