Overview: Molecular Epi

- Lecture 1
  - Genotyping methods:
    - ‘Typing’ vs. ‘branding’
    - The ‘best method’
    - Epidemiology validation of DNA data
- Lecture 2
  - Epidemiologic applications
    - Individual
    - Outbreak
    - Population

Centre international de TB McGill

McGill International TB Centre

Molecular Methods for Typing and Branding the Tubercle bacillus

Marcel A. Behr

Professor, McGill University

Director, McGill Int. TB Centre

marcel.behr@mcgill.ca

Why molecular epidemiology?

- We know that TB is spread between people by aerosols
  - No intermediate host, no arthropod vector, etc.
- We don't typically know:
  - Where, When, From whom?
  - Role of host, pathogen, environment factors?

- **Goal:** Use natural variability in bacterial DNA as tracking tool
  - DNA lineages as stand-alone markers
  - DNA patterns as comparative data

‘Branding’ vs. ‘Typing’

- Branding is top-down process
  - Genus, species, subspecies….
  - Genetic tools can define a bacterial lineage
  - Result is ‘stand-alone’, e.g. BCG, Beijing strain
- Typing is bottom-up process
  - Also called DNA fingerprinting
  - Tools provide patterns
  - Pattern has no stand-alone value
    - A is like B, different than C

Branding vs. Typing

‘This is a VW’

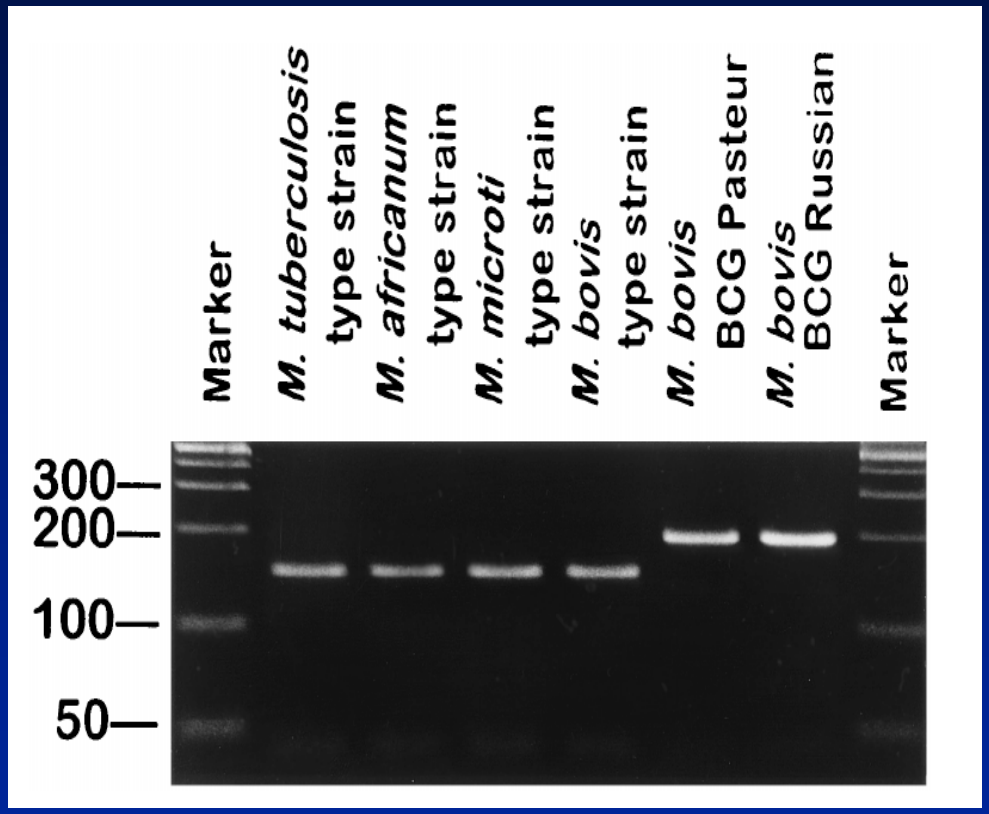
A

B

C

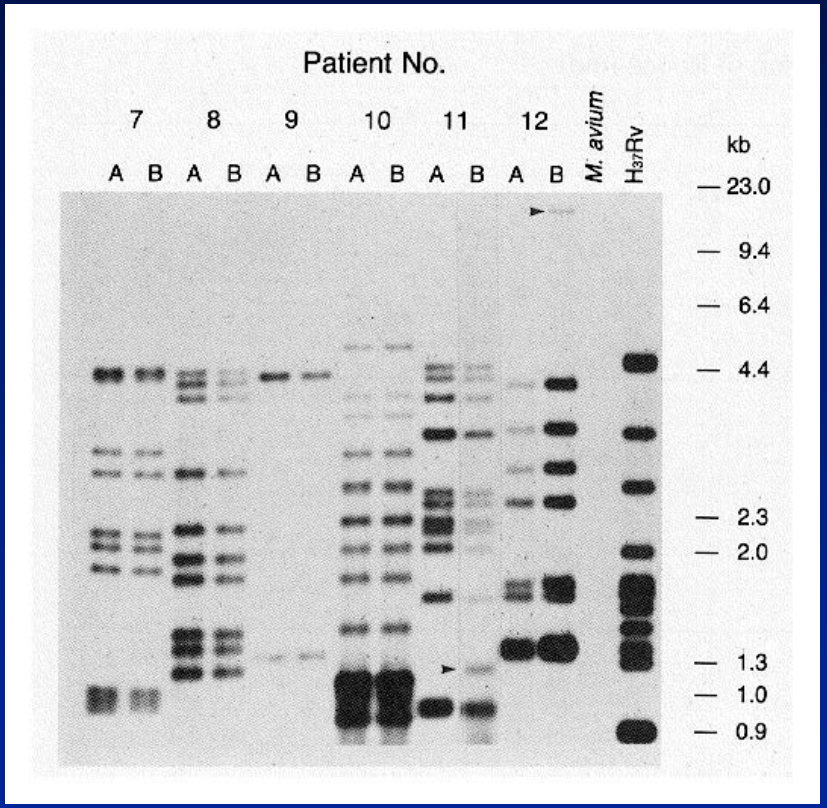
A,B are similar; C is different

Branding vs. Typing



3-primer PCR to detect BCG

Talbot et al., JCM, 1997



RFLP patterns for 6 patients

Small et al., NEJM, 1993

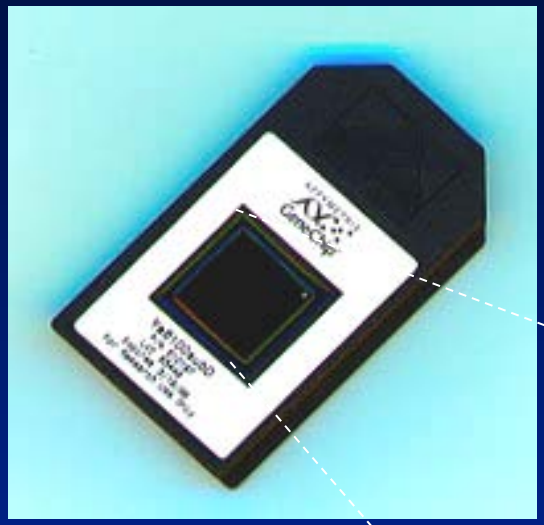
Branding & Typing Methods

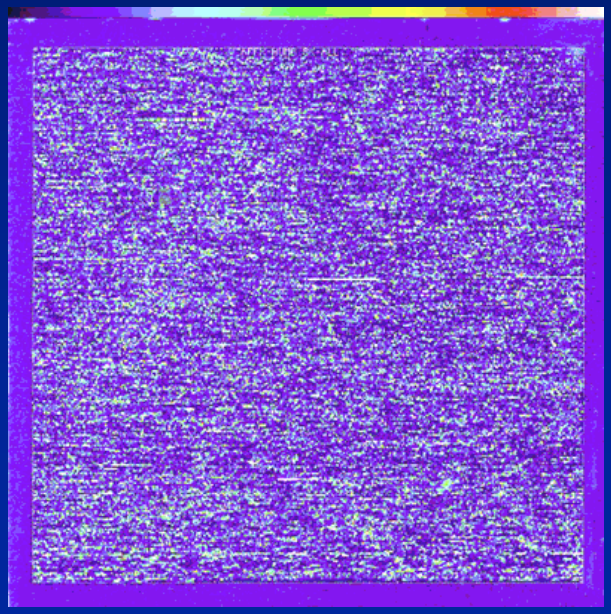
- **Branding:**
  - **Genomic deletions**
  - **Lineage-specific single nucleotide polymorphisms (SNPs)**
- **Typing:**
  - **Restriction fragment length polymorphism (RFLP)**
  - **Mycobacterial interspersed repetitive units (MIRU)**
  - **Spacer oligotyping (spoligotyping)**
- **New method for both:**
  - **Whole genome sequencing (WGS)**

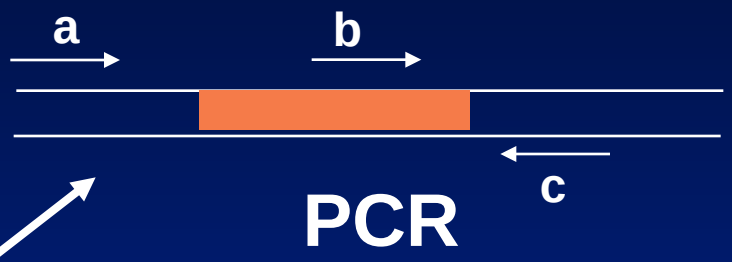
Branding & Typing Mixed Up

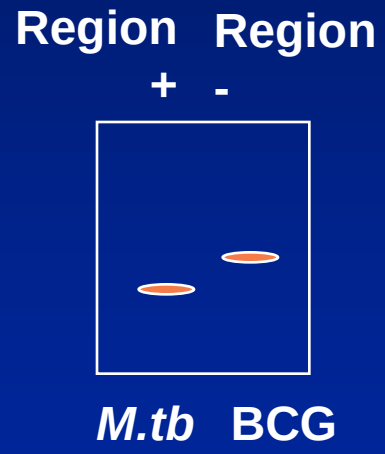
- Typing by branding:
  - Limited value,
    - All you get is ‘BCG’, no patterns to compare
  - If patient & contact have Beijing strain, can we infer that there has been transmission?
- Branding by typing:
  - Tempting but unreliable
  - “Spoligotype looks like Beijing strain”
    - Might not be - convergent patterns described
- Lesson:
  - Before doing study, ask whether you want to know lineages (branding) or to do comparisons (typing)

Branding method 1: Deleted regions detected by GeneChip





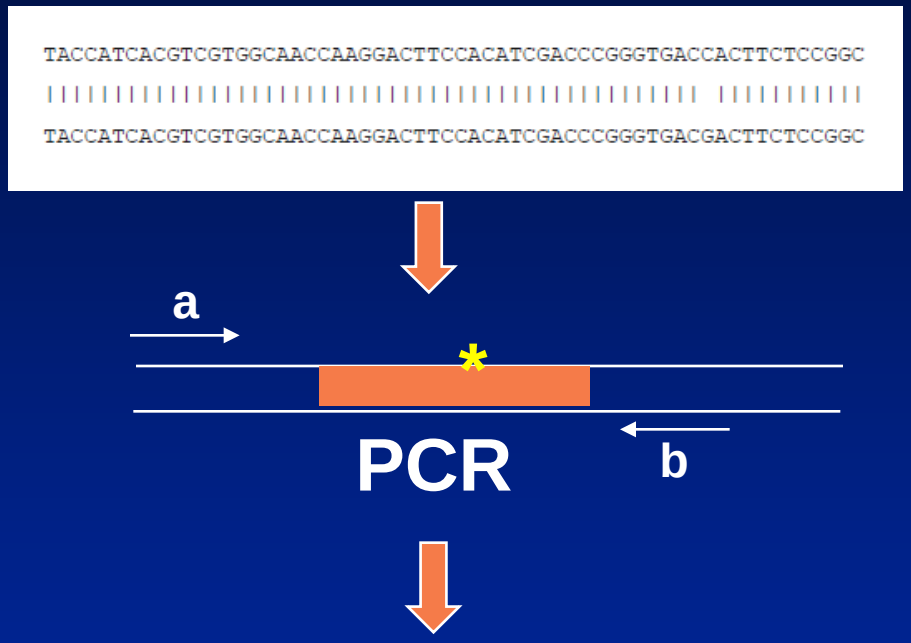




Sequence present: probe+
Sequence absent: probe-

Branding method 2: Single nucleotide polymorphisms (SNPs)

- Many possible methods
- Based on SNPs that define a *M.tb*. lineage
  - Can distinguish *M.tb*. from *M. bovis*
  - Can divide *M.tb*. into major groups
- Can be tailored to detect your local strain



- Send product for sequencing
- Do PCR with SNP-specific probes

Here: *M.tb* (top) or *M. bovis* (bottom)

Deletions vs. SNPs

- *M. tuberculosis* has a clonal population structure
  - Different markers give congruent information
- Choice of method is pragmatic
- Confirm your organism is BCG before doing an infection
  - Do deletion-based PCR
- Test 90 isolates for lineage
  - Suggest real-time PCR with 96-well plate, using probe for SNP

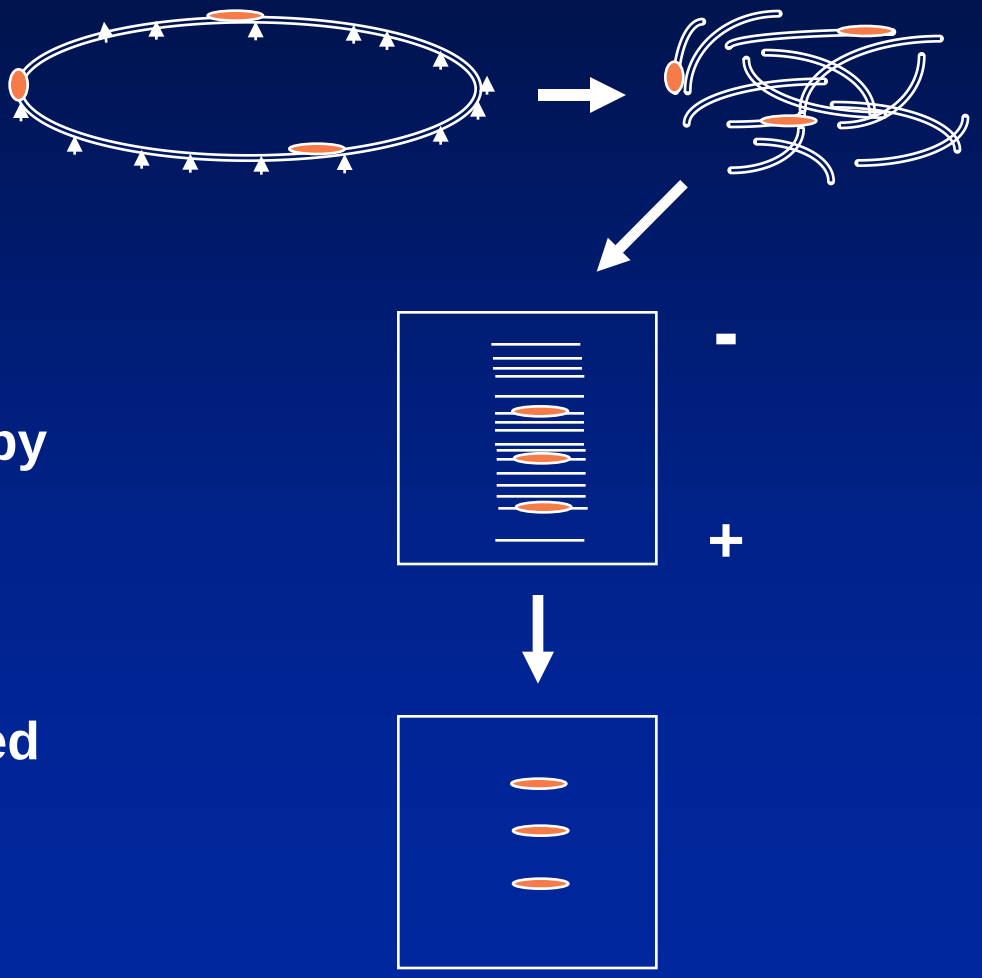
Typing method 1: IS*6110* RFLP

1. Chromosomal DNA
   - restriction site
   - IS*6110* site

2. DNA digested using *Pvu*II

3. Fragments separated by gel electrophoresis

4. Agarose blotted onto nitrocellulose and hybridization performed with labelled IS*6110*



Typing method 2: Mycobacterial interspersed repetitive units (MIRU)

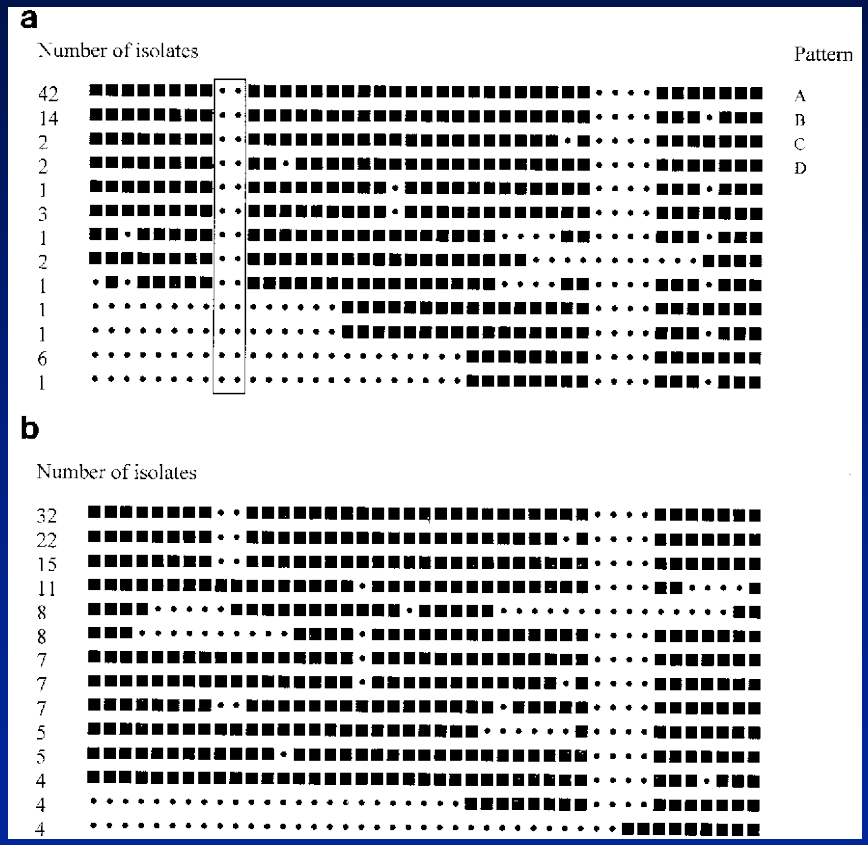
1. Multiple loci in genome have repeats
2. PCR ampilfies the region, but length varies
3. Products measured to enumerate number of repeats at each locus
4. Each isolate is assigned 24-digit code, e.g. 224325……

223425153422232324223352	
224325142324234534423463	
224325143324233534423473	Cluster-1
224325143324233534423473	
224325143324233534423473	
224325143324233534423473	
224325143324233534423473	
224325143324233534423473	
224325143324233534423473	
22432514332423453442–463	Cluster-2
2243251433242345344 23–63	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423463	
224325143324234534423473	
224325173324234434423573	
224325173324234534423473	

MIRU provided by LSPQ

Typing method 3: Spoligotyping

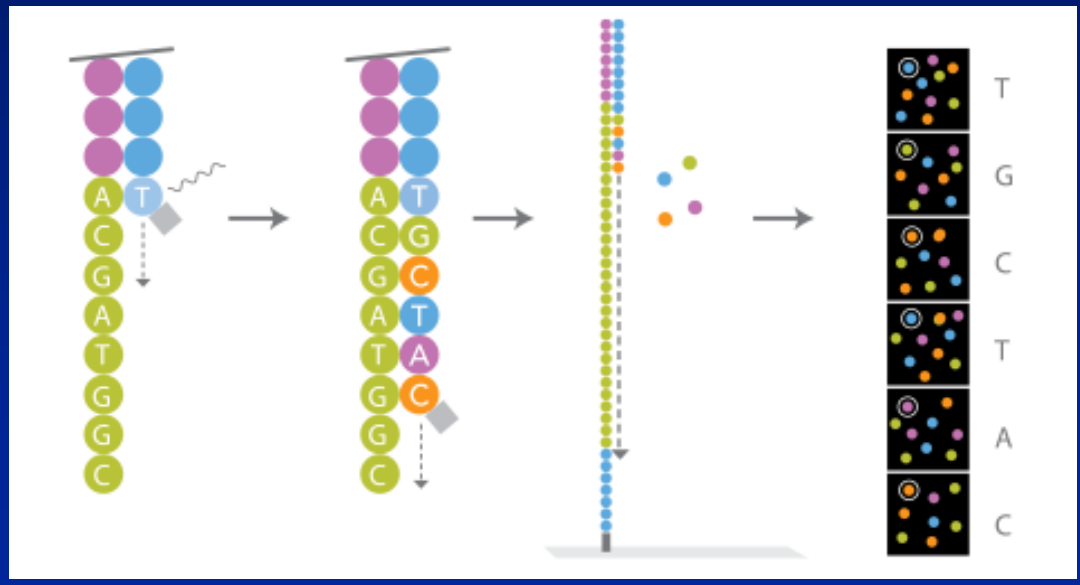
1. Direct repeat region of genome has repeats separated by ‘spacers’
2. Spacers variably present between isolates
3. PCR ampilfies across spacers, with conserved primers for repeats
4. Products hybridized to a membrane with spacers
5. Isolates coded for which spacers present, 110001...



Nguyen et al., JCM, 2003

Typing method 4: Whole genome sequencing (WGS)

1. Genome is 4.4 million base pairs
2. Variability can in theory happen anywhere in genome
3. Two randomly sequenced strains could have > 1000 SNPs
4. By sequencing whole genome, can find more variability than by selecting only portion of genome for typing

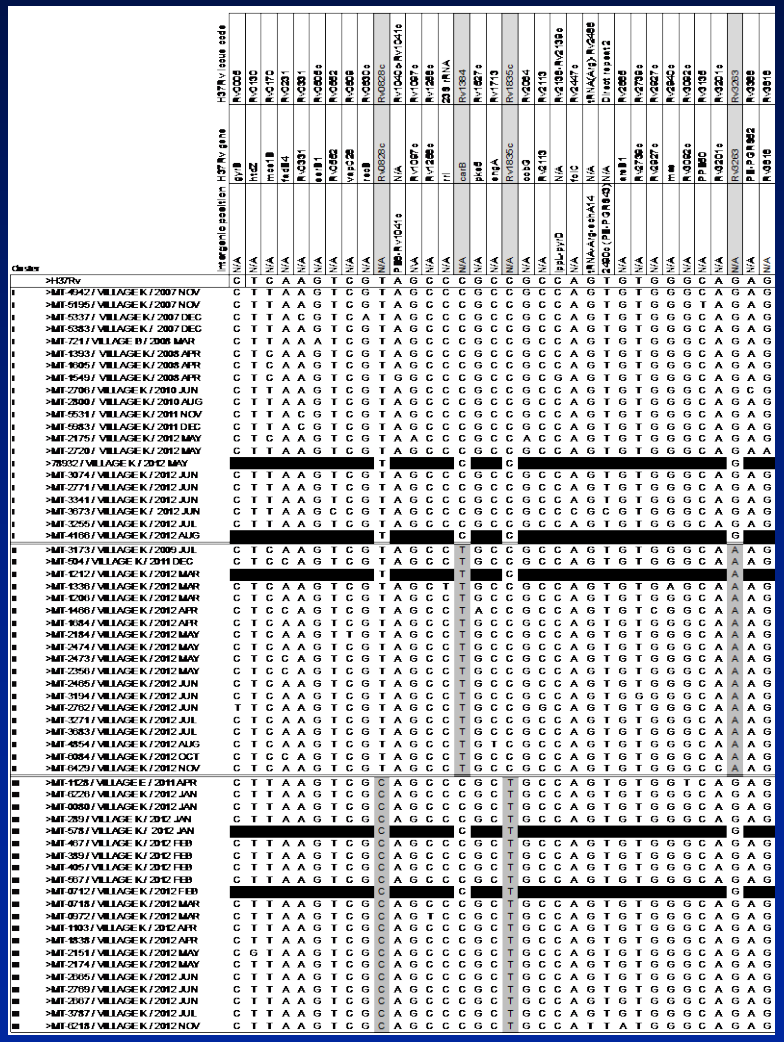


Whole genome sequencing by paired-end sequencing (MiSeq 250 Illumina)

http://www.dkfz.de/gpcf/illumina_hiseq_technology.html

Typing method 4: Whole genome sequencing (WGS)

1. Genome is 4.4 million base pairs
2. Variability can in theory happen anywhere in genome
3. Two randomly sequenced strains could have > 1000 SNPs
4. By sequencing whole genome, can find more variability than by selecting only portion of genome for typing



Lee et al., in submission

Choosing a typing method

- Most appropriate method depends on the question you are asking
  - How much transmission in my city?
  - Does the diagnostic lab have false-positives due to cross-contamination?
  - Are there different strains within a patient?
- Method should be chosen after you decide what you want to study
- Each method has advantages & disadvantages

Advantages / Disadvantages: IS*6110*-based RFLP

- **Advantages:**
  - **Standardized methodology**
  - **Widely used**
  - **High resolution for most strains of *M. tuberculosis***
- **Disadvantages:**
  - **Needs a lot of DNA, this requires culture**
  - **Cannot be done directly on specimen**
  - **Poor resolution for low-copy strains, defined as < 6 IS*6110* copies**

Advantages / Disadvantages: MIRU

- **Advantages:**
  - **Standardized methodology**
  - **Widely used**
  - **Global databases available for strain comparisons**
  - **Applicable to all strains of *M. tuberculosis***
- **Disadvantages:**
  - **Lower discrimination than RFLP**
  - **Tend to say there is more clustering**

Advantages / Disadvantages: Spoligotyping

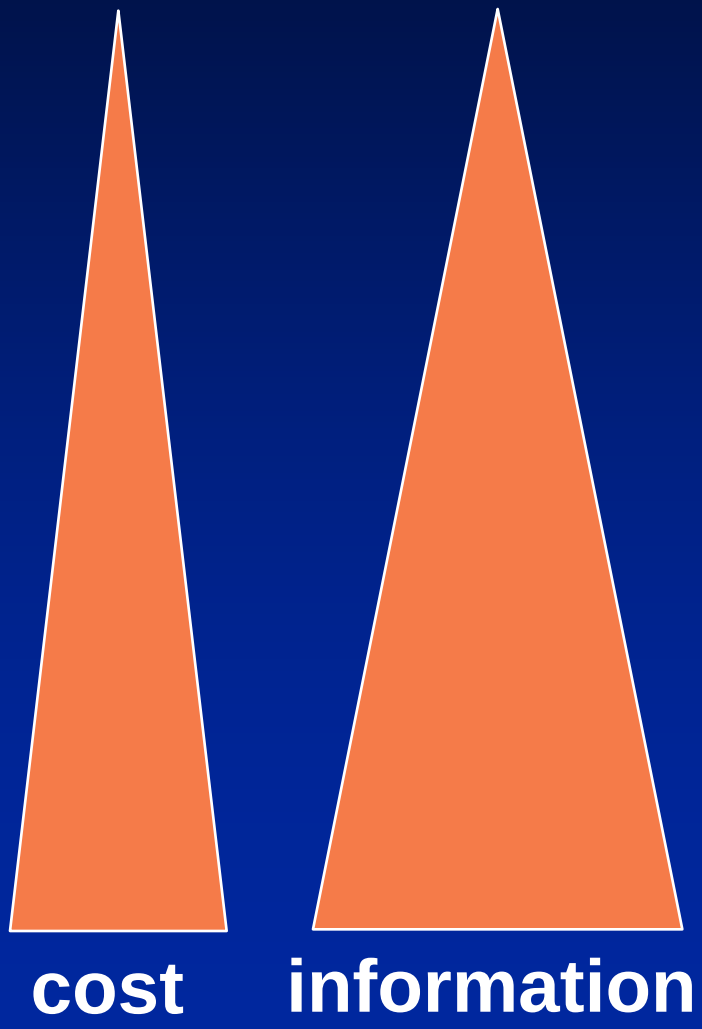
- Advantages:
  - Standardized methodology
  - Widely used
  - Global database available
  - Can be done in high throughput - inexpensive
- Disadvantages:
  - Very low resolution
  - Different patterns can exclude link
  - But same pattern may not be informative

Advantages / Disadvantages: Whole genome sequencing

- **Advantages:**
  - **Highest resolution**
  - **Allows branding AND typing**
  - **Can look for matching (A is like B) AND can infer directionality (A -> B)**
- **Disadvantages:**
  - **Most expensive**
  - **Most rely on a Genome Centre (not in-house)**
  - **Bioinformatics is very challenging**

Typing methods compared

- Spoligotyping
- MIRU/VNTR
- IS6110-based RFLP
- Whole genome sequencing



What is the ‘best method’?

- Best method depends on the question
- How much diversity do you expect?
- What is your pre-test probability of a given result?
- Same principles that apply to selecting a diagnostic test
  - Sensitivity of method to detect a true link
  - Specificity of method to exclude a false link

Typing methods and ‘the truth’

	Transmission	No transmission
Same pattern	‘Match’	
Different pattern		‘No match’

Typing methods and epi data

	Epi link	No epi link
Same pattern	Transmission	???
Different pattern	???	No transmission

Discordant cells:
Interpretation 1: Two methods do not agree
Interpretation 2: Two methods provide complementary information

Epi link, different DNA

- Small differences seen within patients and during outbreaks
  - Infer: Strain evolution
  - More evident with higher resolution of method
    - WGS > RFLP > MIRU > Spoligotype
- Major differences seen within patients and among close contacts
  - In individual: Infer exogenous reinfection
  - In pair: Infer independent infections

DNA match, no epi link

- Low resolution method:
  - May be shared ancestry, but not recent transmission
    - E.g. Quebec strain of *M. tuberculosis*
- High resolution method:
  - Lab cross-contamination
    - If patients unknown to each other, but samples processed at same place/time
  - Transmission
    - E.g. transmission among homeless
    - E.g. transmission by casual contacts

DNA/Epi discordance

- One approach:
  - DNA + epi: confirmed transmission
  - DNA alone: suspected transmission
  - My opinion: this introduces a bias, since the first and second group may not be the same
- Alternative approach:
  - If DNA matching is validated, i.e. not lab error, treat all those with matched strains the same
  - My opinion: if the new method always agrees with the old method, it doesn’t add any value

Genotyping methods

- Multiple techniques enable us to ‘brand’ or ‘type’ *M.tb.* isolates
- We can now ask questions at level of patient, outbreak, population
- Rate-limiting step is epidemiology
  - What is the question?
  - What is the quality of the epidemiologic data?
  - Can we link genotype with a medically or epidemiologically relevant outcome?
- Genotyping is not a goal; it is a tool to help better understand TB

Questions?

Centre international de TB McGill

McGill International TB Centre