Introduction to the TB genome

February 26, 2016

Jay Johnston MD MPH

- No conflicts of interest to declare

Outline

- The *M. tuberculosis* (MTB) genome
- Success and limitations of MTB genetics
- Future prospects with MTB genomics

M. tuberculosis

- *M. tuberculosis* belongs to MTB complex: *M. tuberculosis, africanum, bovis/BCG, cannetti, pinnipedii, microti, caprae*
- Predominately human pathogen
- Characterized by slow growth, complex cell wall, ‘dormancy’, intracellular pathogenesis

MTB evolution

- MTB has evolved in humans for millennia
- Initially thought to arise with animal domestication ~10,000 years ago
- Recent evidence from whole genome sequencing (WGS) points to human infection from 6,000-70,000 years ago[1,2]
- Generally 7 main lineages, each with a distinct geographic distribution[3]

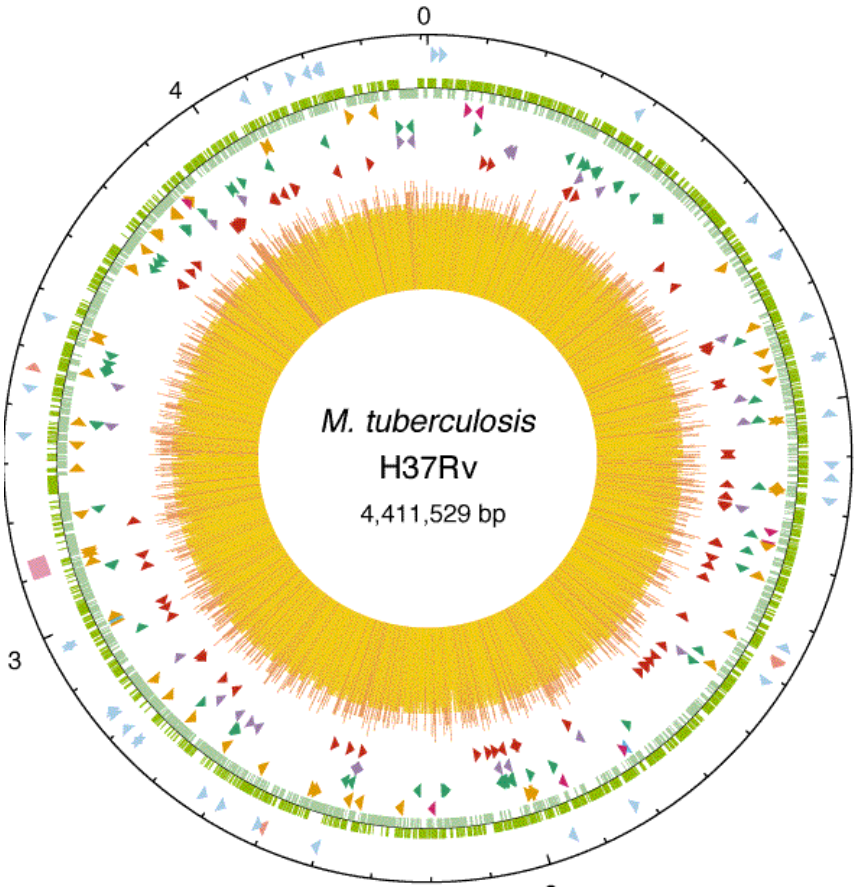
1. Comas *et al. Nat Gen* 2013; 1176-82
2. Bos *et al. Nature* 2014; 514:494-7
3. Galagan JE. *Nat Rev Gen* 2014; 15:307-20

M. Tuberculosis genetics

- MTB genetics well-studied for >30 years
- Publication of the first genome sequence of *M. tuberculosis* H37Rv was in 1998[6]
- At the time, the second largest microbial sequence available (*E. coli 4.6Mb*)

4. Cole *et al. Nature* 1998; 393:537-44.

Mtb genome



- 4.4M base pairs
- ~4000 genes
- Circular chromosome
- No plasmids or extrachromosomal elements
- Large proportion of genes devoted to lipid metabolism

4. Cole *et al. Nature* 1998; 393:537-44

MTB genome: evolution

- MTB evolution is mostly clonal, meaning sequence diversity is generated by SNPs (SNP=single nucleotide polymorphisms)
- Horizontal gene transfer not observed in short periods
- Over longer periods, large genetic shifts have developed through deletion:
  - *M. bovis*: ~66k fewer bp than MTB H37Rv

4. Cole *et al. Nature* 1998; 393:537-44

MTB genome: evolution

- Human MTB sequences highly conserved
- Average strains differ by about 1200 SNPs or 0.03% their genomes[1]
- SNP generation is relatively slow in most (90%) of the TB genome
- ~10% of the genome devoted PE/PPE family
- PE/PPE regions have multiple copies of repeat sequences and generate variability

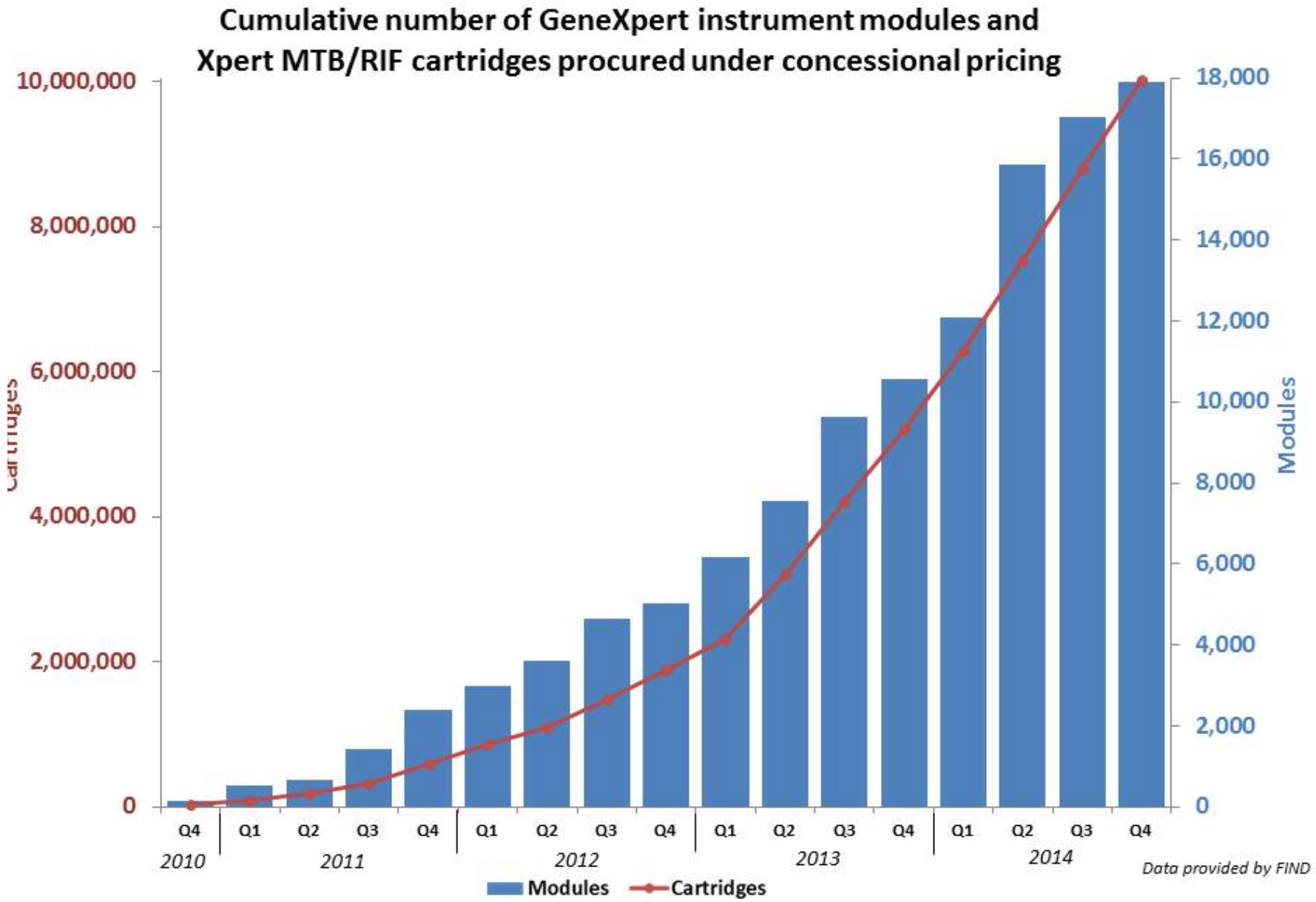
1. Comas *et al. Nat Gen* 2013; 1176-82
4. Cole *et al. Nature* 1998; 393:537-44

How have we benefitted from understanding the MTB genome?

Benefits: predicting resistance

- Mutations conferring INH & RIF resistance were characterized in the mid 1990s[5,6]
- Led to NAATs that lead to early/sensitive identification of MTB pre-culture
- Characterizing SNPs that confer drug resistance has enabled development of assays for TB resistance
- The most notable product to emerge from this has been the GeneXpert system

5. Bodmer *et al. J Antim Ch* 1995; 8: 496-514
6. Zhang *et al. Mol Micr* 1993; 8: 521-524



7. http://who.int/tb/laboratory/mtbrifrollout/en/

Benefits: immune assays

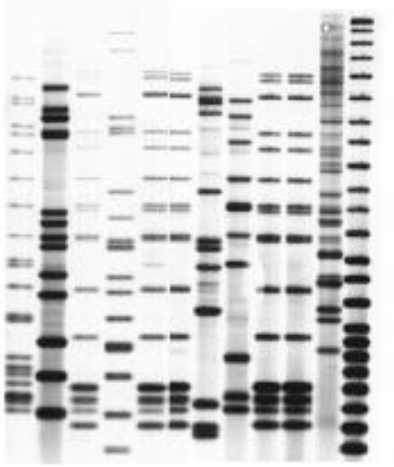
- Comparing *M.bovis*, *M. bovis BCG*, and *MTB H37Rv* identified putative 'virulence regions' present only in *M. tuberculosis*[9]
- The RD1 deletion region contains ESAT-6 and CFP-10, the proteins targeted in Interferon Gamma Immune Assays (IGRAs)

8. Behr et al. Science 1999; 284:1520-3

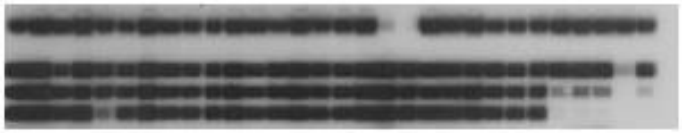
Limitations: diagnostics

- Limited strain typing beyond identifying *M. tuberculosis* and lineage
- Resistance profiling is quite limited
- Immunodiagnostics are still quite limited for active and latent MTB infection

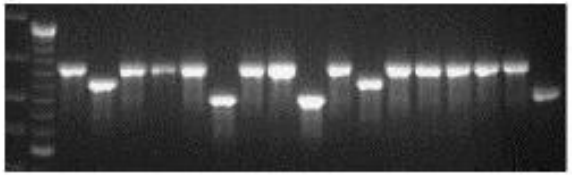
Benefits: MTB genotyping



RFLP



SPOLIGOTYPING



MIRU-VNTR

Benefits: MTB genotyping

- Demonstrates clusters of similar genotyping

  - Identical fingerprint = same strain (transmission)

  - Distinct fingerprint = distinct clone (no transmission)

Benefits: MTB genotyping

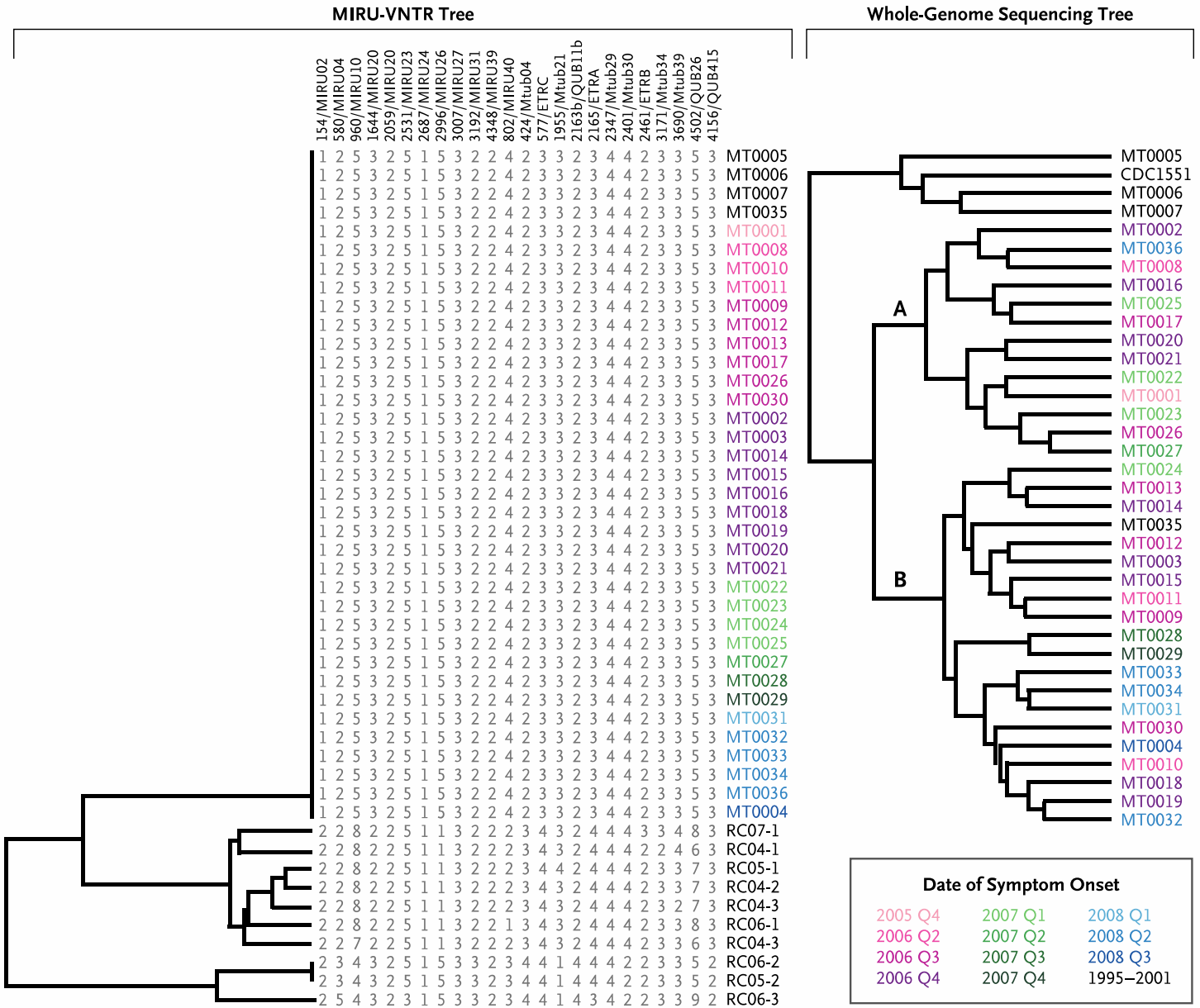
- Demonstrates clusters of similar genotyping
  - Identical fingerprint = same strain (transmission)
  - Distinct fingerprint = distinct clone (no transmission)
- Can discriminate between relapse/reinfection
  - Identical fingerprint = relapse
  - Distinct fingerprint = reinfection

Benefits: MTB genotyping

- Demonstrates clusters of similar genotyping
  - Identical fingerprint = same strain (transmission)
  - Distinct fingerprint = distinct clone (no transmission)
- Can discriminate between relapse/reinfection
  - Identical fingerprint = relapse
  - Distinct fingerprint = reinfection
- Detect laboratory cross-contamination…
  - Identical fingerprint = ?contamination
  - Distinct fingerprint = no contamination

Limitations: MTB genotyping

- Need a large amount of DNA (i.e. culture)
- Limited resolution in regions with low strain diversity (i.e. low incidence regions)
- Binary data provides no information on within-population transmission dynamics
- Only characterizes non-coding regions
- Limited insight into strain pathogenicity/resistance



Gardy et al. NEJM; 2011:364(8).

Whole genome sequencing

Whole genome sequencing

- Identifies numerous coding regions that may confer resistance/virulence
- This can lead to improved resistance assays and immunodiagnostics
- Individual SNP comparison enables us to understand MTB evolution over time
- Enables us to examine TB epidemiology in finer detail than traditional fingerprinting

Whole genome sequencing

- Still requires a relatively large amount of DNA, so culture required for sequencing
- Relies on short read technologies that are prone to error in repetitive regions (maybe the most useful region?)
- With longer reads, we could can span repeat regions and overcome these issues

MTB genomics

- Characterizing MTB genetics/genomics has led to huge leaps in MTB diagnostics, resistance testing and epidemiology
- Building on these successes, I expect that MTB WGS will transform TB diagnostics and public health practice.

Thank you

BCCDC- BCPHMRL Staff
UBC Respiratory Medicine
Jennifer Gardy PhD
Patrick Tang MD PhD
Victoria Cook MD