THE UNIVERSITY OF SYDNEY

CIDM centre for infectious diseases & microbiology Public Health

The first case of XDR TB in NSW: insights gained from whole genome sequencing

Alex Outhred

May 2013

Background

- Born in Ethiopia in late 1970s
- Ancestral country - Somalia
- Arrived in Australia as refugee in 2005
  - accepted in UNHCR "queue" for many years
  - mostly in Kenya, refugee camps
  - TB history major reason for delay
- Married in Australia, but separated
- One child under 2yrs
- Smoker

TB history

- Treated for TB in 1990
  - In Ethiopia
  - Complete disruption after 3 months due to war
- Treated for TB in 2000
  - In Kenya; smear positive.
  - Completed 6 months (supervised)
- Treated for TB in 2003
  - Supervision and duration unclear
  - Apparently culture negative at end of treatment
- No BCG

Clinical presentation

Liverpool ED, August 2010:

- Depression
  - after relationship difficulties and separation
  - (lost contact with child)
- Cough, night sweats and weight loss
  - for at least 16 weeks
  - already cachectic

Investigations

- CXR
  - Multiple cavities
  - Extensive bilateral fibrocalcific changes and scarring
  - Significant volume loss on right
- Sputum
  - 3+ AFB from multiple expectorated sputa
- Vitamin D deficiency

M. tuberculosis phenotype (2010)

- Resistant to:
  - isoniazid
  - rifampicin and rifabutin
  - pyrazinamide
  - streptomycin
  - amikacin
  - capreomycin
  - ciprofloxacin
  - moxifloxacin (using 0.5 and 1.0 mg/L)
  - ethionamide
- Susceptible to:
  - ethambutol *
  - p-aminosalicylic acid
  - clofazimine
  - moxifloxacin (using 2 mg/L)
  - linezolid

Outcome

- Stayed more than 2 years in hospital
  - Gained weight initially
  - Numerous adverse effects from Rx
  - Depression persisted
  - Sputum remained culture positive
  - Weight began to decrease again
- Accommodation found
  - discharged but continued to lose weight
- Readmitted with massive ascites Dec 2012
  - MTB grew from ascitic fluid
  - Died with pulmonary haemorrhage Jan 2013
  - 23 years after initial treatment for TB

Can we learn anything from a tragedy like this?

Sequencing

- Two isolates have been sequenced
  - October 2010
  - December 2011
- Process:
  - Extraction at ICPMR
  - Library preparation at AGRF
    - insert size ~205bp (SD 80-90bp)
  - Illumina HiSeq paired end reads
    - (100bp at each end of fragment)

Analysis

- Various tools used
  - nesoni
  - trimmomatic
  - bwa
  - bowtie2
  - samtools
  - SOAP denovo
  - PAGIT
  - spolpred
  - awk
  - artemis, act
  - TBDREAMDB

AGRF data summary

- Genome coverage 400-600x
  - 2.7 - 3.0 Gb compressed FASTQ per genome
- ~90% of reads map to reference
  - (H37Rv: NC_000962)
  - i.e. 10% of reads do not map

Is the data from the sequencing run valid?

- What are the unmapped reads?
  - Assemble them using SOAP denovo
  - Blast twenty of the larger fragments
    - Most of them hit MTBC other than H37Rv
    - A few hit plants eg. *Ricinus communis*
    - The rest mostly have no hits in Genbank
  - Observation: many hits from *M.canettii*, *M. bovis*, fewer hits against modern TB?
- Try mapping against other reference genomes
  - Similar results to those for H37Rv from a few attempts

Could there be more than one strain of TB in the sequenced library?

- If some reads cannot be mapped to H37Rv, perhaps there is a mixture of MTB and another MTBC.
- How can this be tested?

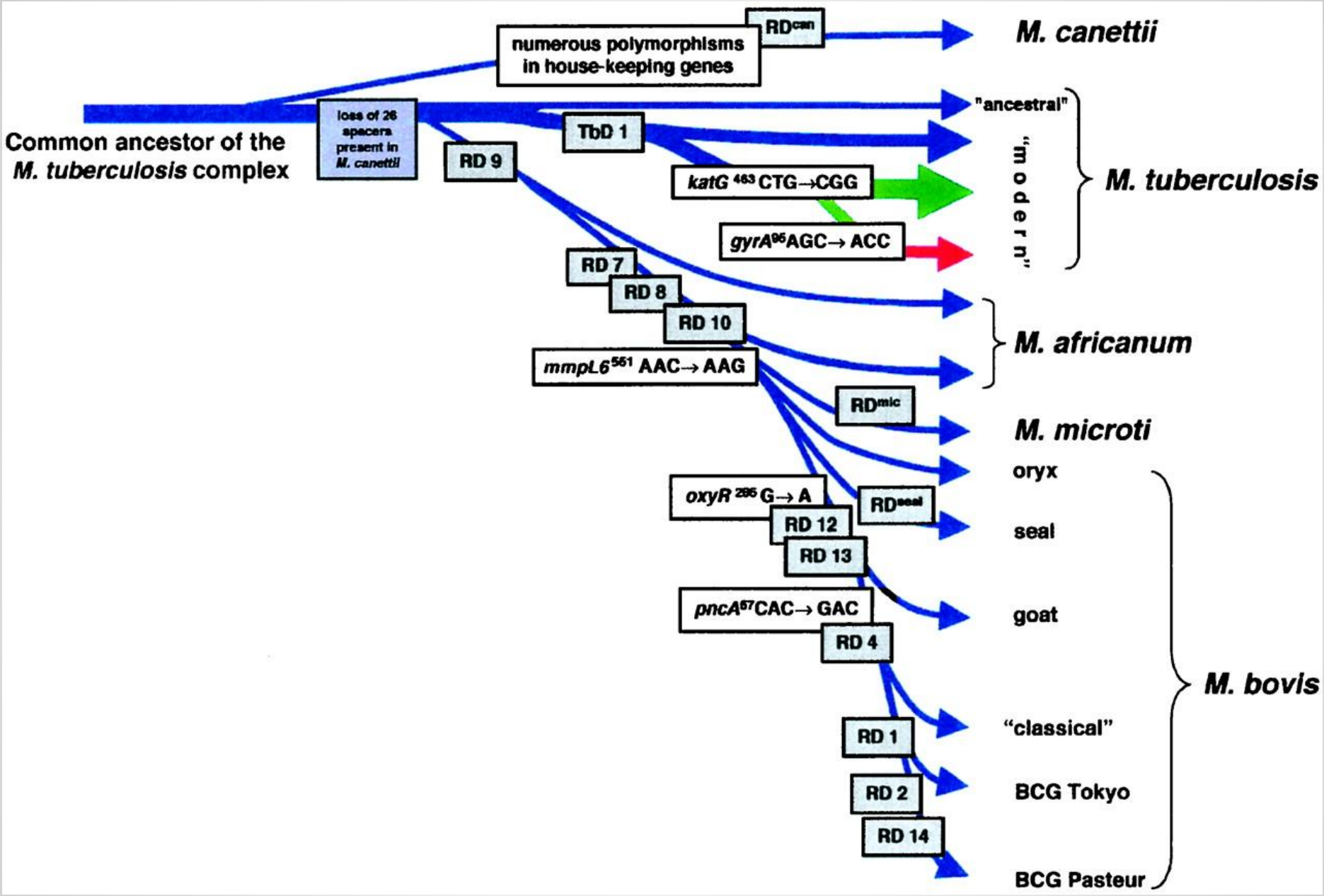
How to check if library is pure?

- Our best idea:

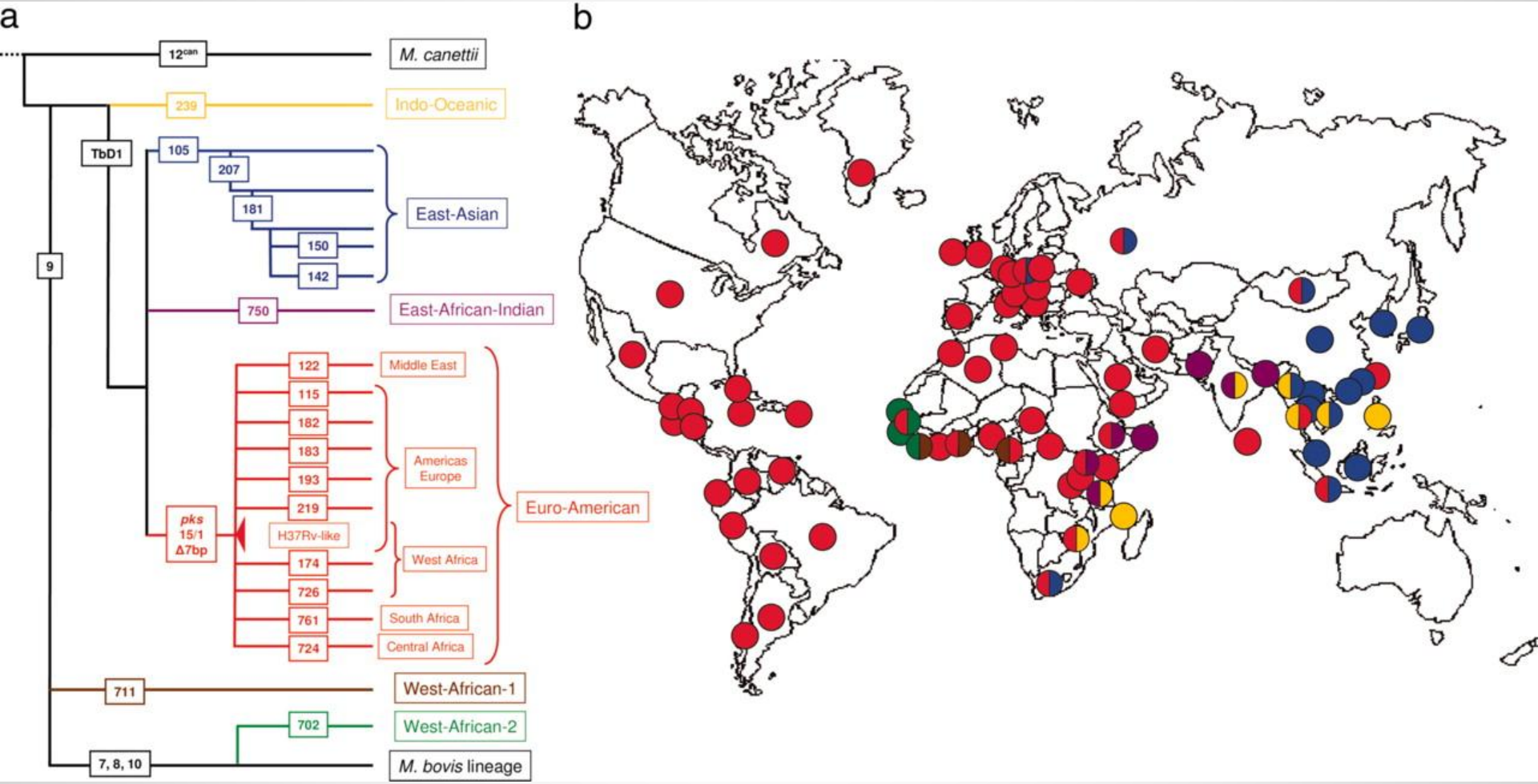
  See if all the reads give a consistent picture of large sequence polymorphisms (LSPs)

Interlude: what are LSPs in MTBC?

- RD001: most famous LSP
  - lost from *M. bovis* genome after serial passage on glycerinated bile-potato medium by Calmette and Guerin before 1921
  - includes genes for ESAT-6 and CFP-10
  - (hence IGRAs)
- Once a chunk has been lost, can it come back?
  - MTBC is clonal, recombination relatively rare, horizontal gene transfer virtually nonexistent



Brosch R, Gordon SV et al. 2002, PNAS 99(6):3684-9, figure 2



Gagneux S, DeRiemer K et al. (2006), PNAS 103(8):2869-73, figure 1

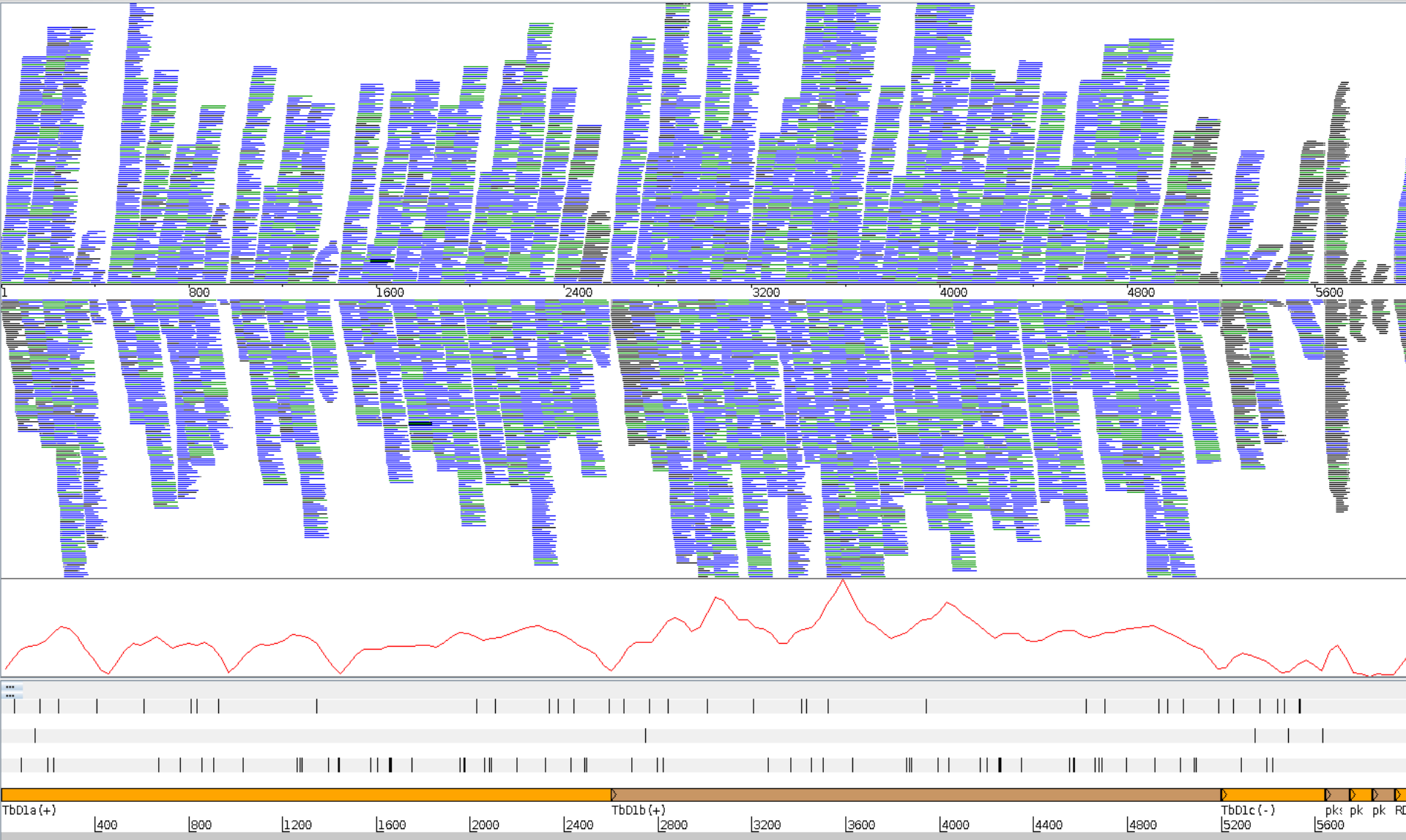
LSP analysis

- Hundreds of LSPs
- MTB lineages can be defined by some
- Collected lineage- and sublineage-defining LSPs:
  - Harder than it sounds - primers published, but not intervening sequence
  - Publicly available genomes biased towards Euro, N. American and East Asian TB, also strains of BCG
  - Eg. CAS/African/Indian lineage by de novo assembly
  - Scripts written to perform "virtual PCR" on genome data to obtain LSP "amplicons"

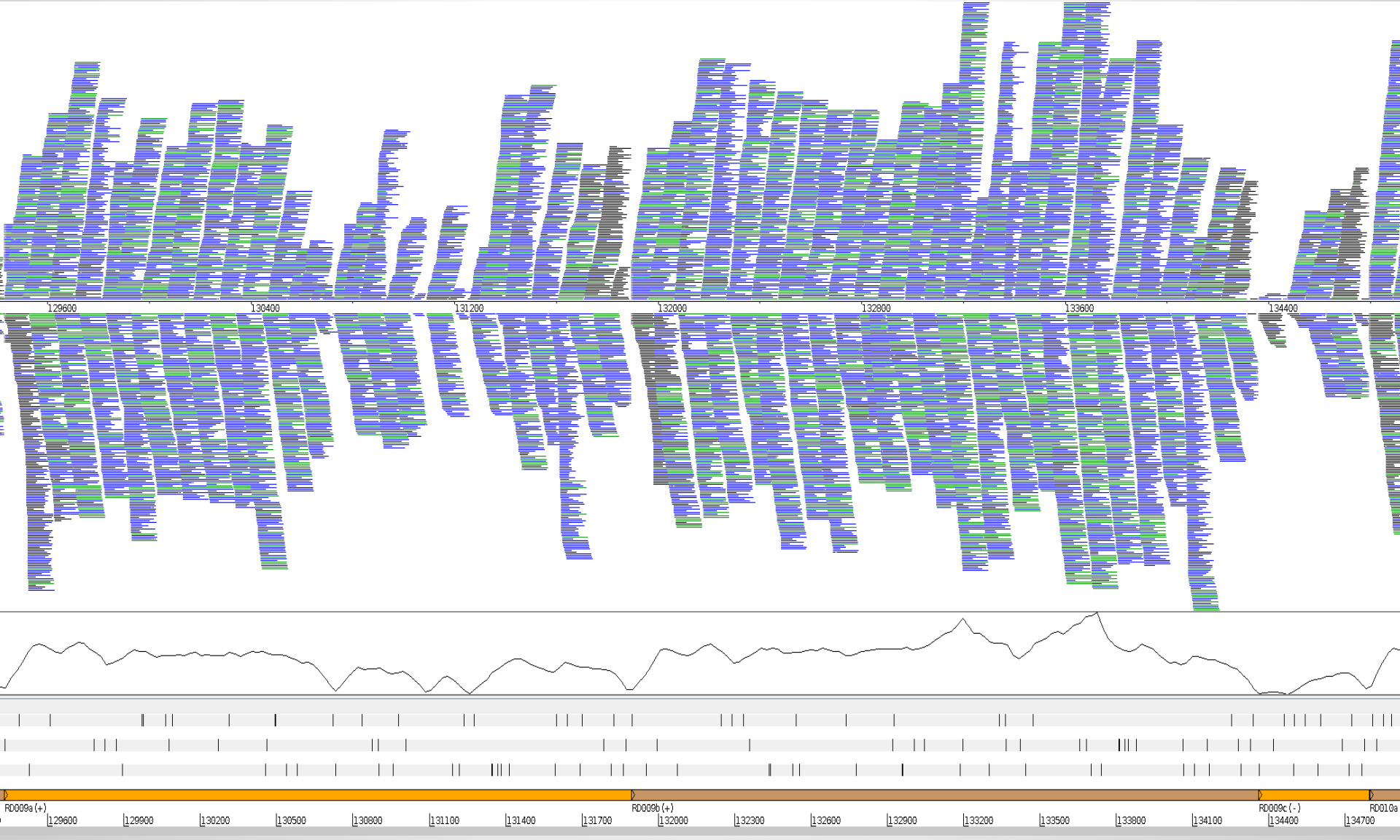
List of LSPs

- LSP(+) & (-) found:
  - TbD1
  - pks15
  - RD001, RD002
  - RD004, RD007
  - RD008, RD009
  - RD010, RD012
  - RD105, RD115
  - RD142, RD174
  - RD181, RD182
  - RD193, RD207
  - RD239, RD702
  - RD711, RD724
  - RD750, RD761
- LSP(-) not found:
  - RD122
  - RD150
  - RD183
  - RD219
  - RD726
- Still looking for strains missing these LSPs...

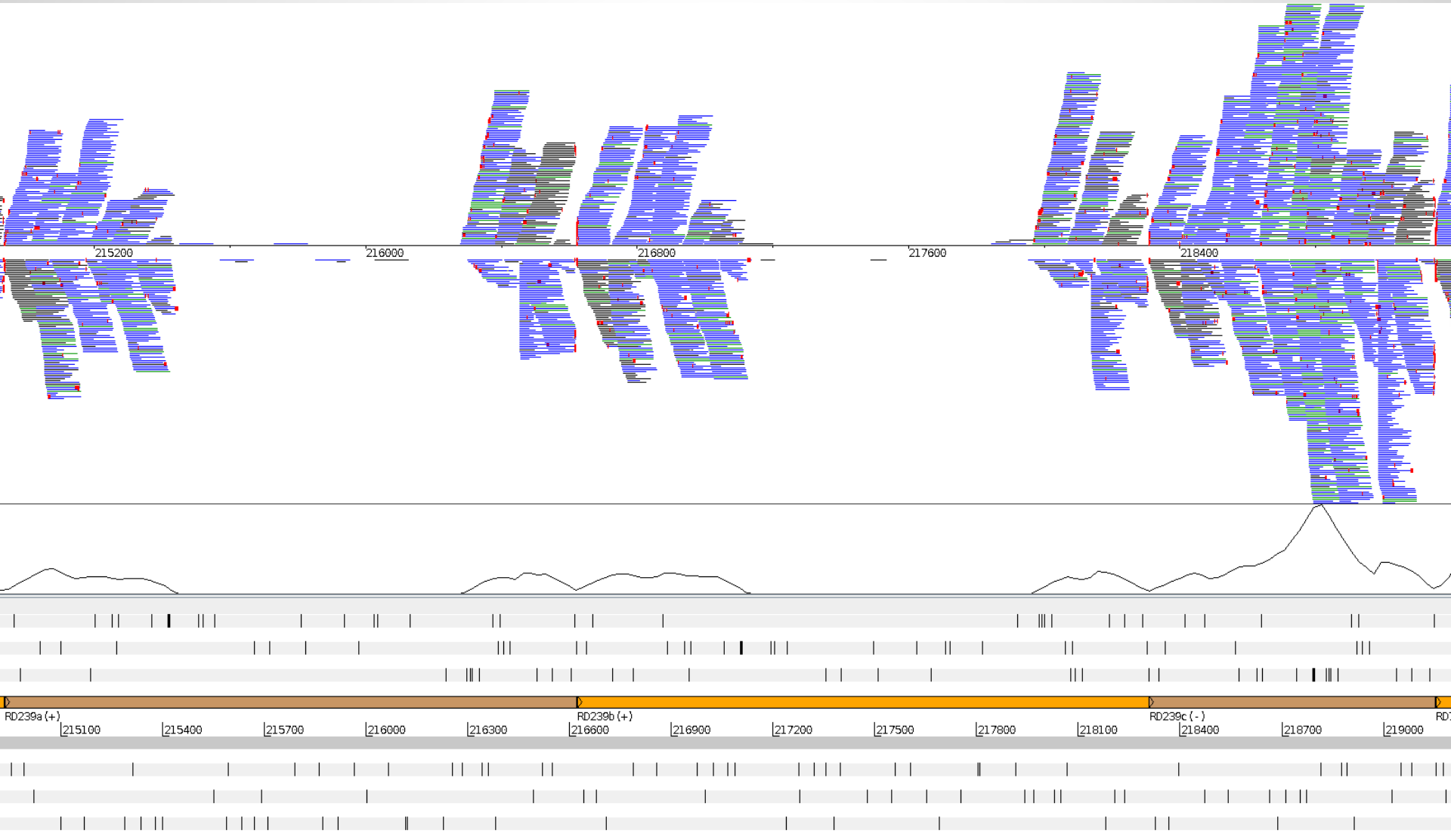
TbD1 is present, pks15 has no deletion

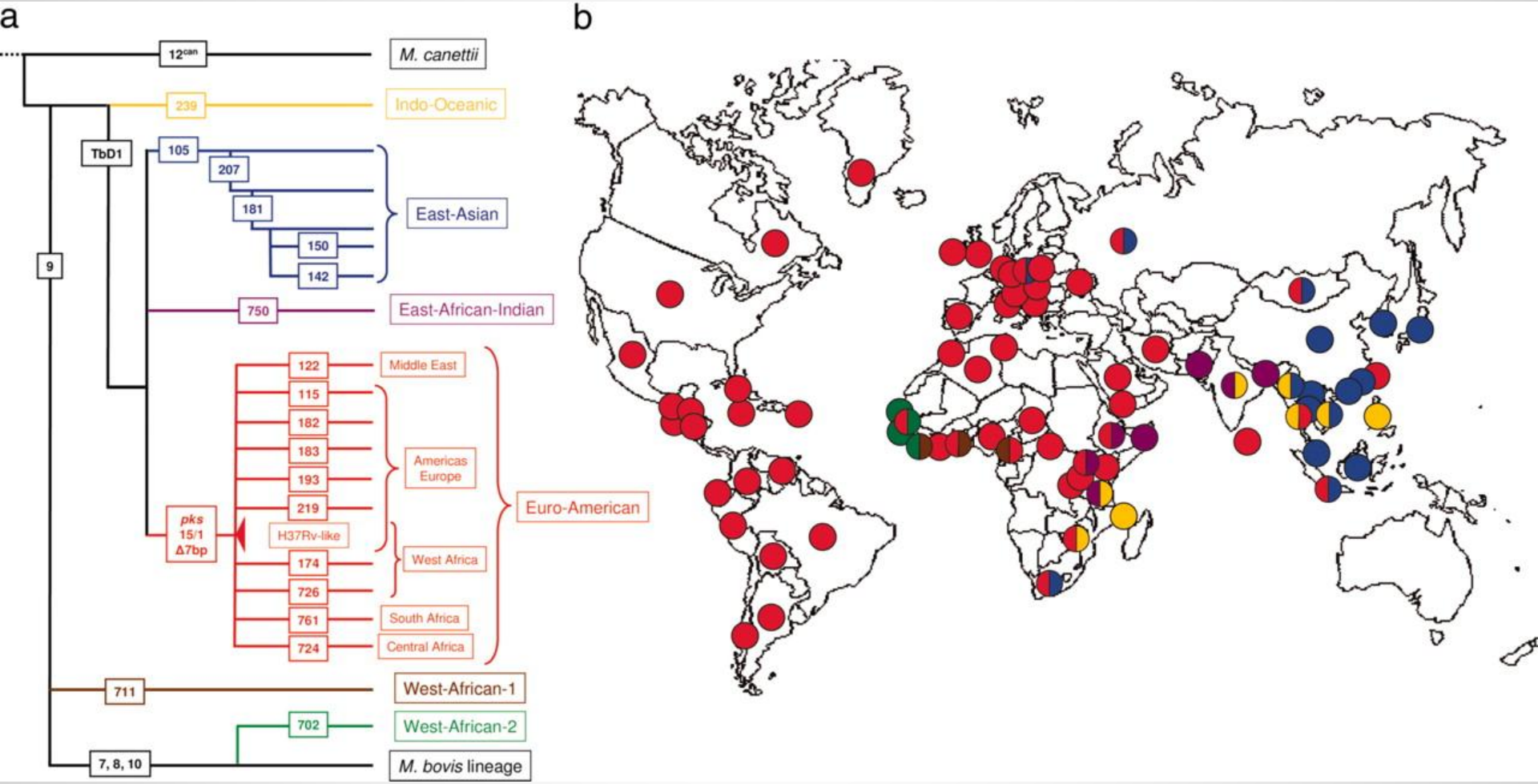


RD9 is present



RD239 is absent





Gagneux S, DeRiemer K et al. (2006), PNAS 103(8):2869-73, figure 1

Lineage 1: "EAI", "Indo-Oceanic"

- "Old" lineage
  - "new" lineages 2, 3, 4 (Beijing, CAS, Haarlem/LAM)
- Common in Philippines, Vietnam, South Asia, and East Africa
- LSP findings match:
  - spoligotype - 477777376413760 (spolpred)
  - MIRU-VNTR
  - Grant Hill-Cawthorne's SNP-based analysis

Confidence in library quality

- No reads in RD239:
  - no evidence of MTB DNA from any lineage other than lineage 1
- Drug resistance SNPs have strong and consistent evidence from read mapping
  - no evidence of MTB DNA that is not XDR
- Therefore analysis can be performed with some confidence

Resistance mutations

- Variants found using nesoni, bowtie2 and samtools to map onto H37Rv
- SNPs reported here were verified by inspection

Isoniazid: recognised association

- KatG: catalase-peroxidase-peroxynitritase T
  - S=>T - Rv1908c base 944 codon 315
  - R=>L - Rv1908c base 1388 codon 463

- FabD: ACP S-malonyltransferase
  - S=>N - Rv2243 base 824 codon 275

Isoniazid: weaker association

- IniA: isoniazid inductible gene protein
  - H=>Q - Rv0342 base 1443 codon 481
- "hypothetical protein" (lipid metabolism)
  - I=>V - Rv1592c base 964 codon 322
- ProA: γ-glutamyl phosphate reductase
  - V=>L - Rv2427c base 418 codon 140
- EmbB: indolylacetylinositol arabinosyltransferase
  - D=>G - Rv3795 base 983 codon 328
  - E=>A - Rv3795 base 1133 codon 378
  - G=>S - Rv3795 base 1216 codon 406

Rifampicin

- RpoB: DNA-directed RNA polymerase subunit beta
  - S=>W - Rv0667 base 1349 codon 450 (EC:531)

- EmbB: indolylacetylinositol arabinosyltransferase
  - D=>G - Rv3795 base 983 codon 328
  - E=>A - Rv3795 base 1133 codon 378
  - G=>S - Rv3795 base 1216 codon 406

Pyrazinamide: recognised association

- PncA: pyrazinamidase
  - D=>E - Rv2043c base 474 codon 158
  - T=>M - Rv2043c base 479-80 codon 160

Ethambutol: recognised association

- EmbC, EmbA, EmbB:
  indolylacetylinositol arabinosyltransferase
  - T=>I - Rv3793 base 809 codon 270
  - N=>D - Rv3793 base 1180 codon 394
  - V=>M - Rv3794 base 616 codon 206
  - P=>S - Rv3794 base 2737 codon 913
  - D=>G - Rv3795 base 983 codon 328
  - E=>A - Rv3795 base 1133 codon 378
  - G=>S - Rv3795 base 1216 codon 406
- RmlD:

  dTDP-6-deoxy-L-lyxo-4-hexulose reductase
  - S=>P - Rv3266c base 769 codon 257

Ethambutol: weaker association

- EmbR: transcriptional regulator
  - D=>N - Rv1267c base 319 codon 107
  - C=>Y - Rv1267c base 329 codon 110

- IniA: isoniazid inductible gene protein
  - H=>Q - Rv0342 base 1443 codon 481

Quinolones: recognised association

- GyrA: DNA gyrase subunit A
  - E=>Q - Rv0006 base 61 codon 21
  - A=>V - Rv0006 base 269 codon 90
  - S=>T - Rv0006 base 284 codon 95
  - A=>V - Rv0006 base 1151 codon 384
  - G=>D - Rv0006 base 2003 codon 668

- GyrB: DNA gyrase subunit B
  - M=>I - Rv0005 base 990 codon 330

Streptomycin

- RpsL: 30S ribosomal protein S12
  - K=>R - Rv0682 base 128 codon 43
- GidB: 16S rRNA methyltransferase
  - S=>F Rv3919c base 299 codon 100
- rrs: 16S ribosomal RNA
  - rRNA A=>G - Rvnr01 base 1401

Amikacin and capreomycin

- rrs: 16S ribosomal RNA
  - rRNA A=>G - Rvnr01 base 1401

Ethionamide: weaker association

- EthA: monooxygenase
  - C=>W - Rv3854c base 1209 codon 403

M. tuberculosis phenotype (2010)

- Resistant to:
  - isoniazid
  - rifampicin and rifabutin
  - pyrazinamide
  - streptomycin
  - amikacin
  - capreomycin
  - ciprofloxacin
  - moxifloxacin (using 0.5 and 1.0 mg/L)
  - ethionamide
- Susceptible to:
  - ethambutol *
  - p-aminosalicylic acid
  - clofazimine
  - moxifloxacin (using 2 mg/L)
  - linezolid

Changes from Oct 2010 to Dec 2011

- Quinolones: GyrA (Rv006)
  - D=>A base 281 codon 94 - extra mutation
  - A=>V base 269 codon 90 - reverted to wild type?

Additional changes of unknown significance (from 2010 to 2011)

- Rv0095c: "hypothetical protein"
  - D=>E base 171 codon 57
- Rv2020c: "hypothetical protein"
  - DDP=>ANP of codons 70..72
- QcrB: ubiquinol-cytochrome C reductase
  - P=>S base 562 codon 188
- Rv2351c: membrane phospholipase C
  - M=>I base 453 codon 151
- Rv3696c: glycerol kinase
  - R=>S base 280 codon 94
- Rv3885c: "hypothetical protein"
  - T=>I base 1286 codon 429

Research questions

- Are any of these changes associated with additional second-line resistance?
  - Or fitness?
- Too many variants against H37Rv:
  - What would we find mapping against a lineage 1 reference strain?
- What additional changes developed between 2011 and 2013?
  - What was the molecular clock during infection/Rx?
  - Do the mutations between these isolates represent random sampling ("quasispecies cloud"), or selection by Rx and immune response?

Summary

- XDR TB
  - not from expected Beijing or Euro-American lineage
  - from more diverse Lineage 1
- Good correlation with phenotype
  - except for initial S/ethambutol - genotype appears resistant
  - multiple gyrA & gyrB could predict further evolution of resistance during quinolone Rx ("MPC" theory?)
- Additional questions arise from analysis of genomic data

Thanks

- Mycobacterium reference lab
  - Peter Jelfs
  - Basel Suliman
- Liverpool chest clinic
  - Pam Banner
  - ABH Crawford
- Supervisors
  - Vitali Sintchenko
  - Ben Marais
  - Grant Hill-Cawthorne