

Mycobacteriology Proteogenomics

DAVID L. TABB, PH.D.

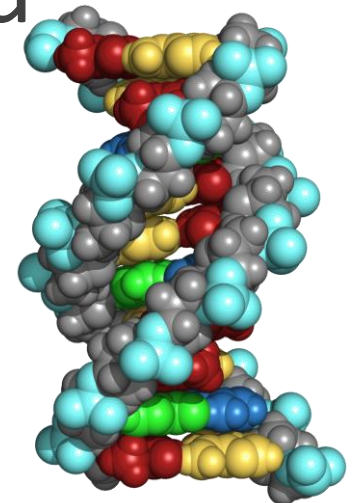
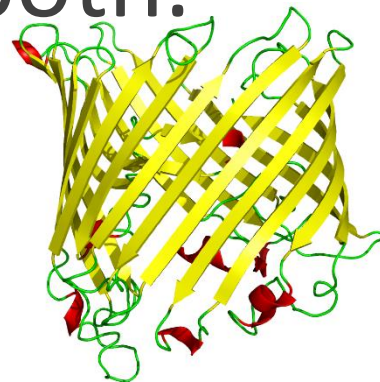
MOLECULAR BIOLOGY AND HUMAN GENETICS

Outline

- Defining proteogenomics
- Gene expression \neq protein expression
- Confirming expression of nsSNVs
- Handling Regions of Difference and insertions

Proteogenomics, defined

A field of biotechnology that combines DNA and/or RNA sequencing data with tandem mass spectrometry data for improved interpretation of both.

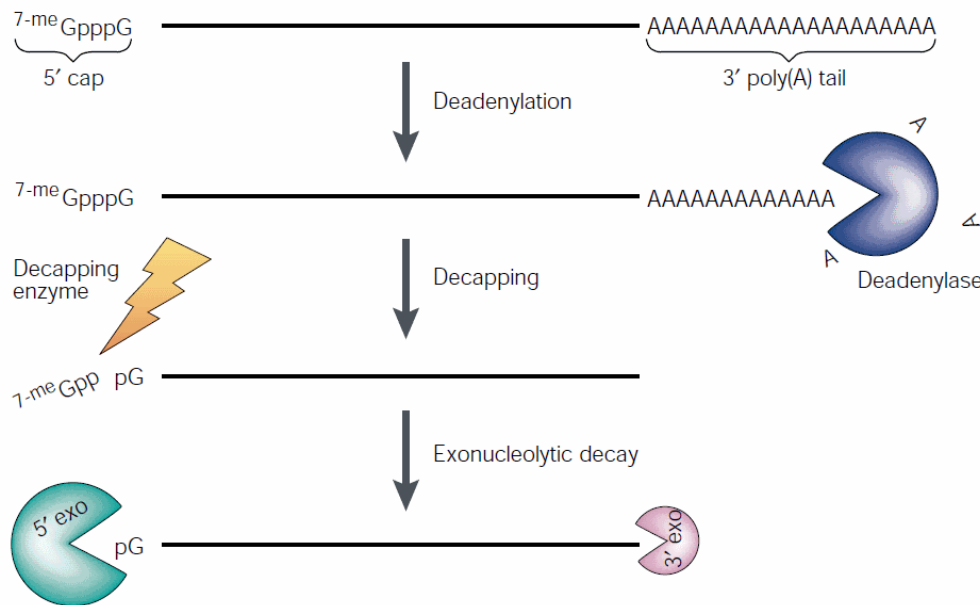


Common goals for Proteogenomics

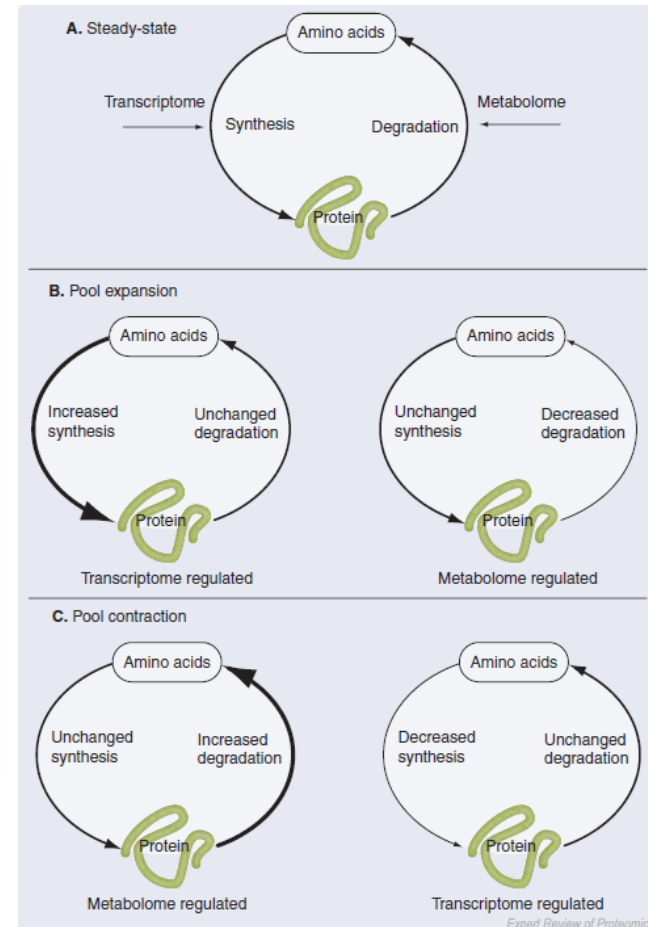
- Detect novel ORFs from genomic sequence
JD Jaffe et al. *Proteomics* (2004) 4: 59-77
- Refine intron/exon boundaries
S Tanner et al. *Genome Res.* (2007) 17: 231-239
- Measure expression of variant sequences
M Bunger et al. *J. Proteome Res.* (2007) 6: 2331-2340
- Compare mRNA and protein quantities
S Ghaemmaghami et al. *Nature* (2003) 425: 737-741

mRNA and Protein turnover: synthesis balances degradation

Typical mRNA degradation:



CJ Wilusz et al. *Nat. Rev. Mol. Cell. Bio.* (2001) 2: 237-246



Protein half-lives vary, yielding complex ratio to mRNA

PRATT ET AL. *MOL. CELL.
PROTEOMICS* (2002) 1:579-591.

Half
life:
31h

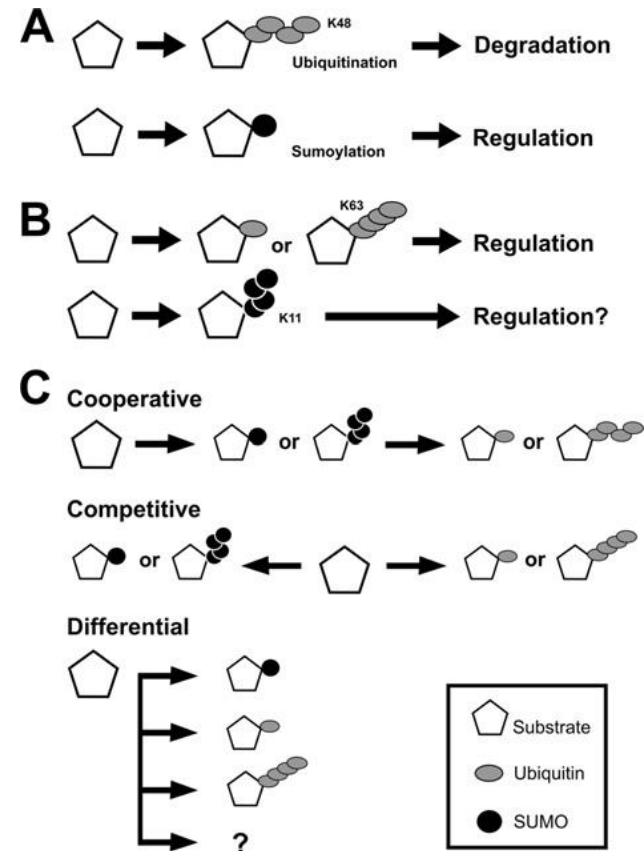
“The average rate of degradation of 50 proteins was 2.2%/h, although some proteins were turned over at imperceptible rates, and others had degradation rates of almost 10%/h.”

TJ GRIFFIN ET AL. *MOL. CELL.
PROTEOMICS* (2002) 1:323-333.

“Insights into the perturbative effects on genes involved in respiration, energy generation, and protein synthesis were obtained that would not have been apparent from measurements made at either the messenger RNA or protein level alone...”

PTMs play key role in degradation and activity

- A. “Classical effects of ubiquitination and SUMOylation.
- B. Expanded effects of new forms of ubiquitination.
- C. Combinatorial effects of the cross-talk of Ub and SUMO modifications.”



Mycobacterial Proteogenomics

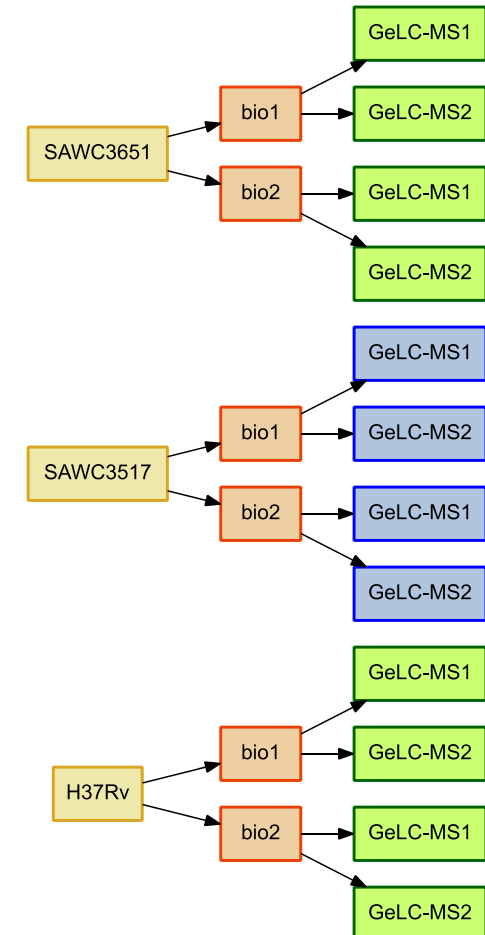
- DS Kelkar et al. *Mol. Cell. Proteomics* (2011) 10: M111.011627
Detected 250 novel peptides for 41 novel genes,
Corrected 79 gene models, and
Moved 33 start sites earlier.
- OT Schubert et al. *Cell Host & Microbe* (2013) 13: 602-612
Detected 29 novel proteins and
Quantified 2195 annotated *M.tb* proteins.
- MG Potgieter et al. *Front. Microbiol.* (2016) 7: 427
Detected 63 novel ORFs for *M. smegmatis* and
Revised 81 N-termini.





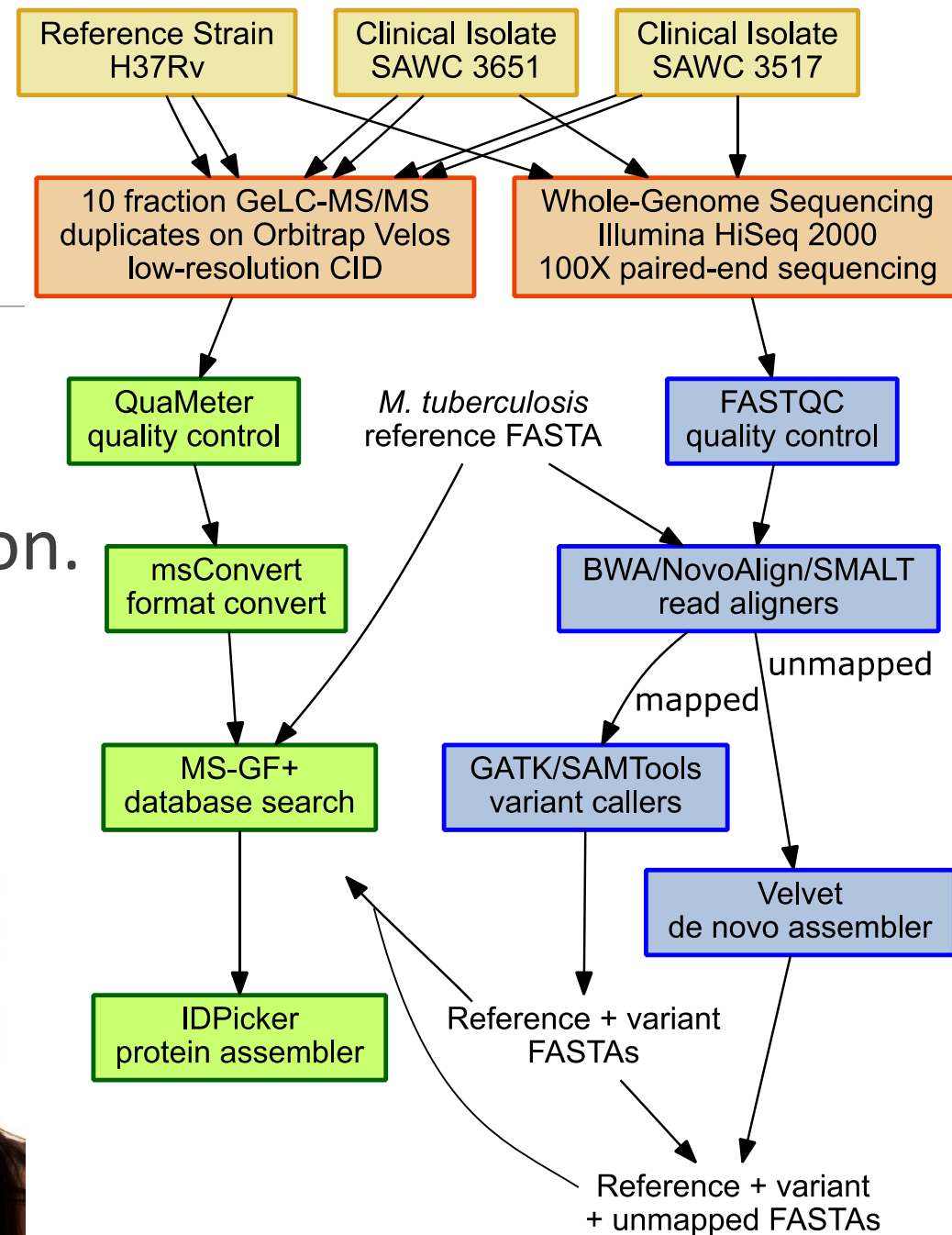
Proteogenomics: which genomic changes in *M.tb* strains are expressed?

- Two LAM *M.tb* strains were cultured from SA patients in the Western Cape.
- Genomic sequencing found major deletion separating them, with more variants separating both from the H37Rv reference strain.
- 10-fraction OrbiTrap Velos GeLC-MS inventoried peptides for six samples.



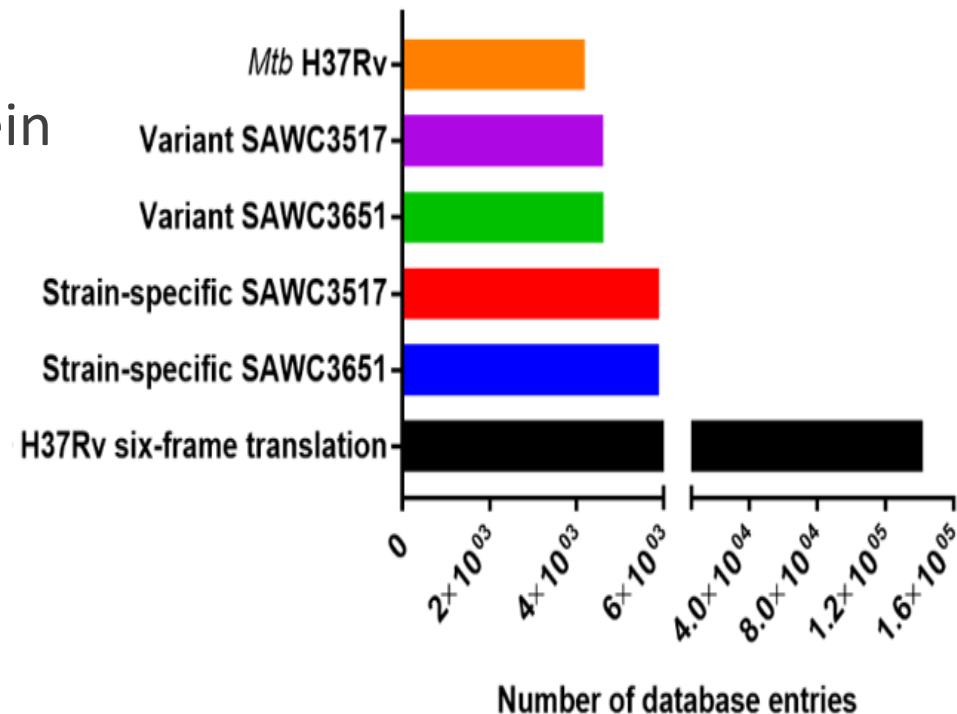
Data flow

- Both pathways receive *M.tb* reference annotation.
- Genome analysis informs proteome analysis, not vice versa.

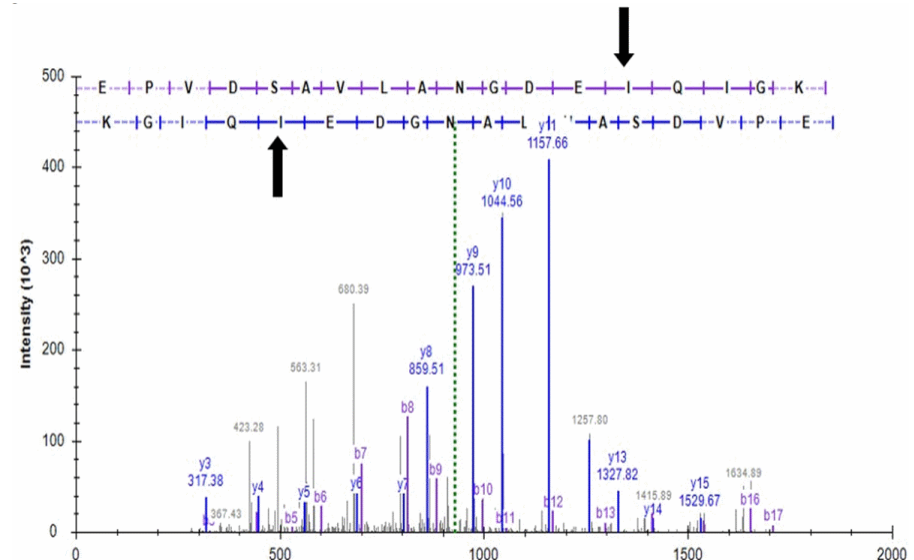
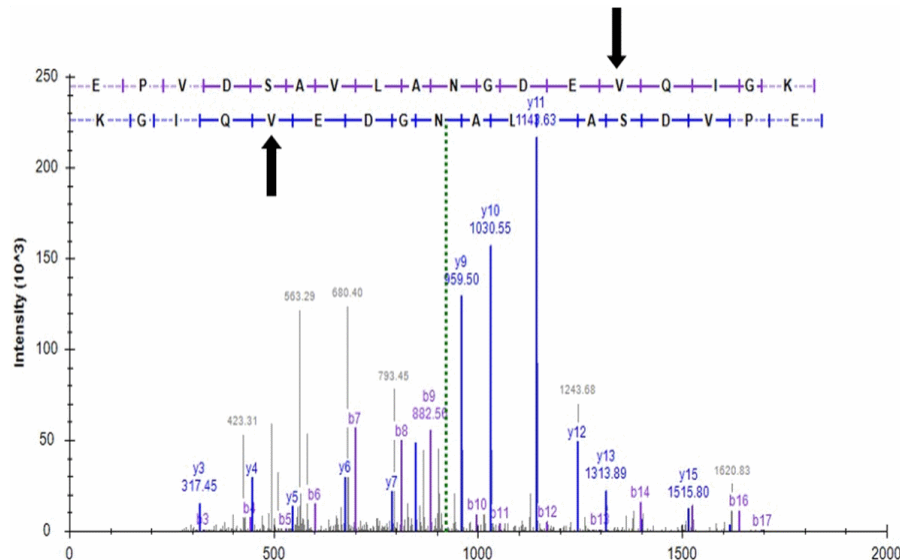


Adding strain-specific sequences impacts search space.

- Reference TubercuList proteome contains 4183 protein sequences.
- Applying nsSNVs to H37Rv made 422/430 additions.
- Adding unmapped regions via six-frame translation added ~1300 entries.
- “Six-frame” DB dwarfed the strain-specific databases.



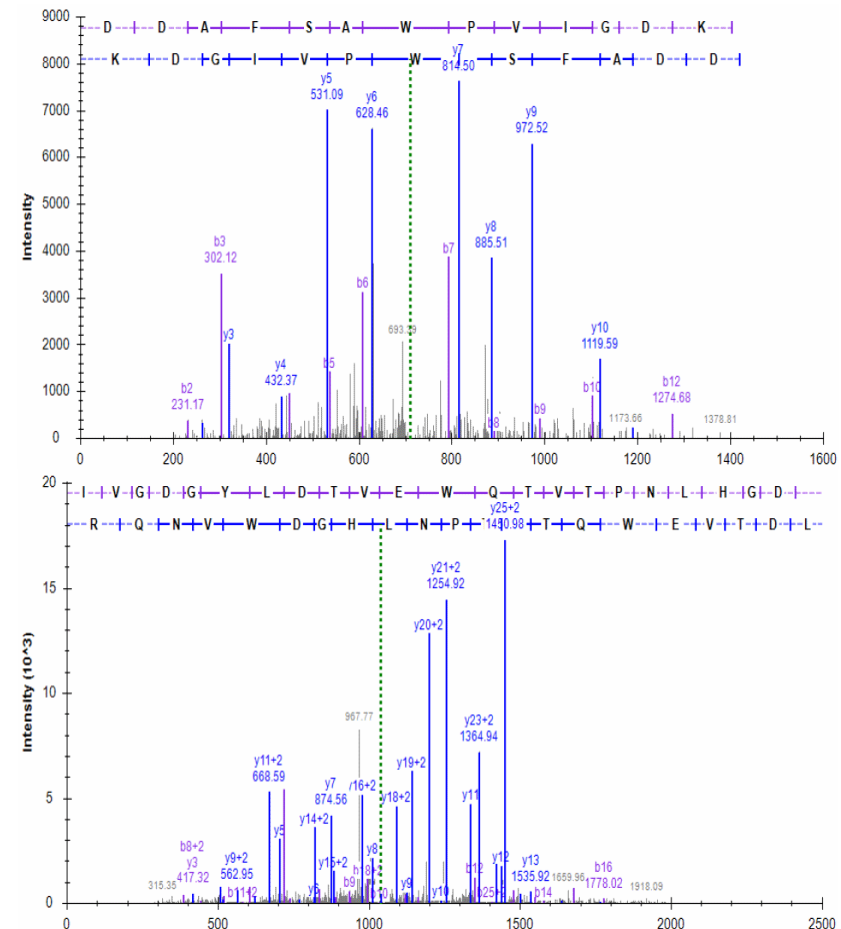
Ion trap data are sufficient to detect variants



- Val becomes Ile at 137 in GarA for SAWC3651, changing both precursor mass and most of the y" ion masses.
- We identified a total of 72 variant peptides, with 59 passing manual inspection of MS/MS evidence.

Interpreting the hits to unmapped DNA

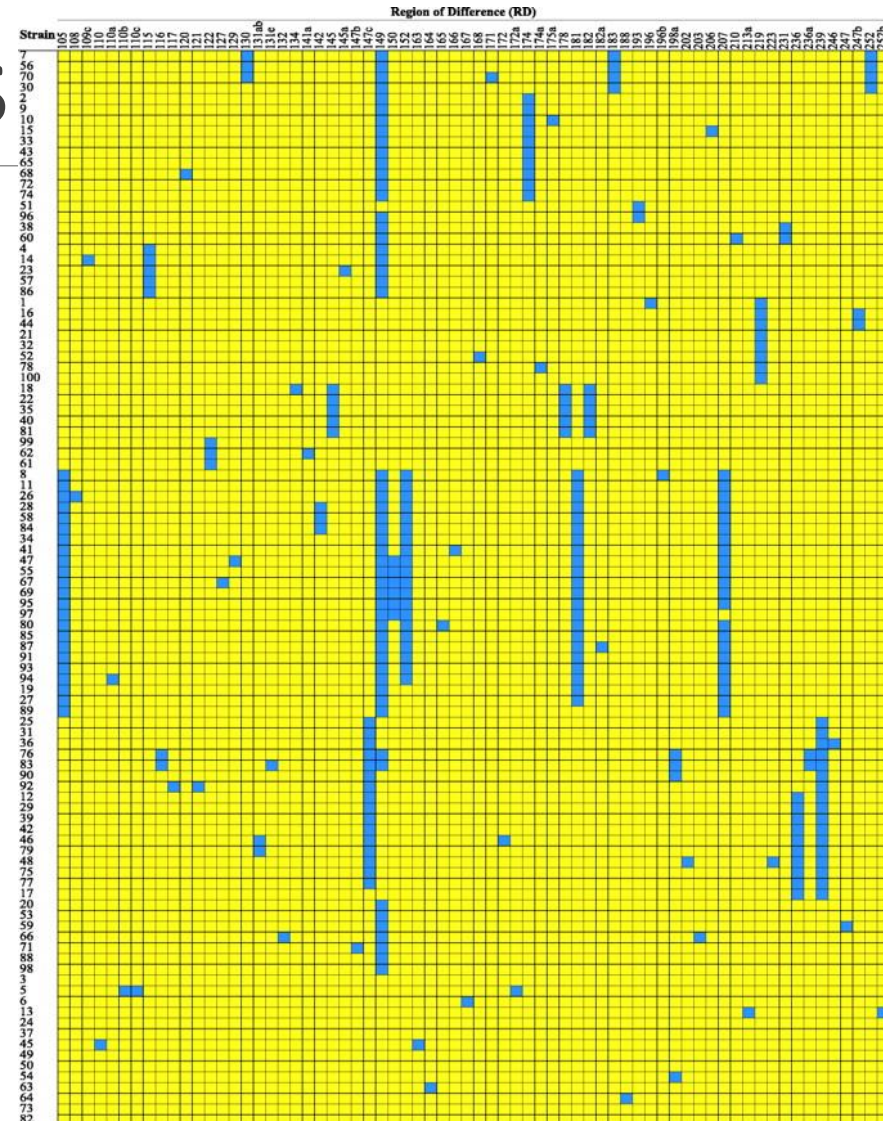
- Some peptides hitting “novel” genes also matched reference sequences.
- Peptides matching deleted regions could be explained by undeleted paralogs.
- 29 peptides supported novel helicase from SAWC3651.



Regions of Difference are variable deletions

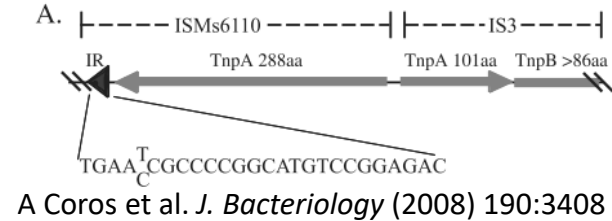
- 68 RD cover 4.2% of H37Rv genome, including 224 protein-coding genes.
- RD appear in syntenic aggregations.

AG Tsolaki et al. *PNAS*
(2004) 101:4865-4870



Overall distribution of RD among 100 clinical isolates of *M.tb*

Can RD be confirmed or overturned through proteomics?



- Peptides were detected that matched to RD152 and a SAWC3517-specific RD; these hits were for IS6110, found in many copies throughout many mycobacteria. No disproof.
- Observed peptides differentiate clinical isolates by matching regions deleted in one.
- One cannot prove absence through proteomics, only produce contrary evidence.

Takeaway Messages

- Turnover implies that mRNA and protein may give different quantities.
- Proteomics can confirm a subset of variant sequences derived through genomics.
- Despite deletion of a region of difference, peptides may yet appear through paralogy.