Evolutionary Forces in *Mycobacterium tuberculosis*

Sébastien Gagneux, PhD

22\textsuperscript{nd} September, 2010
Molecular Typing of *Mtb*

What is the question? 

(i.e. why are you genotyping your strains?)
Use the **Ideal** Marker for Each Question

- **Pragmatic / public health**
  ("traditional" Molepi)
  \[\rightarrow\text{MIRU-VNTR / spoligo}\]

  vs

- **Strain classification**
  (e.g. phenotypic associations)
  \[\rightarrow\text{LSPs / SNPs}\]

  vs

- **Evolutionary studies**
  (e.g. phylogenetics)
  \[\rightarrow\text{DNA seq}\]

Comas *et al.* 2009 *PLoS ONE* 4: e7815
Comparison of 4 Global Phylogenies of *Mtb*

Baker et al. 2004

Gagneux et al. 2006

Gutacker et al. 2006

Filliol et al. 2006

Gagneux & Small *Lancet ID* 2007 7: 328-335
Comparative DNA Sequencing of *Mtb*

Two Data Sets:

**MLSA:** 89 genes, 108 strains

**Illumina seq:** 22 genomes


Comas *et al.* 2010 *Nature Genetics* 42: 498-503
LSPs  
(Gagneux et al. 2006)

MLSA  
(Hershberg et al. 2008)

Comas & Gagneux 2009 PLoS Pathogens 5: 61000600
Global Phylogeny of MTBC

- 108 strains
- ~70kbp/strain
- Parsimony
- HI = 0.0043

Hershberg et al. 2008 PLoS Biology 6: e311

Animal Strains

‘ancient’ (TbD+)

‘modern’ (TbD−)

Hershberg et al. 2008 PLoS Biology 6: e311
22 *Mtb* Genomes

- **East Asia** (Lineage 2)
- **India, East Africa** (Lineage 3)
- **Europe, America, Africa** (Lineage 4)
- **The Philippines** (Lineage 1)
- **Rim of Indian Ocean** (Lineage 1)
- **West Africa 1** (Lineage 5)
- **West Africa 2** (Lineage 6)

*Comas et al. 2010 Nature Genetics 42: 498-503*
Nature of DNA Sequence Diversity

Understanding dN/dS

- nSNP: change in amino acid
- sSNP: no change in amino acid
- \( dN = \frac{\text{nSNPs observed}}{\text{nSNPs possible}} \)
- \( dS = \frac{\text{sSNPs observed}}{\text{sSNPs possible}} \)

- If \( dN/dS < 1 \), purifying selection
  - If \( dN/dS > 1 \), diversifying selection
dN/dS is unusually high in *Mtb*

- *Mtb*: 0.63 (~ 2/3 nSNPs !)
- *E. coli*: 0.09
- Other mycobacteria: ~ 0.16
- *M. canettii*: 0.18

Comas *et al.* 2010 *Nature Genetics* 42: 498-503
Bottlenecks in the Life History of Mtb

• Historical:

• Transmission:

1 to 10 bacteria
*Mtb* Has a **Small**
Effective Population Size

→ Reduced Purifying Selection in *Mtb*
Global Phylogeography of *Mtb*
“Out-of-and-back-to-Africa” of *Mtb*
Long-standing Association Between *Mtb* and Humans

- *Mtb* originated in Africa
- Diversity shaped by human migration & demography
- *Mtb* has phylogeographic structure

→ Host-pathogen co-evolution?
Host-Pathogen Co-Evolution

- Evolutionary ‘arms race’
- Immune evasion
- Examples:
  - HIV
  - HCV
  - *Plasmodium*
  - *Neisseria*
- Immune Evasion in *Mtb*?
22 Genomes - Three Gene Sets:

3,990 genes in H37Rv

- 3,717
  - 760 Essential
  - 2,879 Non-essential
  - 78 T-cell Antigens (491 Epitopes)

273 PE/PPE/mobile
Two Measures of Evolutionary Conservation

- Average Nucleotide Diversity ($\pi$)
- $dN/dS$
Essential Genes Are More Conserved than Non-essential Genes

$p < 0.002$

$p < 0.001$

Comas et al. 2010 Nature Genetics 42: 498-503
Antigens Are **NOT** More Diverse

N. S.
Antigens Are Equally Conserved

N. S.

- Essential: 0.53
- Non-essential: 0.66
- Antigens: 0.5

Overall, the dN/dS values indicate that antigens are equally conserved, as evidenced by the non-significant (N. S.) difference between the groups.
Antigens Consist of Epitopes and Non-Epitopes
Epitopes are More Conserved than Non-epitopes

\[ p < 0.05 \]
Epitopes Most Conserved

Comas et al. 2010 Nature Genetics 42: 498-503
Conclusions

- *Mtb* is more diverse
- Random genetic drift ("chance")
- Longstanding host-pathogen association
- T-cell epitopes are hyperconserved
Proposal for New Nomenclature for Main MTBC Lineages (excl. Animal TB)

Lineage 1: Rim of Indian Ocean
- The Philippines
- Indo-Oceanic
  - Cluster I
  - Cluster group 1
  - Lineage IV
    - EAI (East African-Indian)
      - MANU1

Lineage 2: East Asia
- East Asian
  - Cluster II
  - Cluster group 2
  - Lineage I
    - Beijing, ST523, ST123

Lineage 3: India and East Africa
- East-African-Indian
  - Cluster IIA
  - Cluster group 3
  - Lineage III
    - CAS (Central Asian)

Lineage 4: Europe and Americas
- Euro-American
  - Clusters III and VII
  - Cluster group 3b-5b
  - Lineage II
    - X, Harlem, LAM, Uganda

Lineage 5: West Africa
- M. africanum
  - West-African 1
    - Not done
    - Not done
    - Not done
    - AFR2
    - PGG1

Lineage 6: West Africa
- M. africanum
  - West-African 2
    - Not done
    - Not done
    - Not done
    - AFR1
    - PGG1

Coscolla & Gagneux Drug Discovery Today 2010 in press
Collaborators

- Joel Ernst
- Douglas Young
- Stefan Niemann
- Peter Small
- James Galagan
- Dmitri Petrov
- Marcus Feldman

Swiss TPH, Basel

Funding:
MRC, NIH/NIAID,
Royal Society, SNSF

MRC NIMR, London

- Bijaya Malla
- Mireia Coscollá
- David Stucki

- Sonia Borrell

- Graham Rose
- Tembi Huna
- Iñaki Comas