Mycobacterial population genetics and genomics

Stefan Niemann

Molecular Mycobacteriology Group

National Reference Center for Mycobacteria, Borstel, Germany

"Learning from strains in the field"
Outline

- Population structure of the *M. tuberculosis* complex
- Genetic Diversity
- What does this mean?
**Mycobacterium tuberculosis complex**

- 7 species:
  - *M. tuberculosis*, *M. africanum*, *M. microti*,
  - *M. pinipedii*, *M. caprae*, *M. bovis*,
  - *M. canetti* (*M. prototuberculosis*)

- Immotile, rod shaped bacteria:
  - slow growth rate, high GC, special cell wall

- High similarity on the DNA-level:
  - DNA-DNA relatedness of >99%

- **But**: Large differences in biochemical/phenotypical properties, geographical distribution and importance for TB in humans
Biodiversity

Left: *M. bovis* / right: *M. tuberculosis*

Niacin | Nitrate | Catalase | Bromcresol | Lebek

Large differences in biology, epidemiology and importance for TB in humans
Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination

Srinand Sreevatsan*, Xi Pan*, Kathryn E. Stockbauer*, Nancy D. Connell†, Barry N. Kreiswirth‡, Thomas S. Whittam§, and James M. Musser*†

- 26 structural genes, 842 strains
- 2 MB, only 32 SNPs
## Population structure - SNPs

### gyb discriminatory regions

<table>
<thead>
<tr>
<th>Reference sequences</th>
<th>region 1 (675)</th>
<th>region 2 (756)</th>
<th>region new (1311)</th>
<th>region 3 (1410)</th>
<th>region 4 (1450)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td>GCCCTA C GAGT</td>
<td>AACGCT G CGG</td>
<td>GGCCTG T STGA</td>
<td>TGTAAC C GAAC</td>
<td>CCACG G CGAA</td>
</tr>
<tr>
<td>M. bovis</td>
<td>GCCCTA C GAGT</td>
<td>AACGCT A CGG</td>
<td>GGCCTG T STGA</td>
<td>TGTAAC C GAAC</td>
<td>CCACG T CGAA</td>
</tr>
<tr>
<td>M. africanaum</td>
<td>GCCCTA C GAGT</td>
<td>AACGCT G CGG</td>
<td>GGCCTG T STGA</td>
<td>TGTAAC C GAAC</td>
<td>CCACG T CGAA</td>
</tr>
<tr>
<td>M. microti</td>
<td>GCCCTA T GAGT</td>
<td>AACGCT G CGG</td>
<td>GGCCTG T STGA</td>
<td>TGTAAC C GAAC</td>
<td>CCACG T CGAA</td>
</tr>
</tbody>
</table>

### Strains tested

- M. tuberculosis (n=5)
- M. bovis
- subsp. bovis (n=5)
- subsp. caprae (n=5)
- subsp. C (n=2)
- M. africanaum
- subtype I (n=5)
- subtype II (n=5)
- M. microti
- type llama (n=2)
- type vole (n=1)

**FIG. 1.** DNA sequences of the four discriminatory regions in the gyb gene described by Kasai et al. (9) and of one new region found in this study. Discriminatory base substitutions are shaded.

---

**Evaluation of Genotype MTBC Assay for Differentiation of Clinical Mycobacterium tuberculosis Complex Isolates**

Elvira Richter, Michael Weizenegger, Sabine Rüscher-Gerdes, and Stefan Niemann
Research article

*Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology

Brudey et a. BMC Microbiol 2006
Population structure - Spoligotyping
A new evolutionary scenario for the *Mycobacterium tuberculosis* complex

Population structure - deletions

Brosch et al. PNAS 2002
Population structure - deletions

Variable host–pathogen compatibility in *Mycobacterium tuberculosis*

Sebastien Gagneux, Kathryn DeRiemer, Tran Van, Midori Kato-Maeda, Bouke C. de Jong, Sujatha Narayanan, Mark Nicol, Stefan Niemann, Kristin Kremer, M. Cristina Gutierrez, Markus Hilty, Philip C. Hopewell, and Peter M. Small

Global Phylogeny of *M. tuberculosis*: 6 Main Lineages

PNAS 2006
Population structure - deletions

Gagneux et al. PNAS 2006
Population structure - MIRU

Phylogenetic tree
24 loci MIRU, 352 strains

Wirth et al. Plos Path, in revision
Population structure - MIRU

**TB complex and the coalescent**

Estimated Times (in years) since the most recent common ancestor (TMRCA). Estimates and 95% confidence intervals were calculated with the software Ytime.

<table>
<thead>
<tr>
<th>TMRCA</th>
<th>Age in years</th>
<th>CIs</th>
<th>Hierarchic level</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAM-Beijing</td>
<td>21,300</td>
<td>(14,300-31,600)</td>
<td>Clade 1</td>
</tr>
<tr>
<td>Beijing-CAS</td>
<td>17,100</td>
<td>(11,600-25,400)</td>
<td>Asian TB</td>
</tr>
<tr>
<td>LAM-LAM</td>
<td>7,060</td>
<td>(4,370-11,100)</td>
<td>LAM</td>
</tr>
<tr>
<td>CAS-CAS</td>
<td>9,450</td>
<td>(6,100-14,700)</td>
<td>CAS</td>
</tr>
<tr>
<td>EAI-WA2</td>
<td>32,800</td>
<td>(27,900-38,300)</td>
<td>Clade 2</td>
</tr>
<tr>
<td>EAI-EAI</td>
<td>13,700</td>
<td>(9,100-21,000)</td>
<td>EAI</td>
</tr>
<tr>
<td>M. bovis-M. bovis</td>
<td>5,750</td>
<td>(4,560-7230)</td>
<td>M. bovis</td>
</tr>
<tr>
<td>EAI-LAM</td>
<td>41,500</td>
<td>(29,100-60,000)</td>
<td>TB complex</td>
</tr>
<tr>
<td>EAI-Beijing</td>
<td>37,500</td>
<td>(25,800-55,100)</td>
<td>TB complex</td>
</tr>
</tbody>
</table>
Multilocus sequence analysis

- 108 strains
- 1 *M. canettii* strain (Outgroup)
- 89 Gene sequenced:
  - Housekeeping (N = 25)
  - Virulence (N = 15)
  - Antigene (N = 49)
- 65,829 bp / strain
- ~ 7.1 Mb total

370 SNPs (+ 118 SNPs in *M. canettii*)

High rate of non-synonymous SNPs (62%)
Genomic Diversity - MLST

Ancient

Modern

PGG 1

PGG 3

PGG 2
Overall genomic diversity in human strains is not different to animal strains

Hershberg et al. PLOS Biol., in review
MTBC strains are under extremely reduced purifying selection
We sequenced a susceptible and an MDR TB isolate from Nukus City:

Both isolates were isogenic Beijing strains as determined by traditional genotyping methods (IS6110typing; spoligotyping)
## Genetic variation and drug resistance

<table>
<thead>
<tr>
<th>Drug resistance</th>
<th>H37Rv + K-A</th>
<th>MDR</th>
<th>mutation</th>
<th>Capillary data</th>
<th>Known resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>rifampicin</td>
<td><em>rpoB</em></td>
<td><strong>CTG</strong> - <strong>CCG</strong></td>
<td>L458P</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>isoniazid</td>
<td><em>katG</em></td>
<td><strong>AGC</strong> - <strong>ACC</strong></td>
<td>S315T</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>pyrazinamide</td>
<td><em>pncA</em></td>
<td><strong>TGG</strong> - <strong>TGA</strong></td>
<td>W119stop</td>
<td>✔️</td>
<td>new</td>
</tr>
<tr>
<td>ethambutol</td>
<td><em>embB</em></td>
<td><strong>CAG</strong> - <strong>CGG</strong></td>
<td>Q497R</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>streptomycin</td>
<td><em>rpsL</em></td>
<td><strong>AAG</strong> - <strong>AGG</strong></td>
<td>K43R</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
Genotype and Disease
Variability of *M. tuberculosis* genotypes

*Mycobacterium tuberculosis* Beijing
Genotype Strains Associated with Febrile Response to Treatment

Reinout van Crevel,* Ron H.H. Nelwan,† Wilma de Lenne,* Yelilsan Veeraragu,‡ Adri G. van der Zanden,§ Zulkifli Amin,† Jos W.M. van der Meer,* and Dick van Soolingen¶

*EID, 2001*

A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes


Different Strains of *Mycobacterium tuberculosis* Cause Various Spectrums of Disease in the Rabbit Model of Tuberculosis

Yukari C. Manabe,1,2* Arthur M. Dannenberg, Jr.,3,4,5,6 Sandeep K. Tyagi,1 Christine L. Hatem,1 Mark Yoder,1 Samuel C. Woolwine,1 Bernard C. Zook,7 M. Louise M. Pitt,8 and William R. Bishai1,2,5

*Infect Immun, 2003*

Genetic background of *M. tuberculosis* genotypes influences virulence and transmissibility
Population structure – Eastern Europe

Prospective study 2003 – 2005
326 patients

Beijing genotype and MDR-TB:

OR 20.9
(95% CI 10.5-41.7 ), p<0.001

Panini et al. submitted
23% of all MDR cases caused by one clone (Cluster 5)

Highly resistant – Super virulent phenotype

Panini et al. submitted
Follow up study of 266 patients eligible for DOTS-Plus treatment in districts Nukus and Chambay 2003-2006

- 87 MDR-TB (no OFX resistance)
- 18 developed OFX resistance during treatment
  - 4 infected with a different strain
    - 3 XDR-TB
      - 3 treatment success
    - 1 MDR-TB + OFX res
      - Died
  - 1 with mixed strain infection
    - 1 MDR-TB + OFX res
      - Treatment failure (died)
  - 13 infected with the same strain
    - 7 XDR-TB
      - 2 died
      - 5 treatment failure
    - 6 MDR-TB + OFX res
      - 4 died
      - 2 treatment success

- 1 Death with XDR-TB 8 months after treatment end

Cox et al. NEJM, in revision
Summary

- *M. tuberculosis* complex: Phylogeographical population structure with distinct lineages
- Lineages can be identified by molecular methods e.g. MIRU classification
- Evolved from an *M. prototuberculosis* like ancestor approx. 40,000 years ago
- Lineages have specific pathogenic characteristics
- Beijing genotype strongly associated with drug resistance/transmission
- Clonal expansion of particular MDR strains with super transmissible phenotype
- Molecular mechanisms mainly unclear
Summary

- Reduced purifying selection results in **high functional diversity**
- **Ultrafast genome sequencing** with high genome coverage (98%) and high quality possible
- **Whole genome resistance genotype** assessable
- **Microevolution** of epidemic strains can be monitored
- **High amount of genetic diversity** in strains identical by DNA fingerprint

→ Importance for virulence??
Thanks to

TB-Team Borstel

MIRU work

P. Supply
Institute Pasteur
Lille

T. Wirth
Muséum National d'Histoire Naturelle
Paris

MLST work

P. Small, S. Gagneux
Institute for Systems Biology, Seattle
MRC, London

Genome sequencing

J. Archer, C. Koeser
University Cambridge

All other cooperation partners

Funding:
Welcome to the MIRU-VNTRplus web application!

Molecular typing of bacteria from the Mycobacterium tuberculosis complex (MTBC) is essential for epidemiological purposes such as investigating the spreading of specific genotypes. Recently, mycobacterial interspersed repetitive units (MIRU) typing has become an important method, as it allows high-throughput, discriminatory and reproducible analysis of clinical isolates. Because of its portable data format, MIRU typing has the potential to be a versatile tool for individual strain identification based on large reference databases.

For MIRU-VNTRplus, a collection of 197 strains representing the major MTBC lineages was used. For each strain epidemiologic and genotype information was stored together with copy numbers of 24 MIRU loci, spoligotyping patterns, regions of difference (RD) profiles, single nucleotide polymorphisms (SNPs), susceptibility data and ESRI RFLP fingerprint images.

Via this freely accessible service users can compare their strain(s) with the reference strains or analyze their strains without using the database content. Comparisons can be based on single MIRU-, spoligo-, RD-typing data, or by a combination of different datasets. Several distance coefficients are available, including Nei's D, Jaccard's and categorical. Based upon the respective distance matrix, a dendrogram can be calculated using UPGMA or neighbor-joining clustering algorithms. The resulting trees may be exported in various data formats.

To use the web application, select one of the commands below:

- Compare your strains with reference database
  - Add a single strain
  - Import multiple strains from file
  - Browse reference strains

- Analyse your strains, without using reference database
  - Import strains from a file
  - Browse your strains

- Change settings
  - Change MIRU-VNTR loci set
  - Change MIRU-VNTR loci order
  - Remove user strains
  - Reset default parameters
Virulence in model systems

Low dose aerosol infection model in mice

Cooperation with S. Ehlers, Kerstin Walter, Molekulare Infekt., Borstel
Virulence in model systems

TNF inducing potential in murine macrophages
(Infection with clinical isolates, t=24h, mean of two experiments)

Cooperation with N. Reiling, Molekulare Infekt., Borstel
M. tuberculosis complex

- M. tuberculosis: eugonic growth, positive for nitrate reduction and niacin accumulation
- M. africanum: dysgonic growth, variable results for nitrate reduction and niacin accumulation
- M. bovis: dysgonic growth, negative for nitrate reduction and niacin accumulation, resistant to PZA
- M. microti: S-shaped cell morphology, very slow growth
- M. canetti: "smooth" glossy colonies, positive for nitrate reduction and niacin accumulation (M. prototuberculosis)
- And some more: M. caprae, M. pinnipedii….
M. tuberculosis complex

Immotile, rod shaped bacteria: slow growth rate, high GC, special cell wall, high similarity on the genome level (99.5%)
Multilocus Sequence Analysis

nSNPs/sSNPs by gene class and phylogenetic branch

<table>
<thead>
<tr>
<th></th>
<th>nSNPs</th>
<th>sSNPs</th>
<th>nSNPs : sSNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>231</td>
<td>139</td>
<td>1.66</td>
</tr>
<tr>
<td><strong>Genes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housekeeping</td>
<td>94</td>
<td>59</td>
<td>1.59</td>
</tr>
<tr>
<td>Virulence</td>
<td>46</td>
<td>23</td>
<td>2.00</td>
</tr>
<tr>
<td>Surface</td>
<td>91</td>
<td>57</td>
<td>1.60</td>
</tr>
<tr>
<td><strong>Clades</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancient Human</td>
<td>77</td>
<td>50</td>
<td>1.54</td>
</tr>
<tr>
<td>Modern Human</td>
<td>107</td>
<td>57</td>
<td>1.88</td>
</tr>
<tr>
<td>Animal</td>
<td>40</td>
<td>28</td>
<td>1.43</td>
</tr>
<tr>
<td><strong>Branches</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td>100</td>
<td>66</td>
<td>1.52</td>
</tr>
<tr>
<td>External</td>
<td>131</td>
<td>73</td>
<td>1.79</td>
</tr>
</tbody>
</table>