

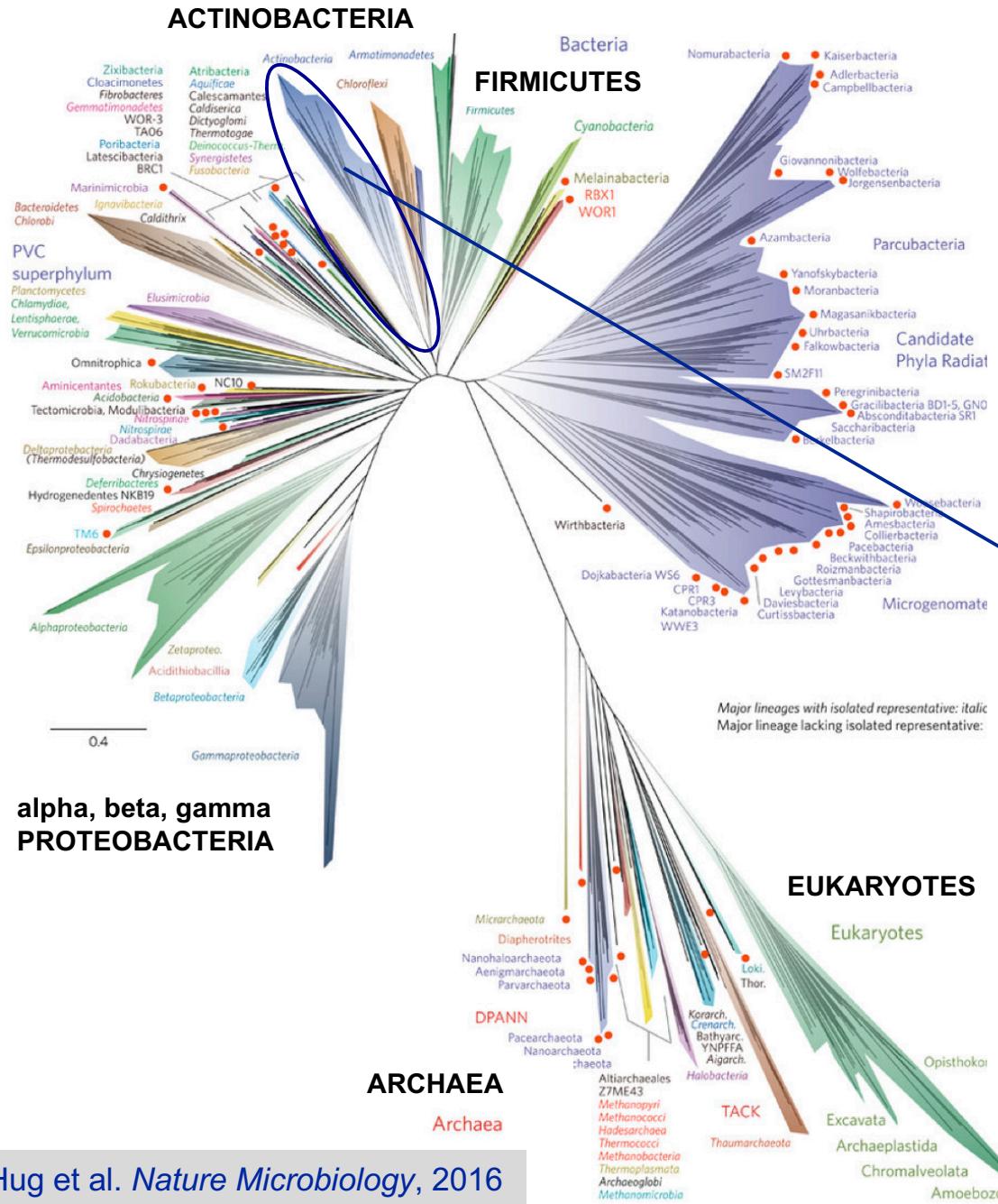
# **The evolution of the tuberculosis agent: a genome-based approach to mycobacterial evolution and pathogenesis**

**Roland Brosch**

**Unit for Integrated Mycobacterial Pathogenomics**

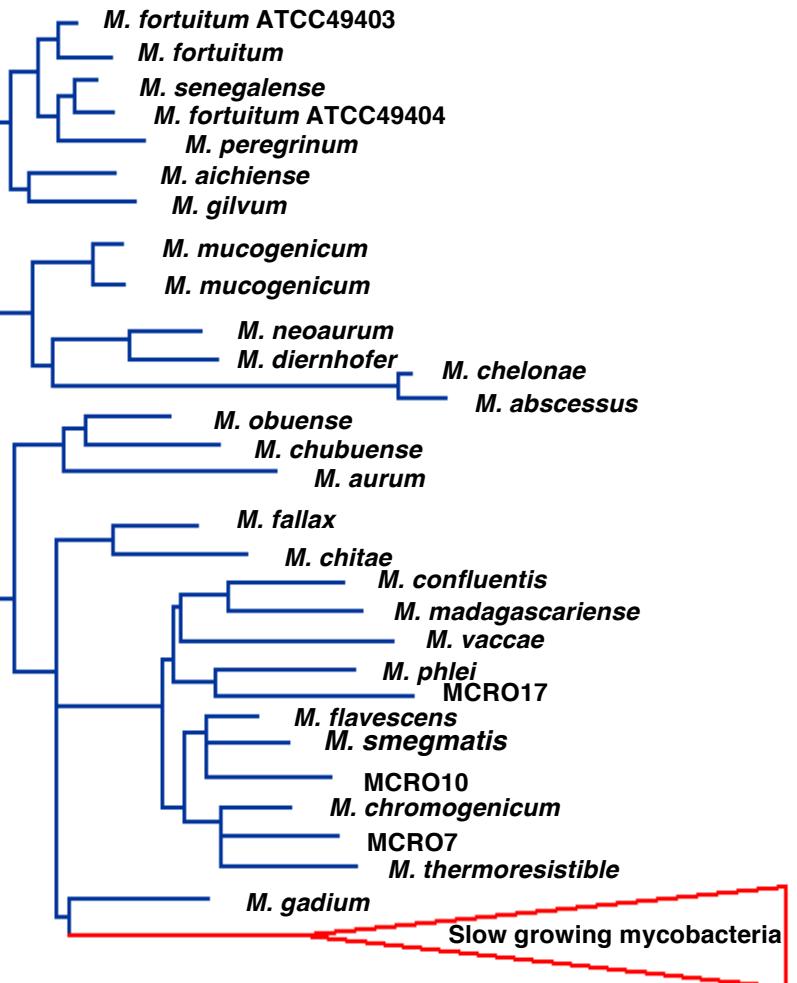
**Bioinformatics and Genome Analyses Course  
Institut Pasteur Tunis, Tunisia. September 18 – December 15, 2017**





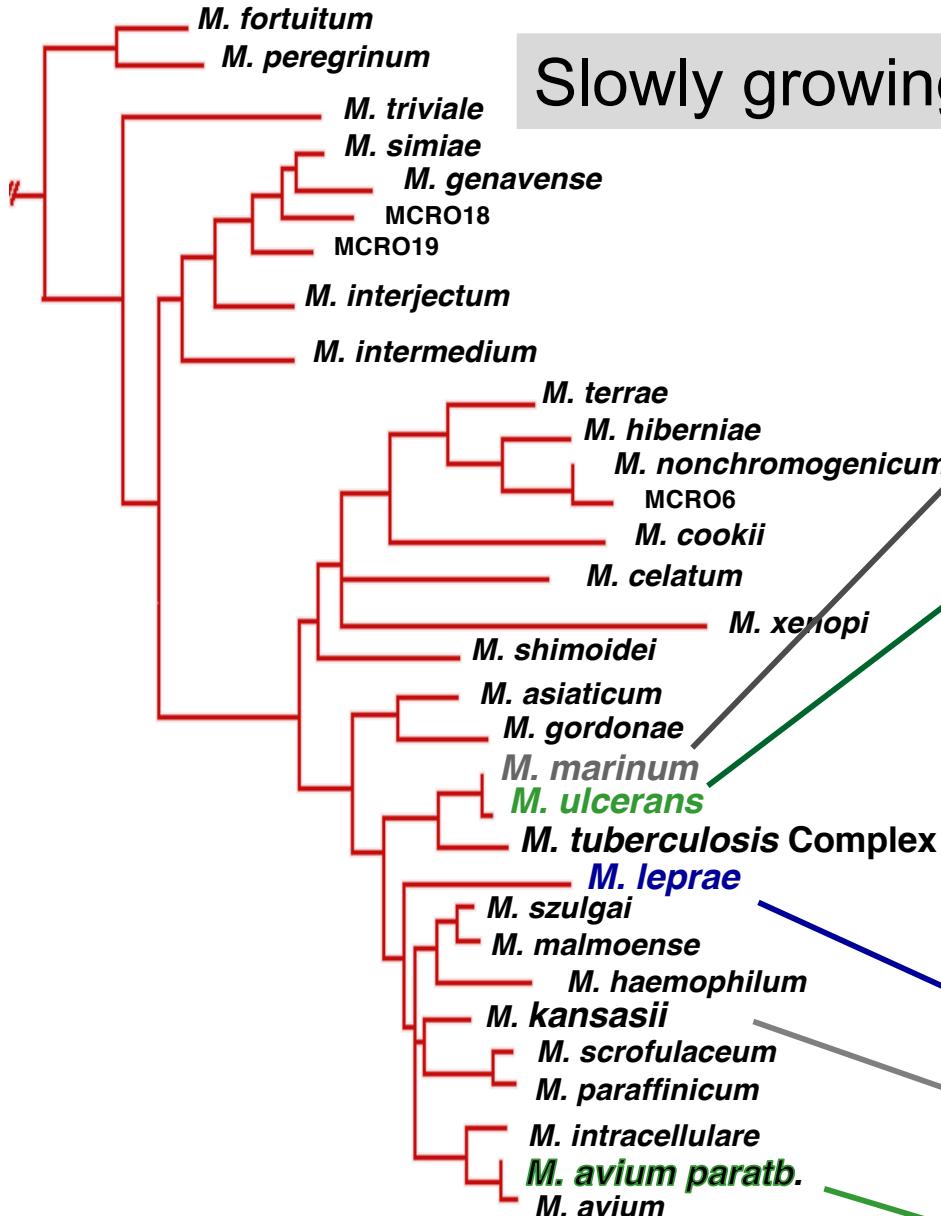
# Mycobacteria are part of the phylum Actinobacteria

consist of fast-growers and slow-growers



16S rRNA tree - Springer et al., *J. Clin. Microbiol.*, (1996)

# Slowly growing mycobacteria harbour pathogens



• fish pathog & human opportun. pathogen

• etiological agent of Buruli ulcer

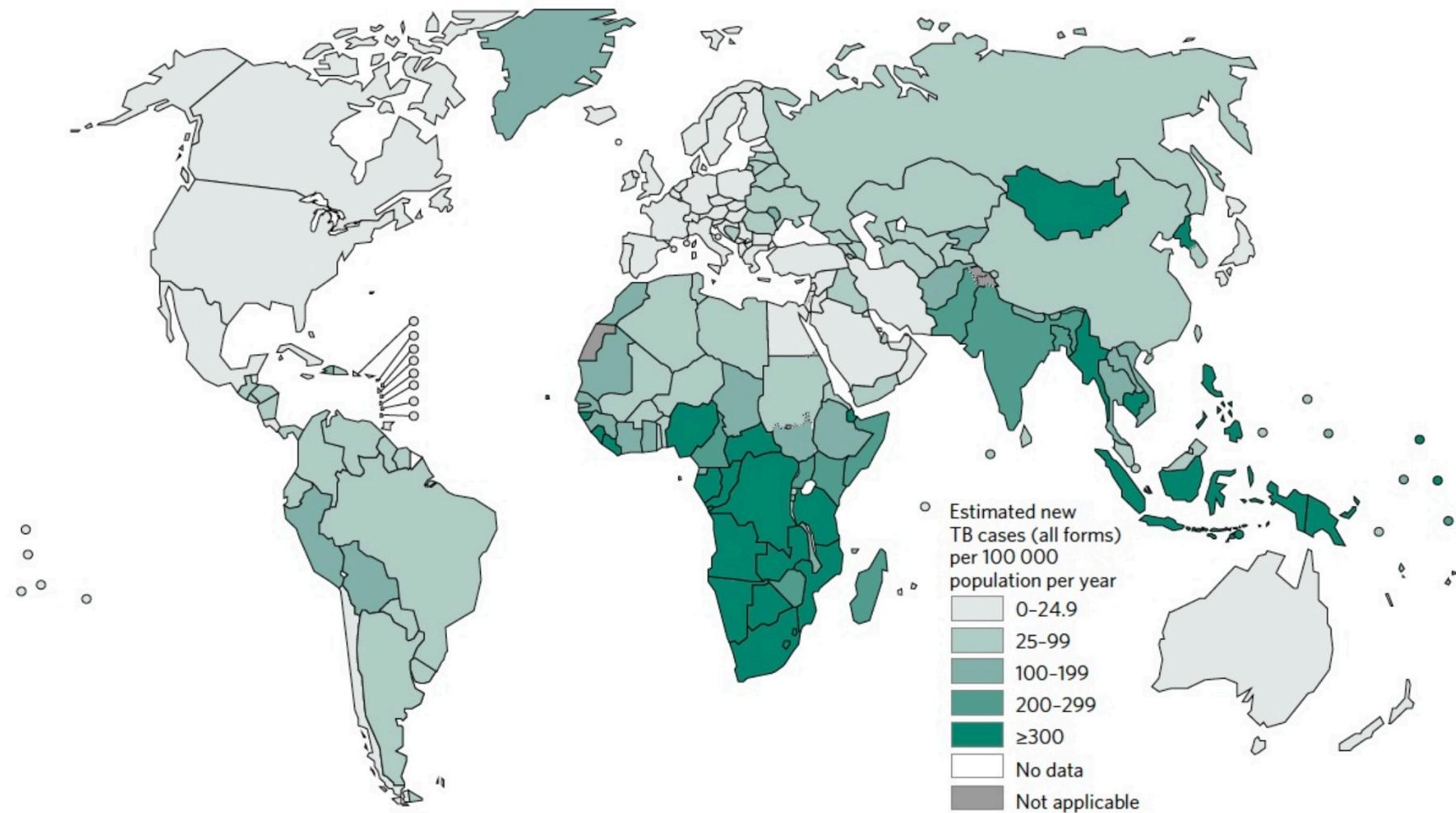
• etiological agent of Tuberculosis  
(in different mammalian species)

• etiological agent of Leprosy

• human opportun. pathogen

• agent of Johne's disease

## Geographical distribution of tuberculosis



**2015 Deaths: 1.8 million (> 4000 per day); New cases 10 million;  
leading killer of people with HIV infection**

## Pulmonary tuberculosis, advanced - chest x-ray

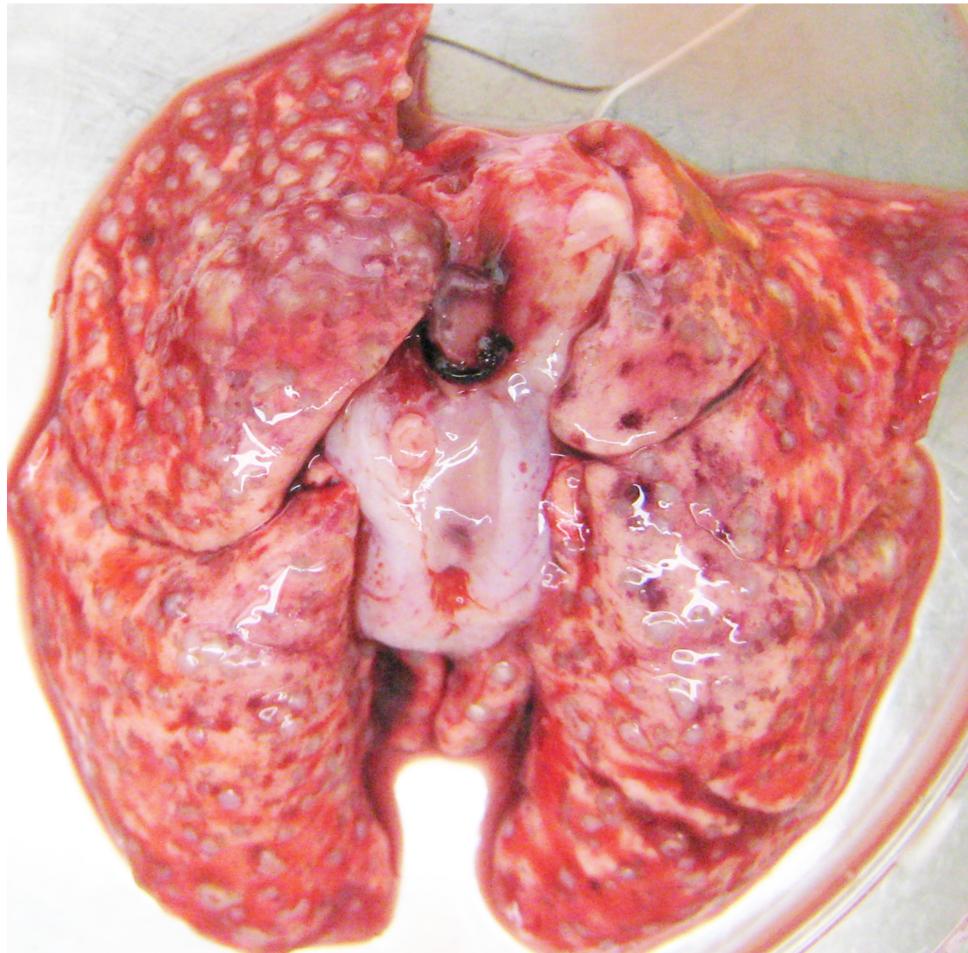


There are multiple light areas (opacities) of varying size that run together (coalesce). Arrows indicate the location of cavities within these light areas.

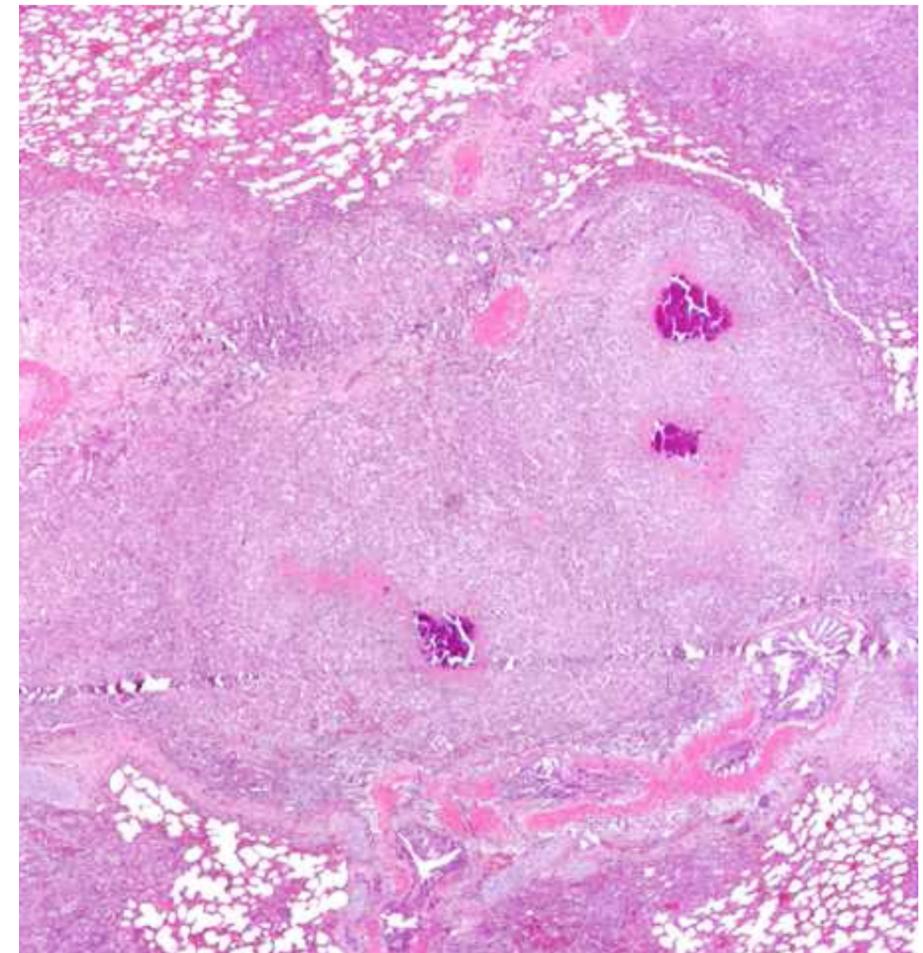
ADAM

<https://medlineplus.gov/ency/imagepages/1607.htm>

# How did *M. tuberculosis* evolve into such a key pathogen ?



Guinea pig lungs, infected by *M. tuberculosis*



Granuloma in guinea pig lung (4x)

## Main questions:

**Genomics enabled to get new insights into  
the evolution of *M. tuberculosis*:**



released 1998

Cole ST, Brosch R, Parkhill J, + et al.,  
(1998), Nature 393, 537-544.

# Genome of *M. tuberculosis* H37Rv

4,411,529 bp     GC 65.6% ≈ 4000 genes

Varied metabolic potential: aerobic/microaeroph./anaerobic

11 Serine Threonine Protein Kinases,

11 two component systems, eg. PhoP/R

Abundance of genes involved in lipid metabolism

Large gene families - PE/PPE, ESAT-6, *mmpL*



The Wellcome Trust  
Sanger Institute

# <http://genolist.pasteur.fr/TubercuList/>



# <https://mycobrowser.epfl.ch/>

https://mycobrowser.epfl.ch/genes/Rv0262c

Hôtels : Paris. Ré... WEBcampus Protein BLAST: se... Pathogénomique ... Google http://webcampu... news.ORF.at MaGe Login Mag... Sytadin : l'état du...

Mycobrowser Mutants Releases About M. tuberculosis Gene name or function Search Feedback

Gene Rv0262c in *Mycobacterium tuberculosis* H37Rv

Available Tracks

filter tracks

Annotation  
 GC Content  
 Reference sequence

Essentiality (DeJesus et al. 2017) 4

Essential regions  
 Growth advantage regions  
 Growth defect regions  
 Non-essential regions

RNA-Seq 2

Stationary phase (Uplekar et al. 2013) 2

Normalized coverage

Genome Track View Help Share Fullscreen

0 500,000 1,000,000 1,500,000 2,000,000 2,500,000 3,000,000 3,500,000 4,000,000

NC\_000962.3 NC\_000962.3:313308..315853 (2 Go)

313,750 315,000

Reference sequence Zoom in to see sequence Zoom in to see sequence

Annotation aac Rv0263c Rv026

virulence, detoxification, adaptation lipid metabolism conserved hypotheticals pseudogenes information pathways cell wall and cell processes stable RNAs insertion seqs and phages PE/PPE intermediary metabolism and respiration unknown regulatory proteins

Go to browser

General annotation

Type CDS

Function Confers resistance to aminoglycosides (gentamicin, tobramycin, dibekacin, netilmicin, and 6'-N-ethylnetilmicin).

Product Aminoglycoside 2'-N-acetyltransferase Aac (Aac(2')-IC)

Gene summary information

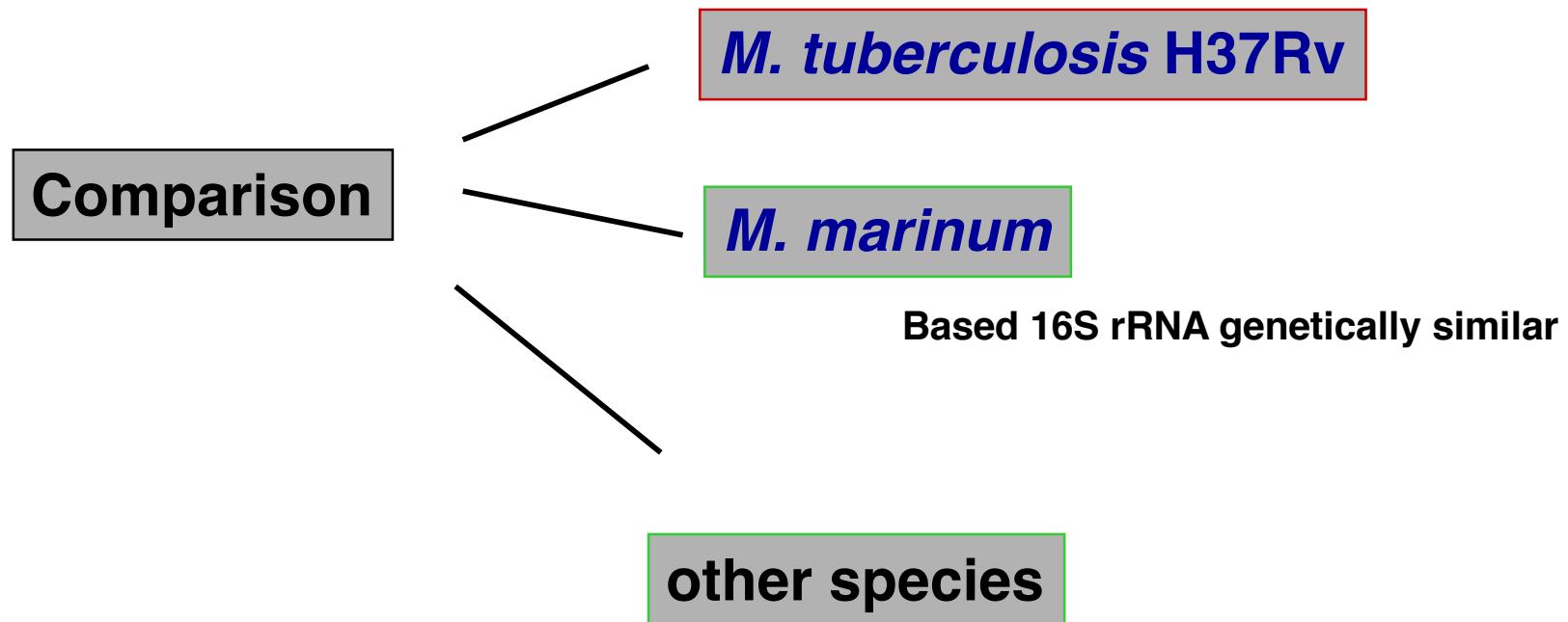
Gene name aac  
Gene length 546 bp  
Identifier Rv0262c

EPFL  
ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

©2017 EPFL, 1015 Lausanne  
Contact Admin

# Comparative genomics between different species

To extract the biological information from genome sequence data

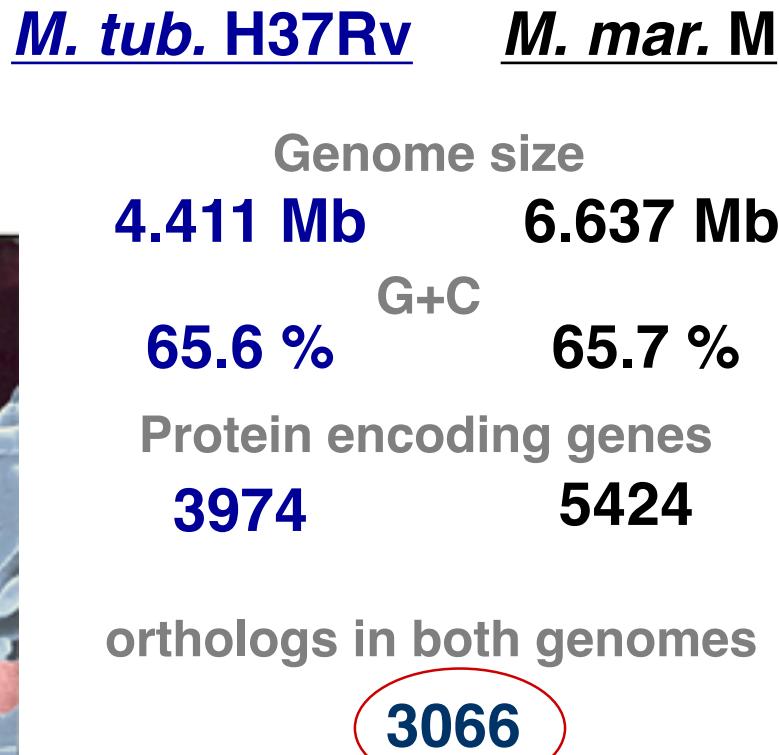


... allows insights into the more “distant” events ...

# Genome comparison *M. tuberculosis* - *M. marinum*



Cole et al., (1998), *Nature*



Letter

Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*

Timothy P. Stinear,<sup>1,10</sup> Torsten Seemann,<sup>2</sup> Paul F. Harrison,<sup>2</sup> Grant A. Jenkin,<sup>1</sup> John K. Davies,<sup>1</sup> Paul D.R. Johnson,<sup>3</sup> Zahra Abdellah,<sup>4</sup> Claire Arrowsmith,<sup>4</sup> Tracey Chillingworth,<sup>4</sup> Carol Churcher,<sup>4</sup> Kay Clarke,<sup>4</sup> Ann Cronin,<sup>4</sup> Paul Davis,<sup>4</sup> Ian Goodhead,<sup>4</sup> Nancy Holroyd,<sup>4</sup> Kay Jagels,<sup>4</sup> Angela Lord,<sup>4</sup> Sharon Moule,<sup>4</sup> Karen Mungall,<sup>4</sup> Halina Norbertczak,<sup>4</sup> Michael A. Quail,<sup>4</sup> Ester Rabinowitz,<sup>4</sup> Danielle Walker,<sup>4</sup> Brian White,<sup>4</sup> Sally Whitehead,<sup>4</sup> Pamela L.C. Small,<sup>5</sup> Roland Brosch,<sup>6</sup> Lalita Ramakrishnan,<sup>7</sup> Michael A. Fischbach,<sup>8</sup> Julian Parkhill,<sup>4</sup> and Stewart T. Cole<sup>9</sup>

<sup>1</sup>Department of Microbiology, Monash University, Clayton 3800, Australia; <sup>2</sup>Victorian Bioinformatics Consortium, Monash University, Clayton 3800, Australia; <sup>3</sup>Department of Infectious Diseases, Austin Hospital, Heidelberg 3080, Australia; <sup>4</sup>Welcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; <sup>5</sup>Department of Microbiology, University of Tennessee, M409 Wabers Science, Knoxville, Tennessee 37996-0445, USA; <sup>6</sup>Institut Pasteur, UP Pathogénomique Mycobactérienne Intégrée, Paris 75015, France; <sup>7</sup>Department of Microbiology, University of Washington, Seattle, Washington 98195, USA; <sup>8</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA; <sup>9</sup>Global Health Institute, EPFL, Station 15, CH-1015, Lausanne, Switzerland

*Mycobacterium marinum*, a ubiquitous pathogen of fish and amphibia, is a near relative of *Mycobacterium tuberculosis*, the etiologic agent of tuberculosis in humans. The genome of the M strain of *M. marinum* comprises a 6,636,827-bp circular chromosome with 5424 CDS, 10 prophages, and a 23-kb mercury-resistance plasmid. Prominent features are the very large number of genes (57) encoding polyketide synthases (PKSs) and nonribosomal peptide synthases (NRPSs) and the most extensive repressive operons reported for mycobacteria-restricted PE and PPE proteins, and related ESX secretion systems. Some of the new genes confer a metabolic advantage seen to have been acquired horizontally. *M. marinum* is unique widely as a model organism to study *M. tuberculosis* pathogenesis, and genome comparisons confirm the close genetic relationship between these two species, as they share 3000 orthologs with an average amino acid identity of 85%. Comparisons with the more distantly related *Mycobacterium avium* subspecies *paratuberculosis* and *Mycobacterium smegmatis* reveal how an ancestral mycobacterium evolved into *M. tuberculosis* and *M. marinum*. *M. tuberculosis* has undergone genome downsizing and extensive lateral gene transfer to become a specialized pathogen of humans and other primates without retaining an environmental niche. *M. marinum* has maintained a large genome so as to retain the capacity for environmental survival while becoming a broad host range pathogen that produces disease strikingly similar to *M. tuberculosis*. The work described herein provides a foundation for using *M. marinum* to better understand the determinants of pathogenesis of tuberculosis.

Supplemental material is available online at [www.genome.org](http://www.genome.org). The sequence data from this study have been submitted to GenBank under accession nos. CRO00854 (chromosome) and CRO00895 (pM23 plasmid).

In 1924, Joseph D. Atkinson isolated a Mycobacterium from tubercles observed predominantly in the spleen and liver of diseased fish that had died in the Philadelphia Aquarium and named it *Mycobacterium marinum* (Atkinson 1927). *M. marinum* was subsequently shown to also be a human pathogen when it was isolated again much later in a swimming pool-associated outbreak of human granulomatous skin lesions, although in this report the *Mycobacterium* was mistakenly given a new species name, *Mycobacterium balnei* (Lindl and Norden 1954), a name that is no longer used.

Several phylogenetic studies have shown that *M. marinum* is most closely related to *Mycobacterium ulcerans* (99.7% nucleotide identity) and then *M. tuberculosis* (>89% nucleotide identity) (Tenjin et al. 1998; Stinear et al. 2000; Devulder et al. 2005; Guy van Pittius et al. 2006; Yip et al. 2007). Multi-locus sequence typing of *M. marinum* isolates from around the world and our own suggest clearly show that *M. ulcerans* has recently evolved from a *M. marinum* progenitor by acquisition of a virulence plasmid and reductive evolution (Stinear et al. 2004, 2007; Yip et al. 2007). *M. ulcerans* cause the devastating human skin disease Buruli ulcer, which is quite distinct in its pathology from other mycobacterial diseases, primarily because of the production from

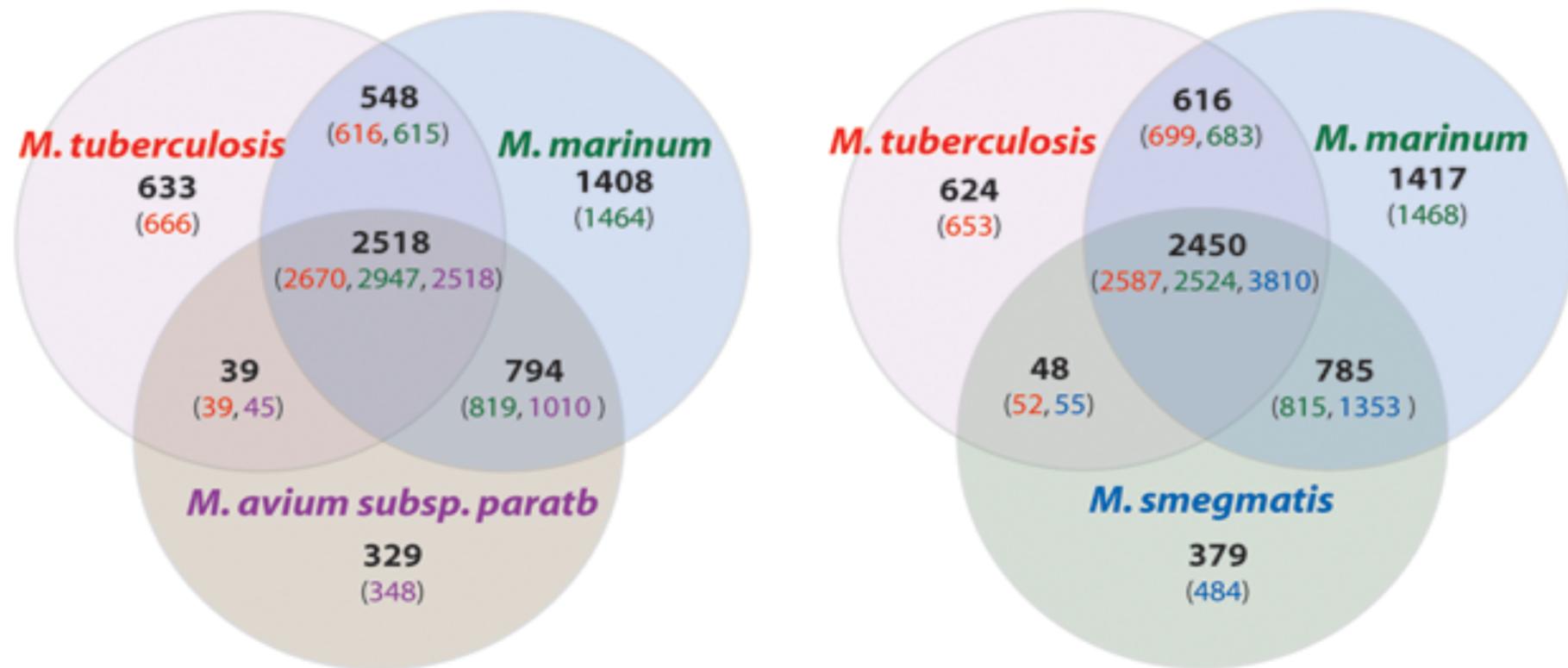
18000-000 ©2008 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/08 www.genome.org

Genome Research 1  
www.genome.org

Stinear et al., (2008), *Genome Res.*



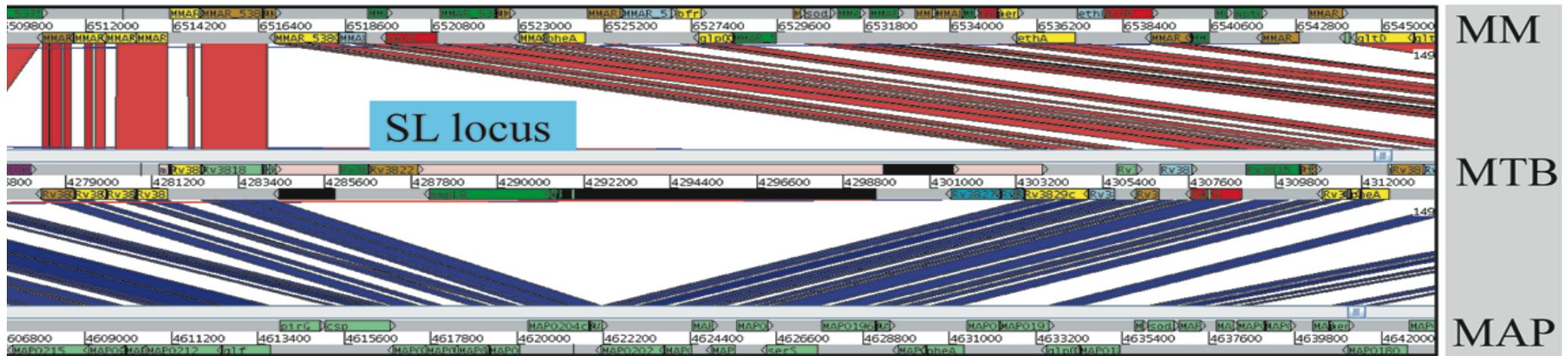
# Identification of ca. 600 *M. tuberculosis*-specific CDS



- Mycobacterial core genome (not including *M. leprae*) ~ 2500 genes
- At least 80 regions appeared by *M. tuberculosis* have been gained through HGT

Stinear *et al.*, 2008, *Genome Res.*

# Example of *M. tuberculosis*-specific CDSs

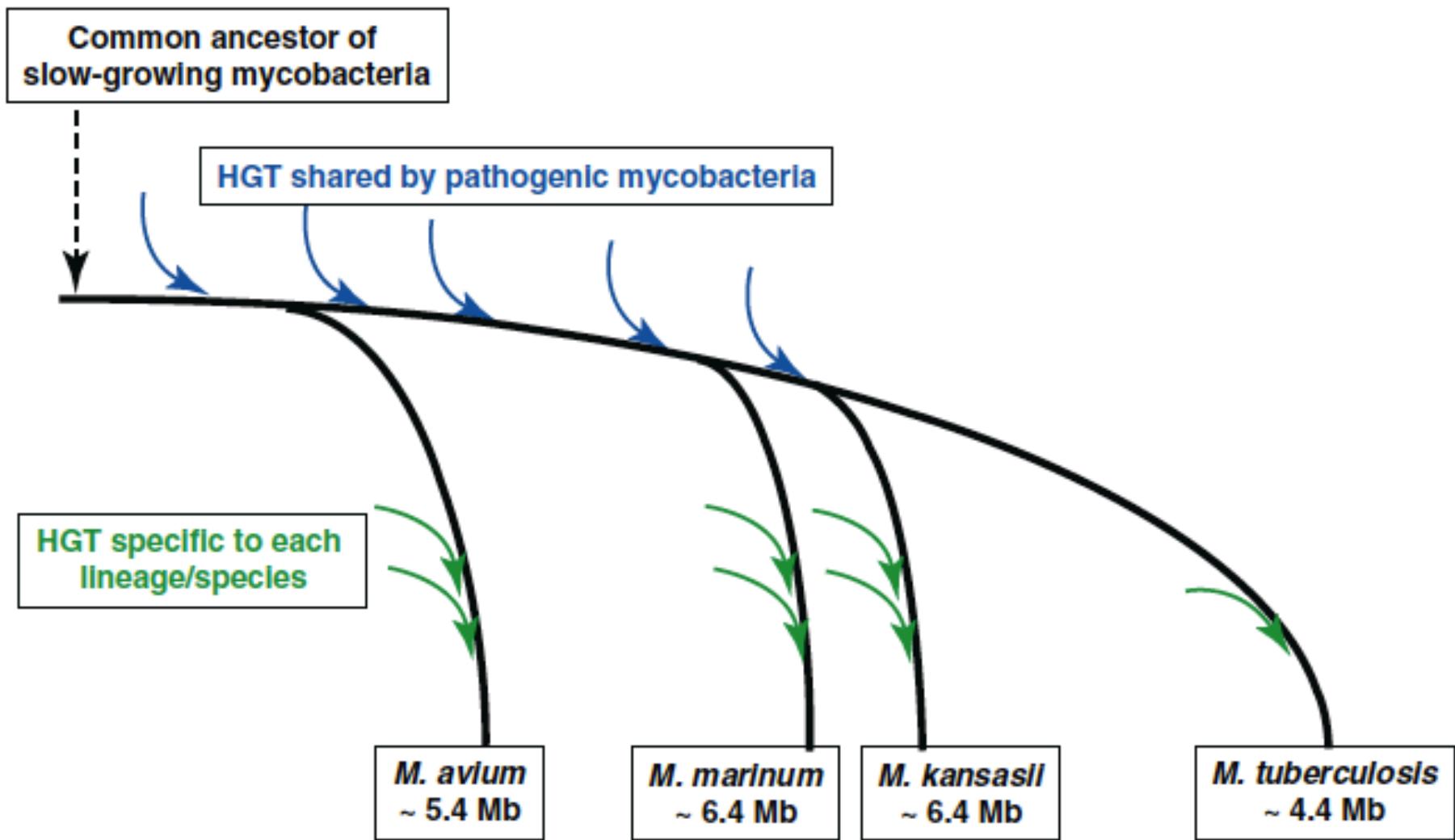


This three-genome comparison by the Artemis Comparison Tool (ACT) shows specific genes in *Mtb* that code for enzymes involved in the synthesis of Sulfo-Lipids.

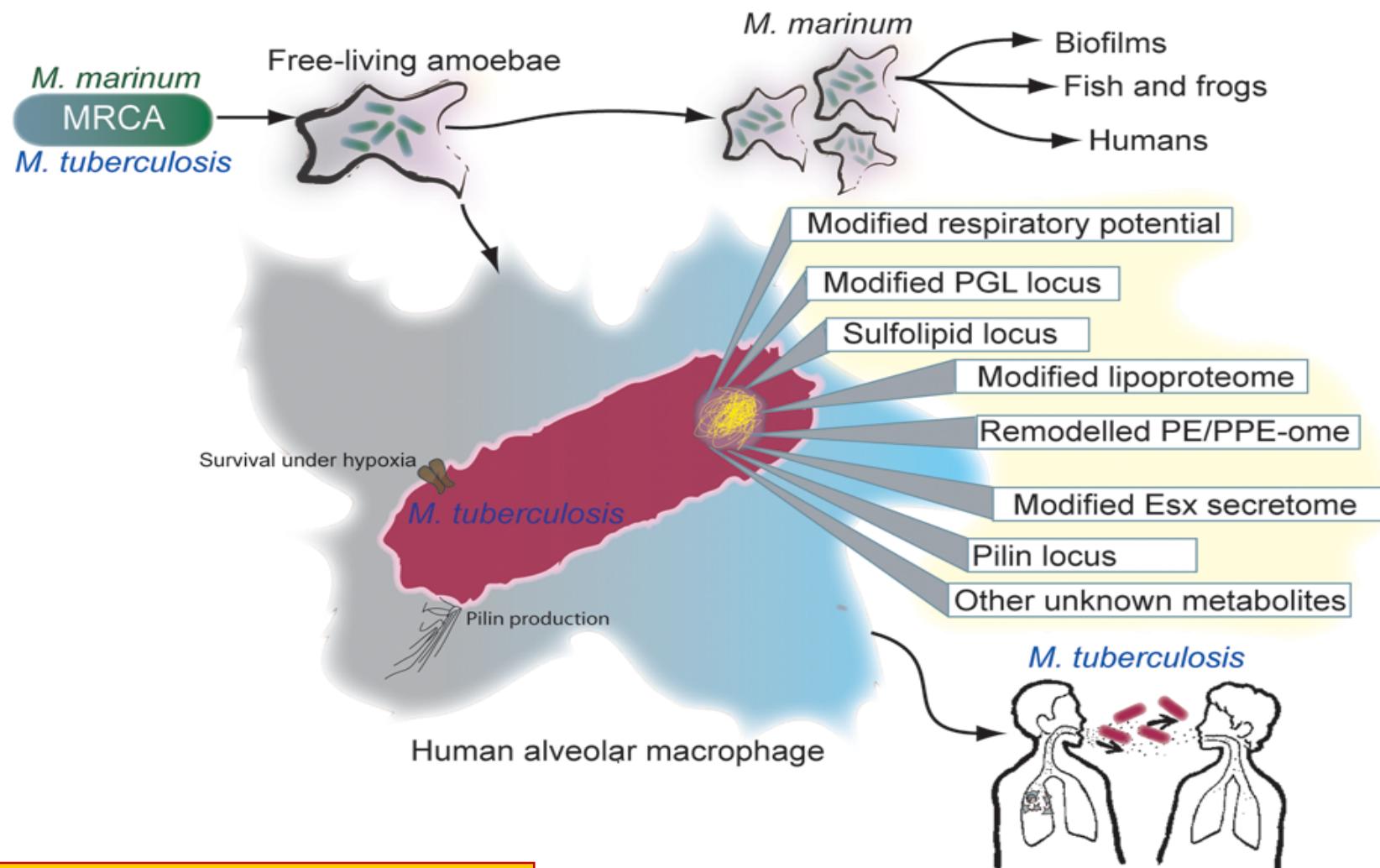
- other examples:
  - virS locus
  - pilin gene
  - fumarate reductase locus

Stinear et al., 2008, *Genome Res.*

## Situation is similar with respect to *M. kansasii*

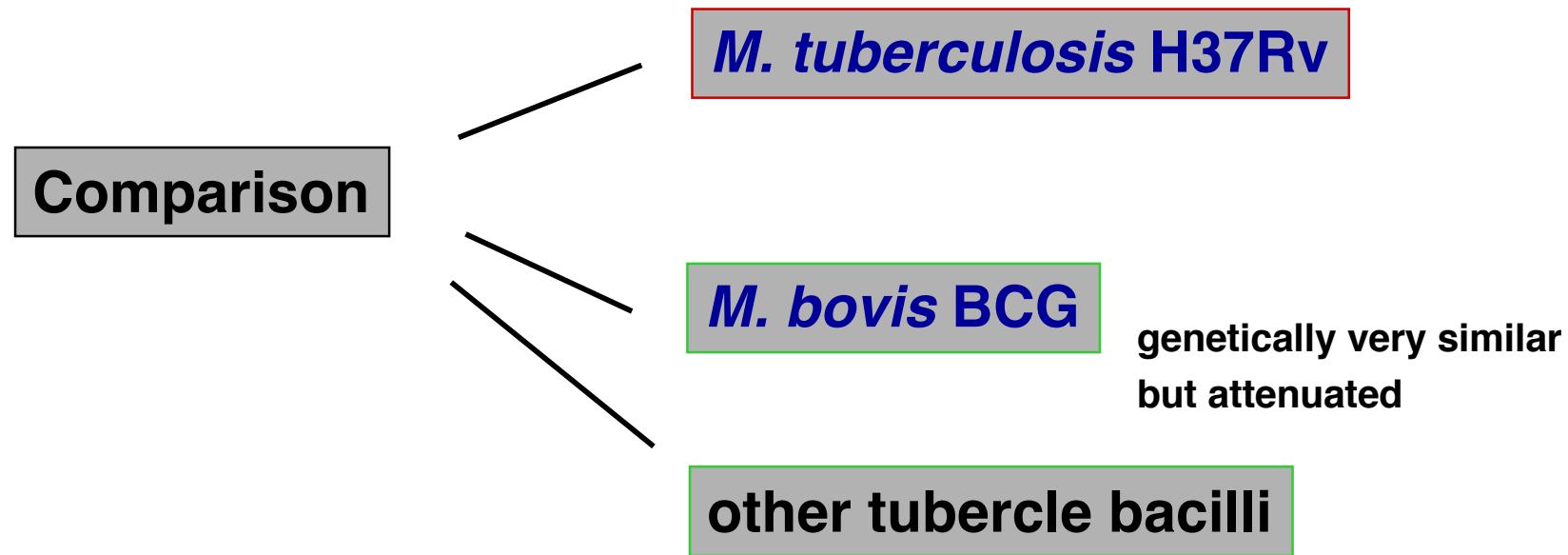


# ? possible early evolution of *M. tuberculosis*



Gordon et al., (2009), *Bioessays*

# Comparative genomics within the tubercle bacilli



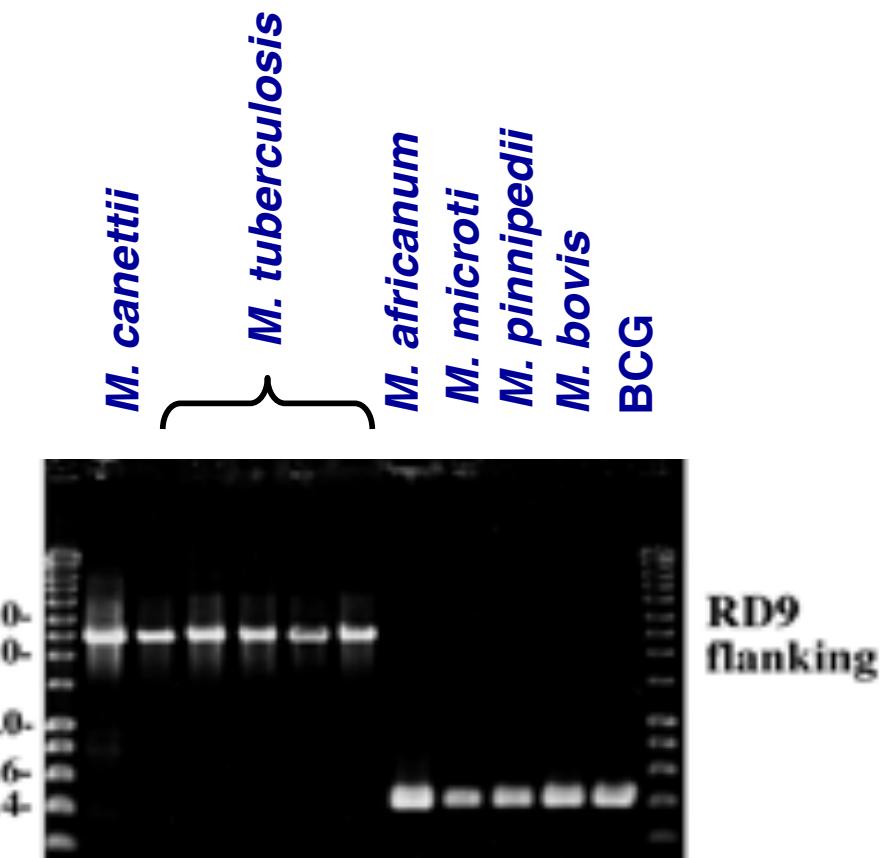
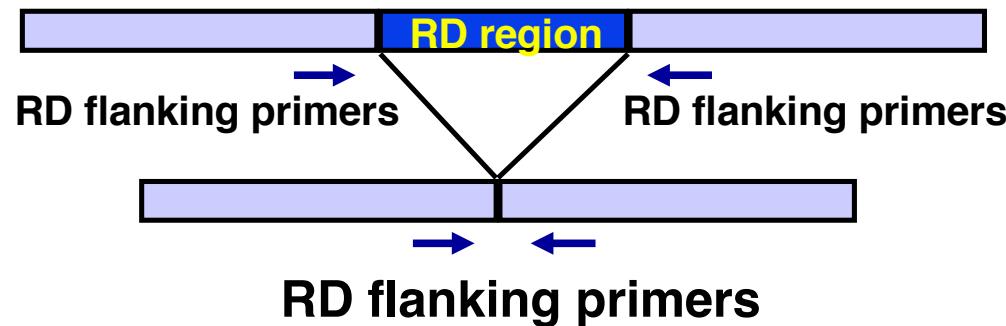
20 Regions of Difference (RD) described

... the more “recent” events ...

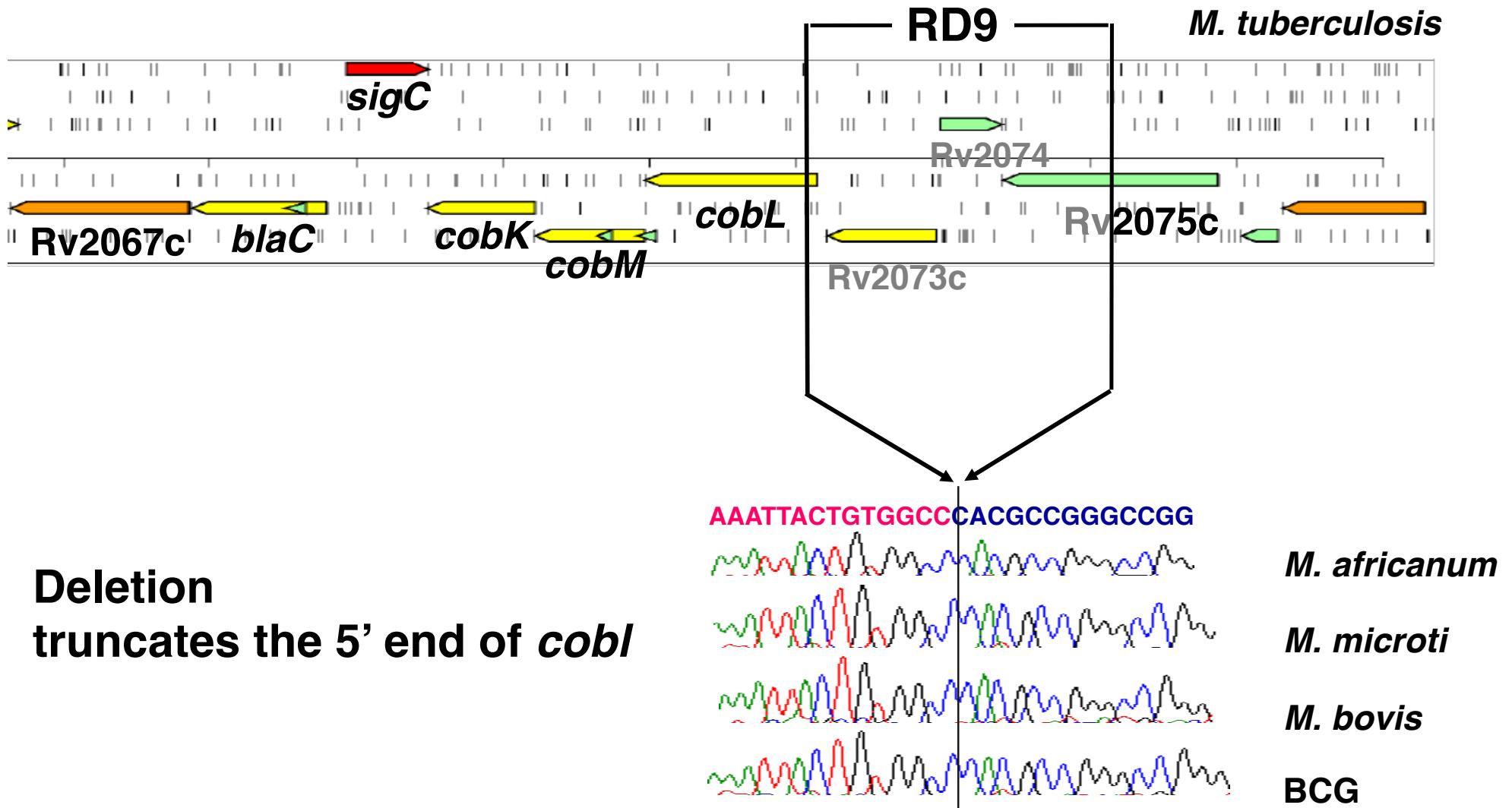
Gordon *et al.*, 1999 Mol Micro; Behr *et al.*, 1999 Science;

# Comparative genomics identifies Regions of Difference (RD)

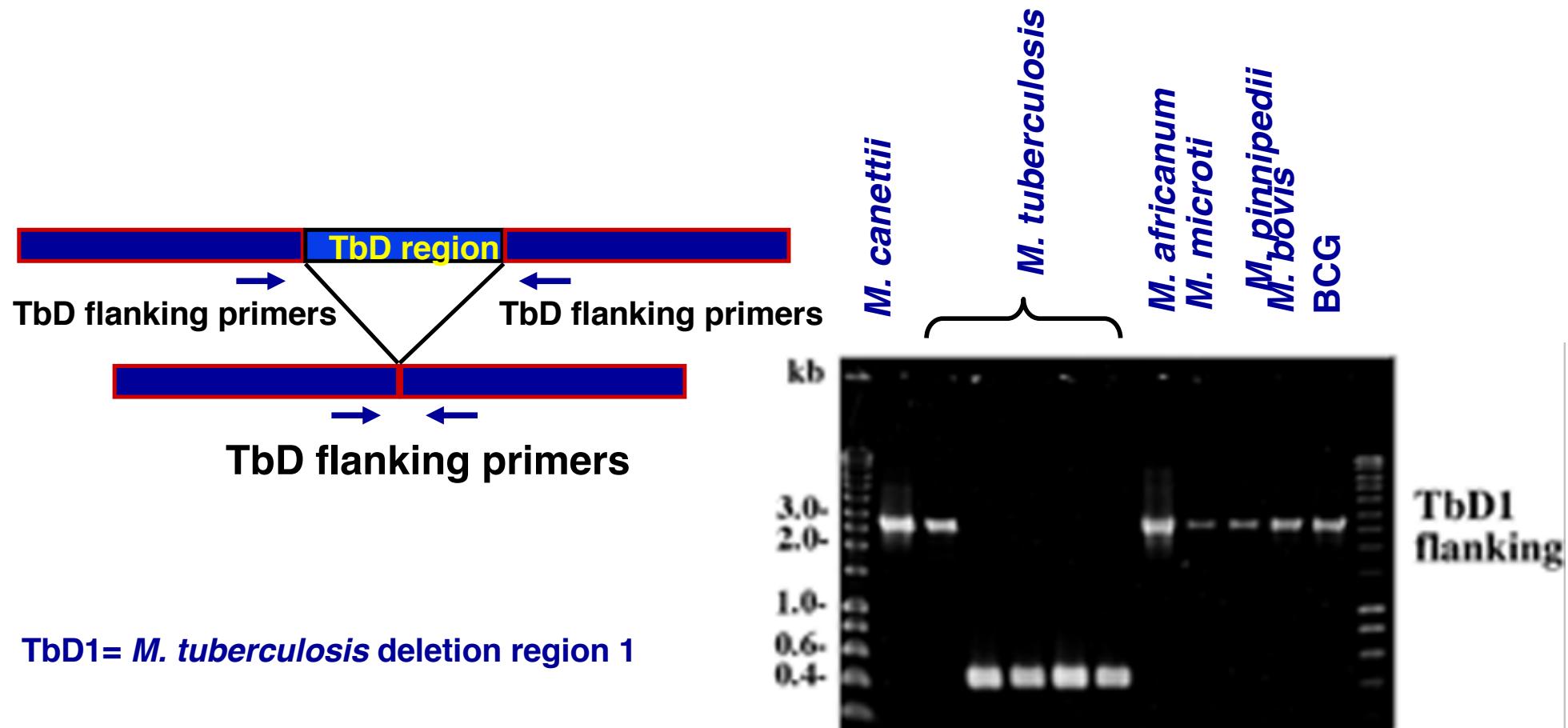
## Example of RD9



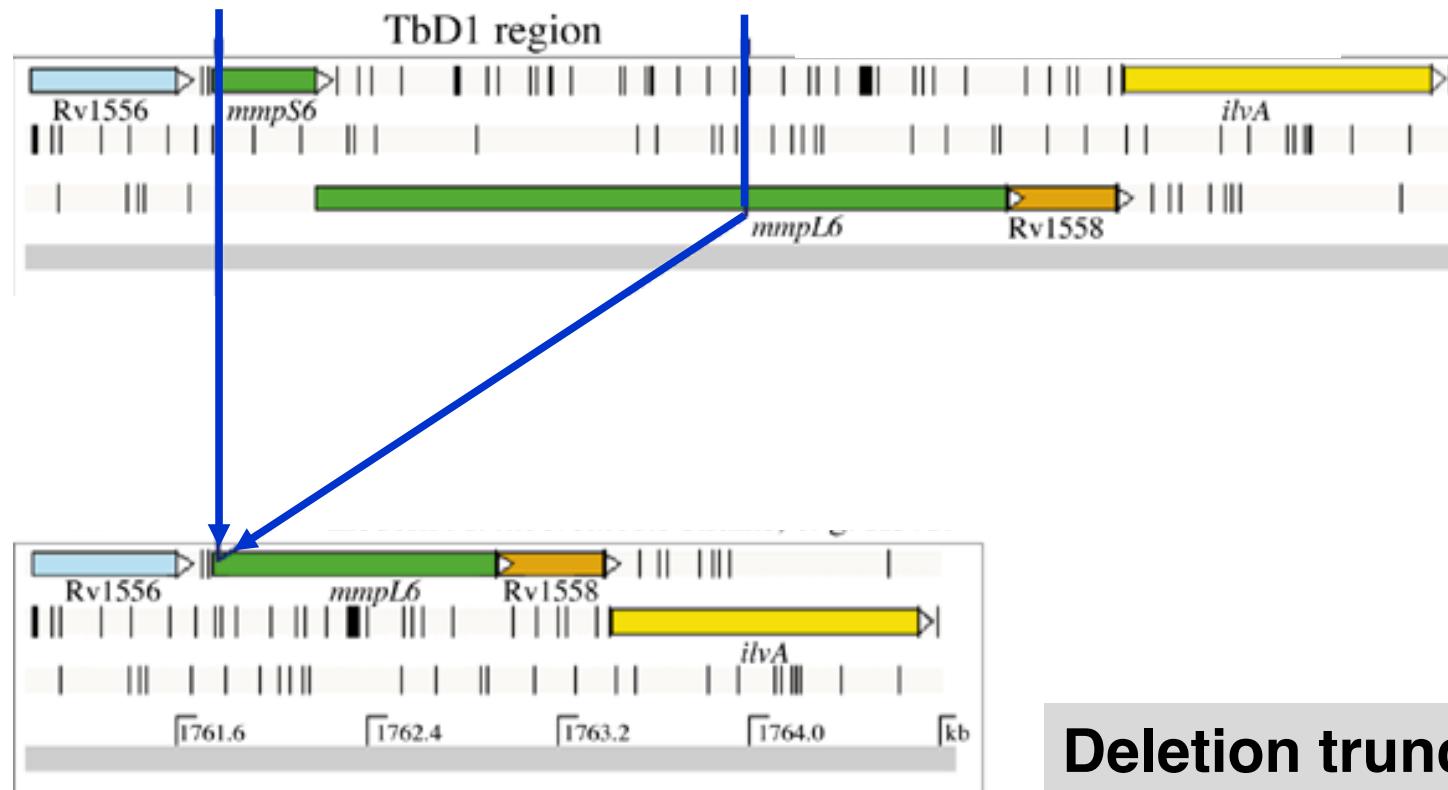
## RD9 region absent from *M. africanum* - *M. microti* - *M. bovis* lineage



# Comparative genomics identifies *M. tuberculosis* deletion region 1 (TbD1)



# Region TbD1 absent from many *M. tuberculosis* strains



*M. tuberculosis* H37Rv,  
*M. tuberculosis* CDC1551,  
*M. tuberculosis* Beijing 210  
*M. tuberculosis* C  
*M. tuberculosis* F11

*M. bovis*  
*M. africanum*  
*M. microti*  
BCG

“ancestral” *M. tub.*  
prevalent in SE Asia

Deletion truncates and  
fuses *mmpS6* and *mmpL6*  
to a new ORF in “modern”  
*Mtb* strains

# **Deletion events and the *M. tuberculosis* complex**

**most junction regions of RDs are located within genes**



**deletion events --> not insertion events**

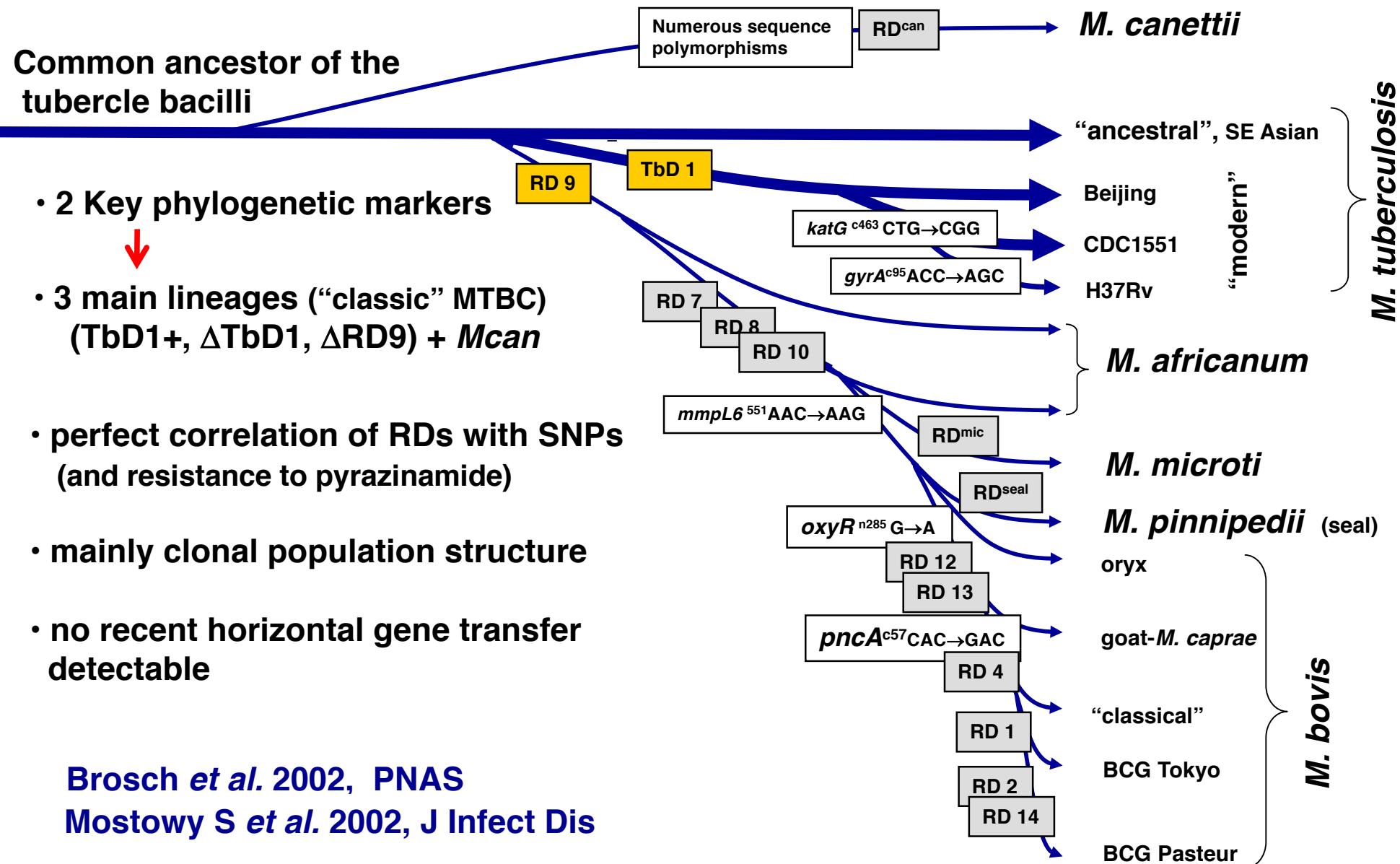


**conserved deletions  
among members of the *M. tuberculosis* complex**

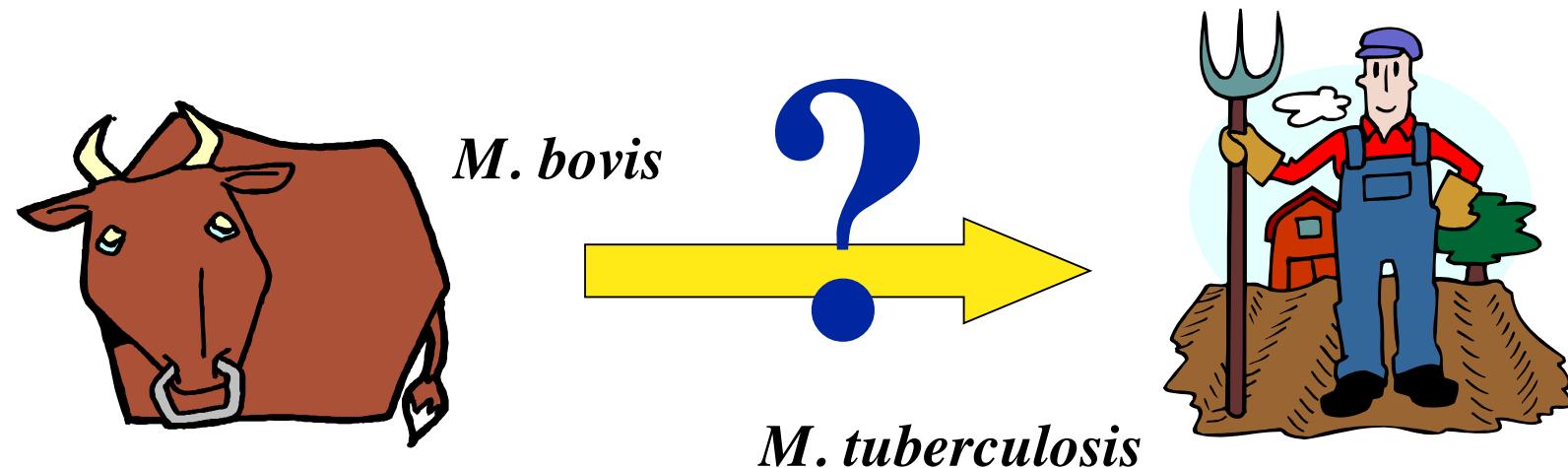


**most probably occurred  
in a common progenitor strain**

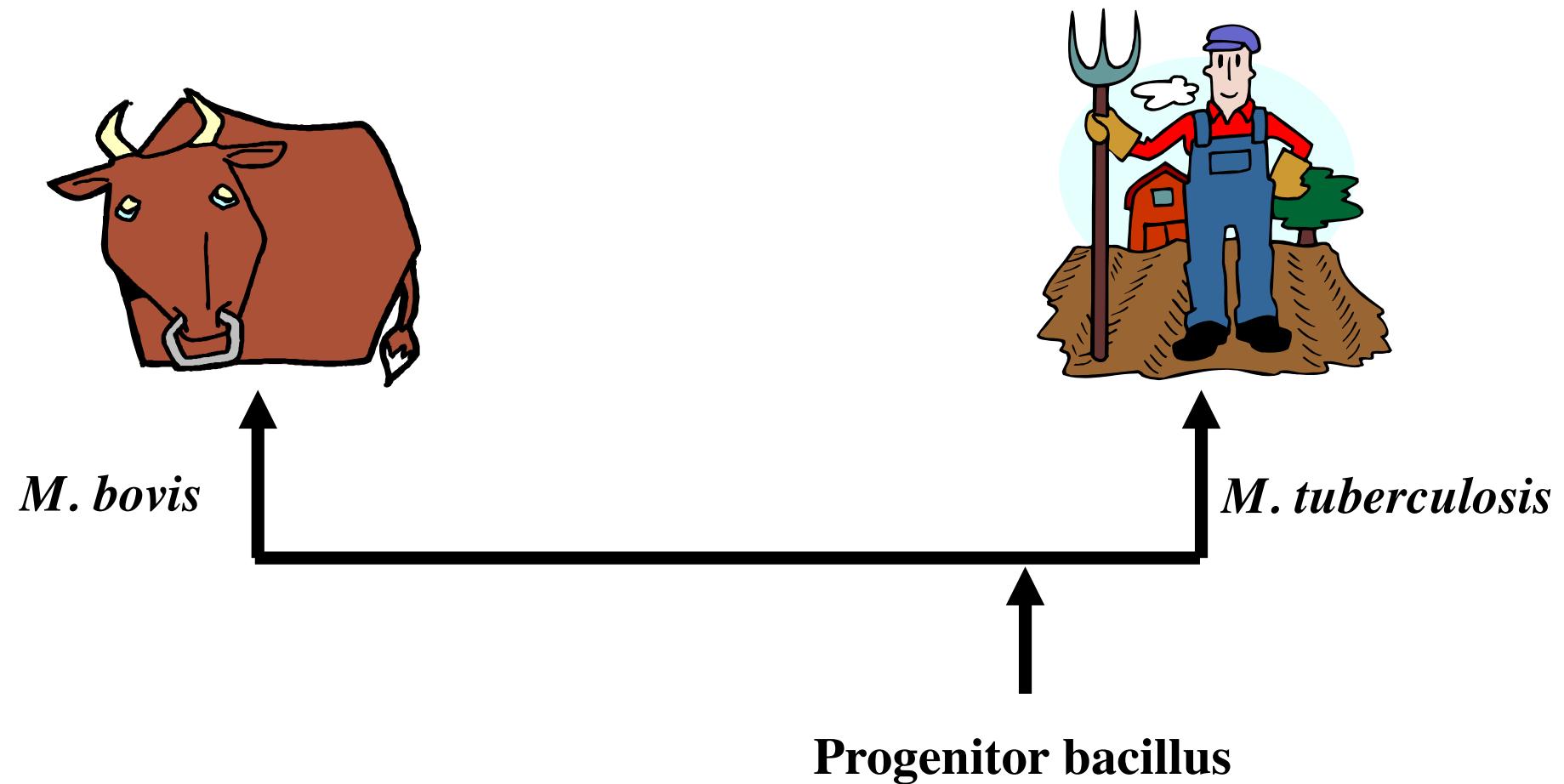
# Evolutionary pathway of the *M. tuberculosis* complex (MTBC)



**Textbook:**  
***M. tuberculosis* is derived from *M. bovis***



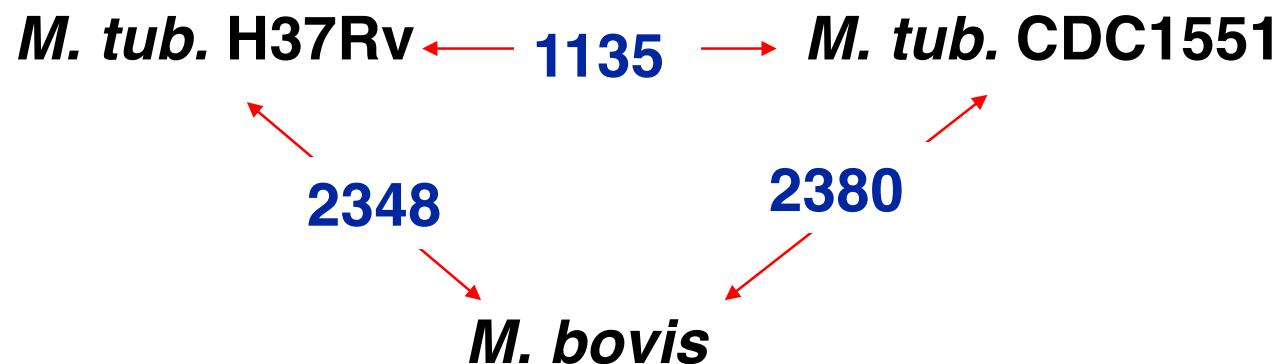
# Evolution of the *M. tuberculosis* complex



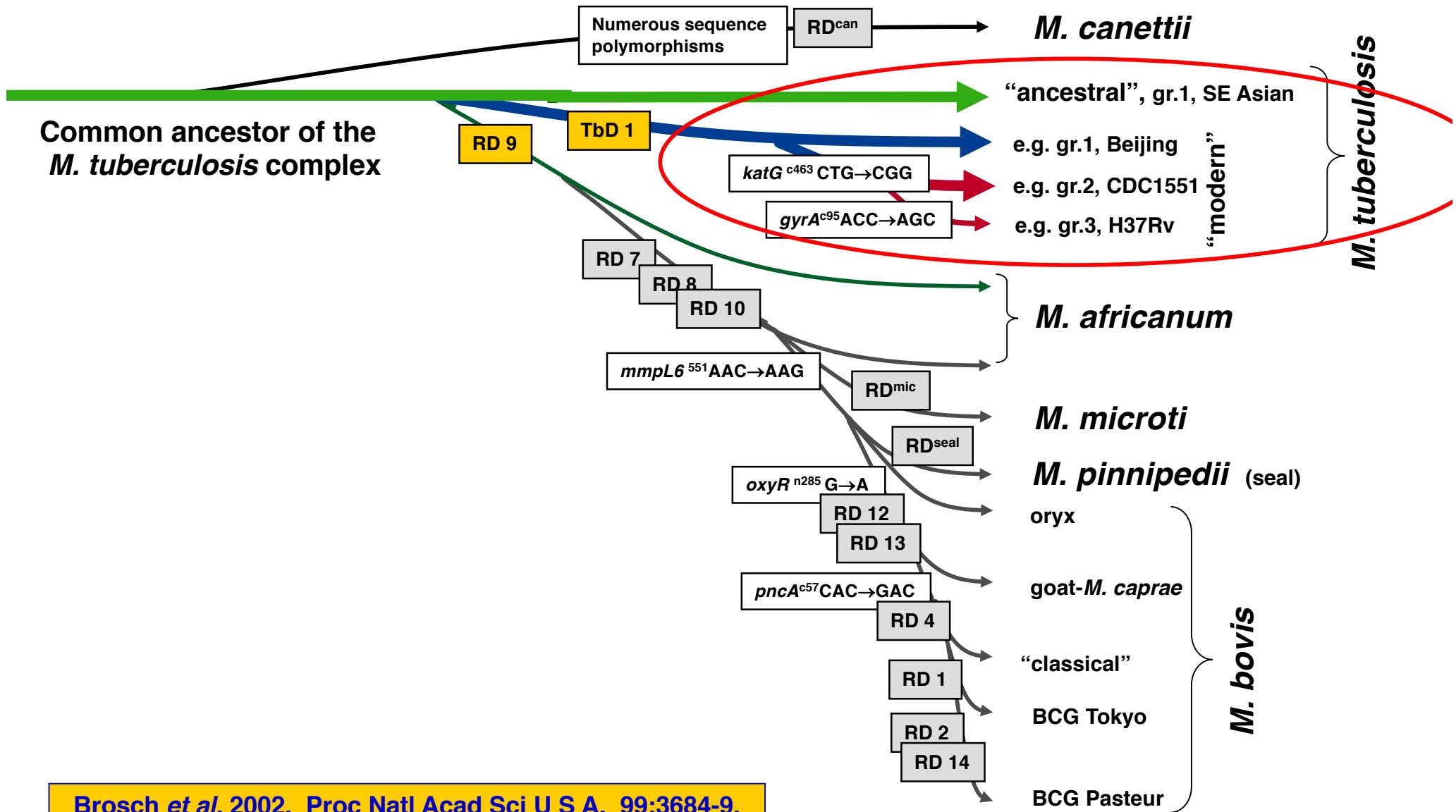
..... confirmed by genome analysis of *M. bovis*

<i>M. tuberculosis</i> H37Rv	4,411,532 bp
<i>M. tuberculosis</i> CDC1551	4,403,836 bp
<i>M. bovis</i> AF2122/97	4,345,492 bp

**S**ingle **N**ucleotide **P**olymorphisms



# Focus on the *M. tuberculosis* cluster



# Ancient-DNA analysis suggests that the oldest human TB case known in Britain was caused by a TbD1-deleted *Mtb* strain



FIG. 1. Excavation of burial 7 at Tarrant Hinton, showing the skeleton in situ.

*pncA169*  
~~IS1081 (113 bp)~~  
TbD1 flanking  
TbD1 internal

Iron age; 2,200 years old

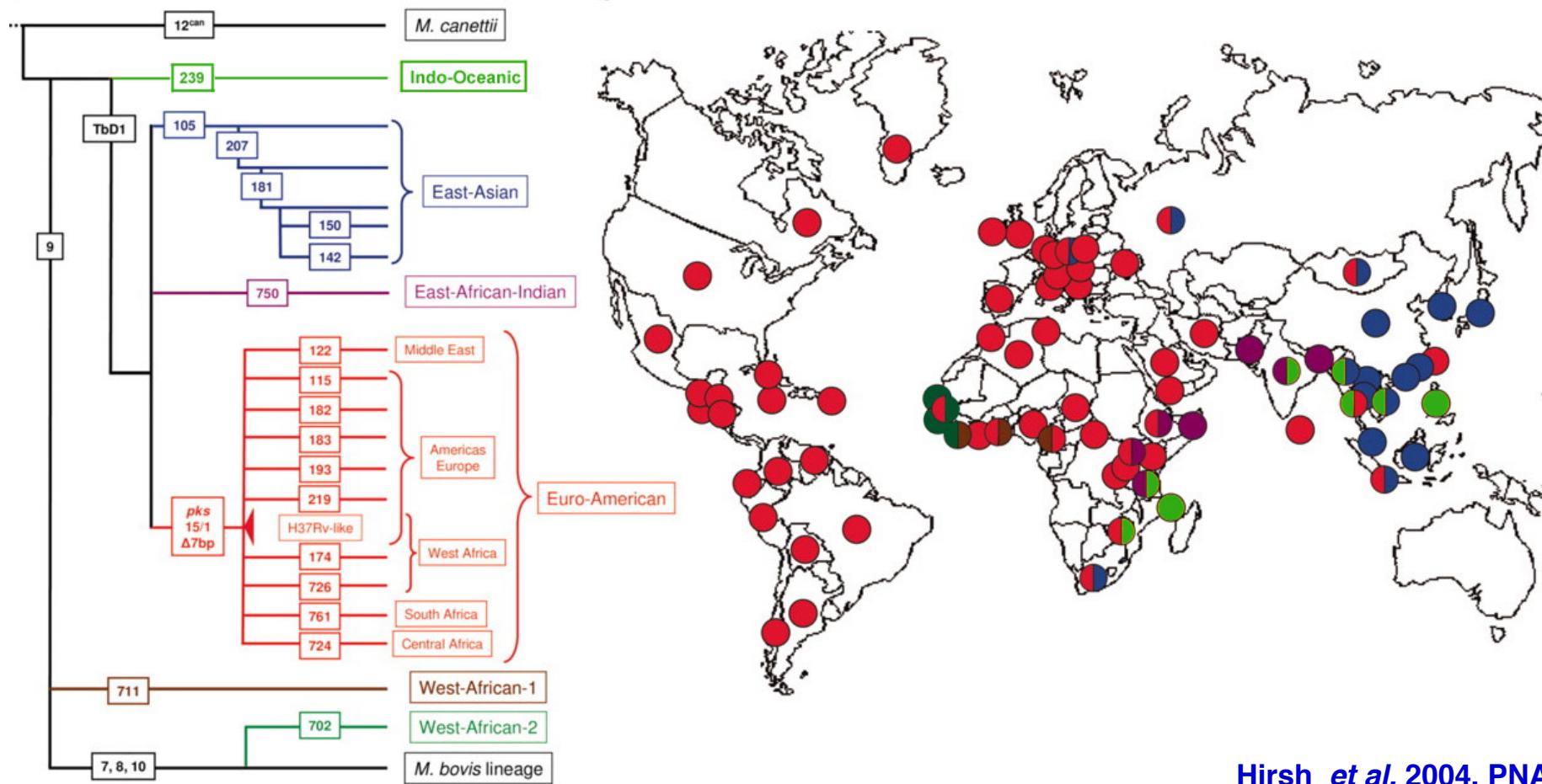
PCR results obtained from Tarrant Hinton burial 7

Result for <sup>a</sup> :			
Lumbar vertebra (LV2)	Rib 1	Rib 2	
+	+	—	
+G	+s	—	
+C	ND	ND	
+s	+s	+	
+s	ND	ND	
—	—	—	

<sup>a</sup> +, PCR positive; —, PCR negative. s denotes identity of PCR product confirmed by sequencing. G, guanine; C, cytosine; ND, not determined.

Taylor *et al.*, JCM, 2005

## RD-analyses (LSP) allowed to reveal stable association between strains of *M. tuberculosis* and their human host populations /geographical distribution



Hirsh *et al.* 2004, PNAS

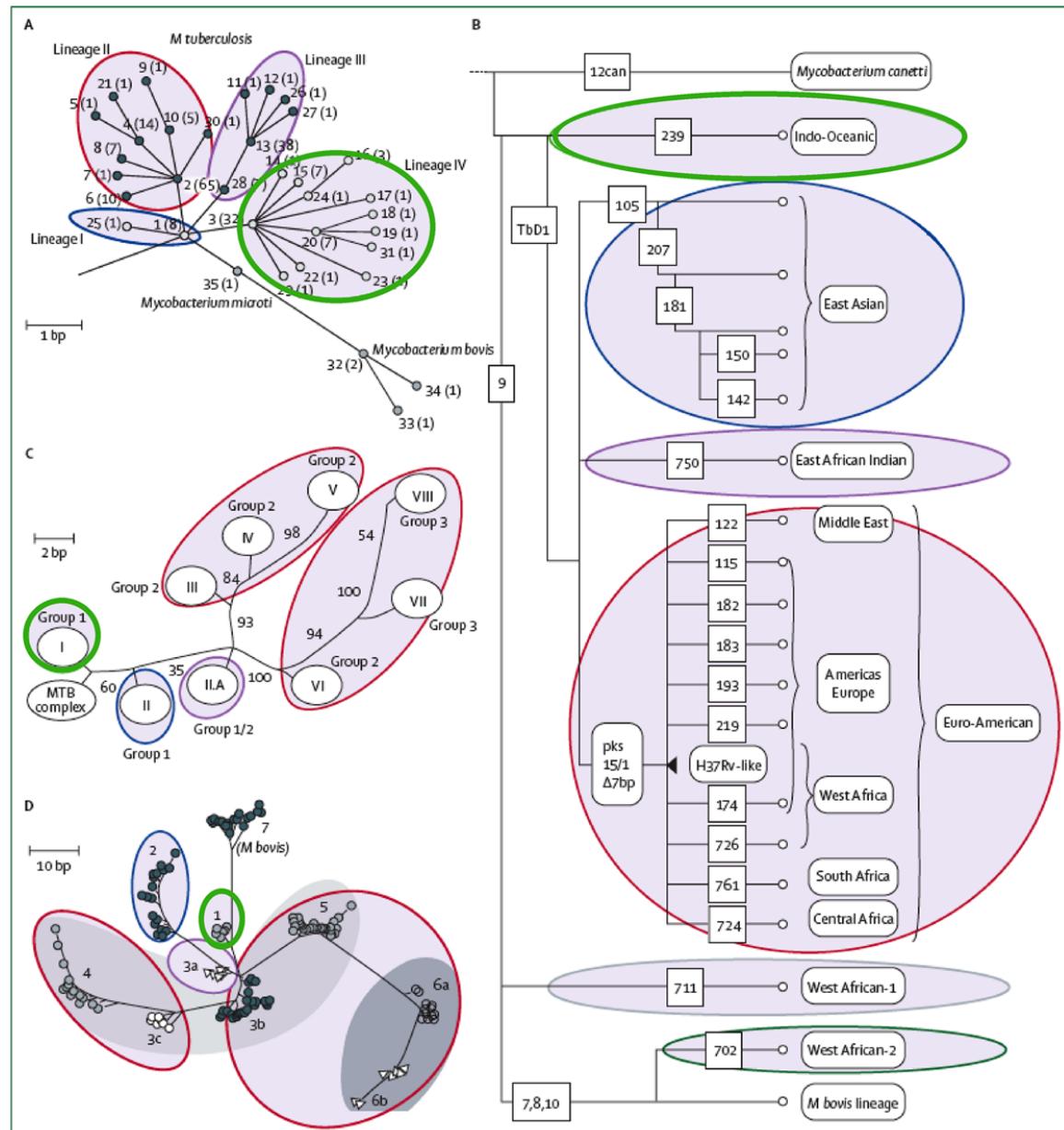
Gagneux *et al.* 2006, PNAS

# Strict correlation of RDs and SNPs

Baker *et al.*, 2004

Gutacker *et al.*, 2006

Filliol *et al.*, 2006



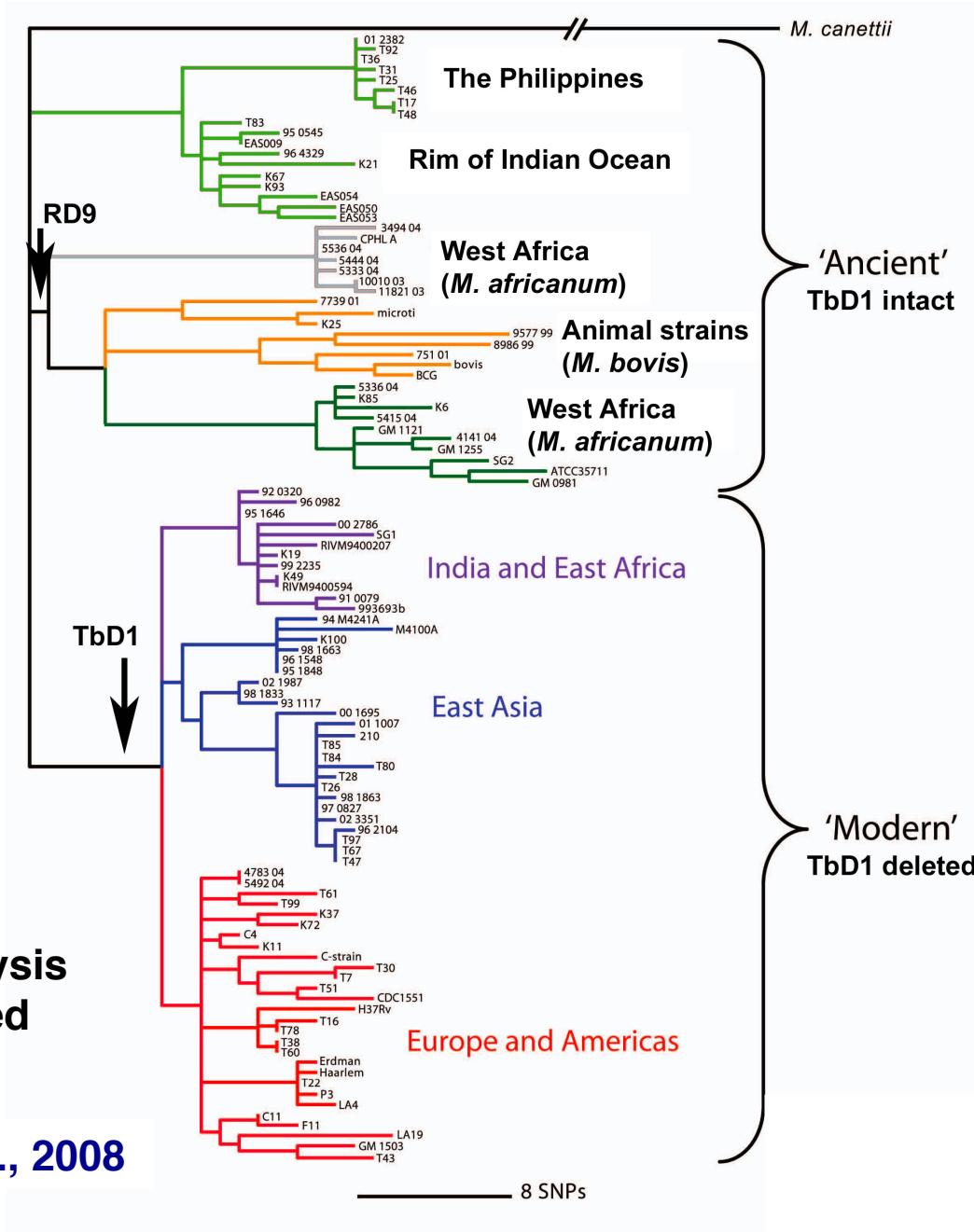
after Gagneux & Small., Lancet Infect Dis., 2007

# Phylogenetic scheme confirmed and refined by SNP based methods

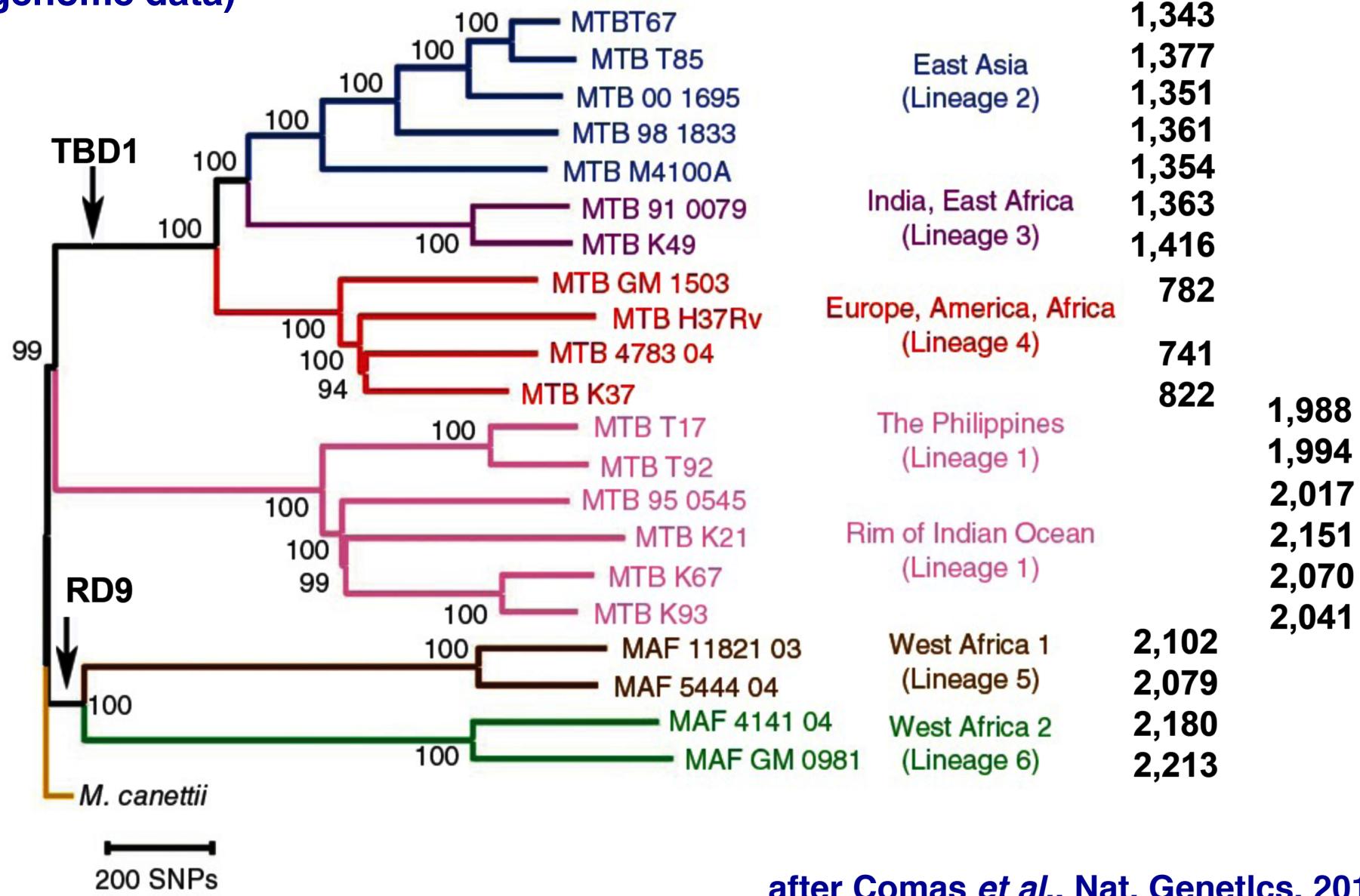
Baker et al., 2004,  
Gutacker et al., 2006,  
Filliol et al., 2006

and Multi-Locus-Sequence-Analysis  
(MLSA) based on 89 concatenated  
gene sequences

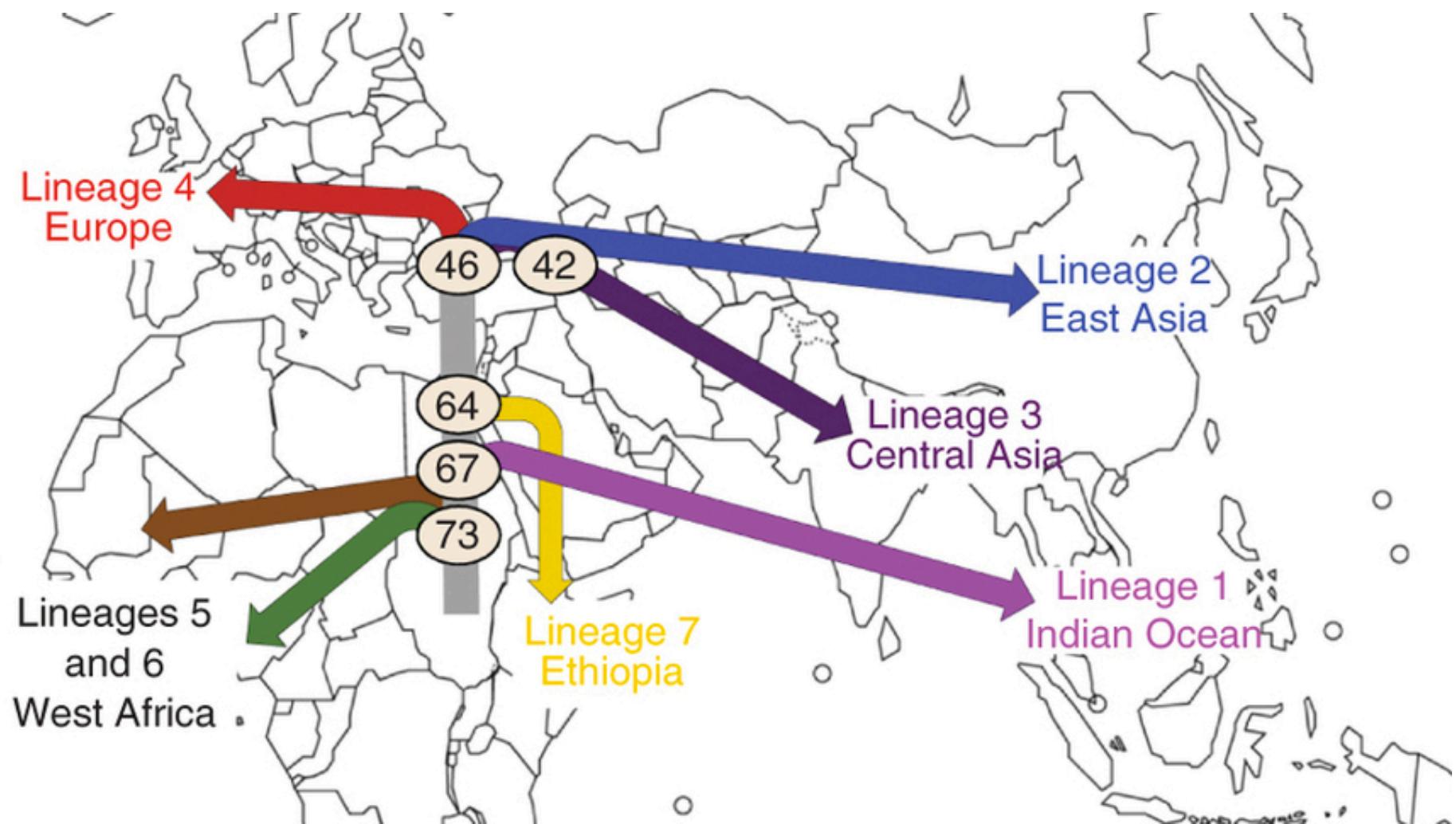
after Hershberg et al., PLoS Biol., 2008



## Broad diversity of worldwide *M. tuberculosis* strains (genome data)

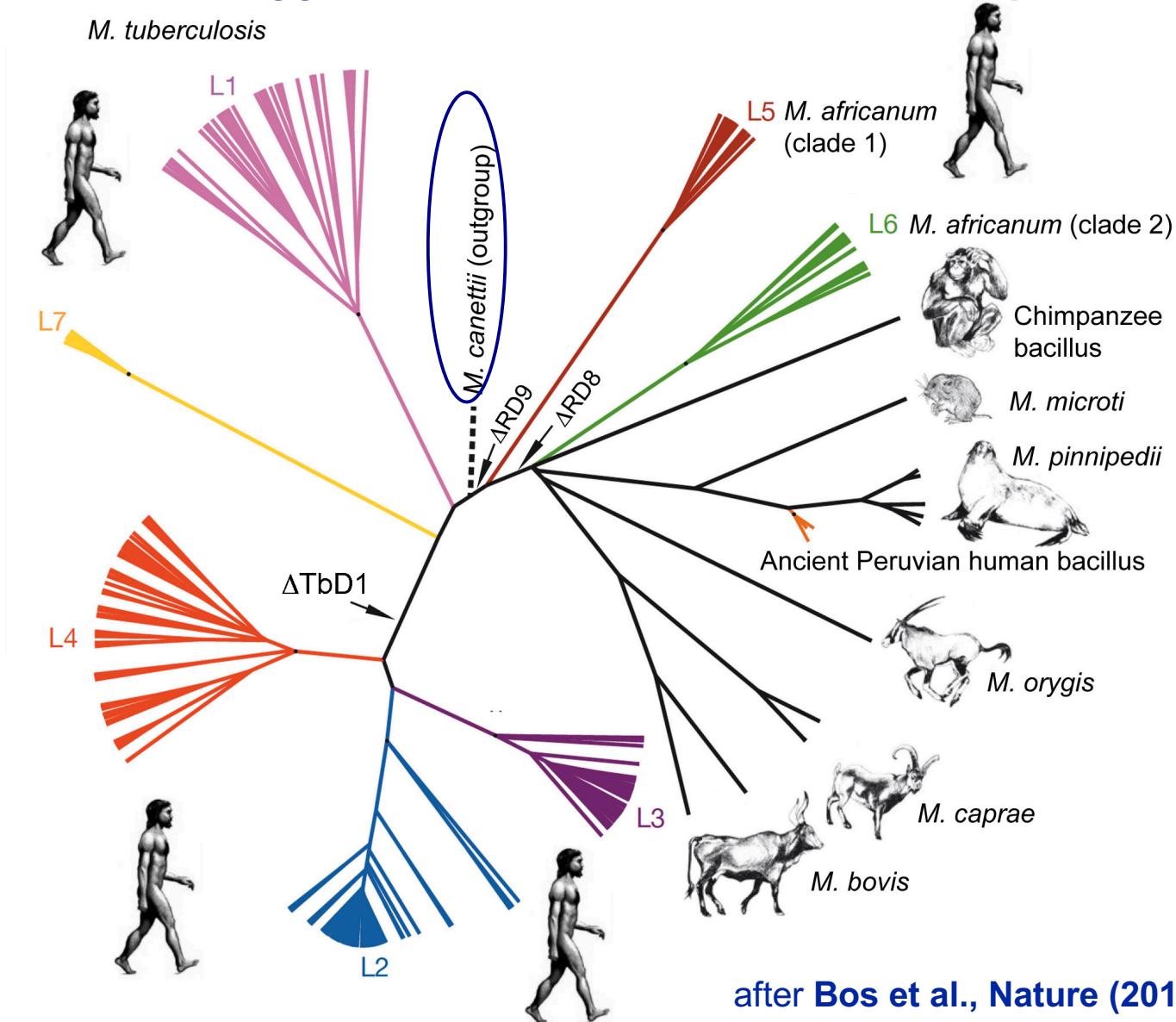


## Classical *M. tuberculosis* complex is estimated to have evolved ~ 70 000 y ago □



Comas et al., Nat. Genetics (2013)

# Analysis of ancient DNA from a *M. pinnipedii*-like *M. tuberculosis* complex member suggests that the *M. tuberculosis* complex is ~ 6 000 y old □



after Bos et al., Nature (2015)

Comas et al., Nat. Gen 2010,  
Comas et al., Nat. Gen 2013

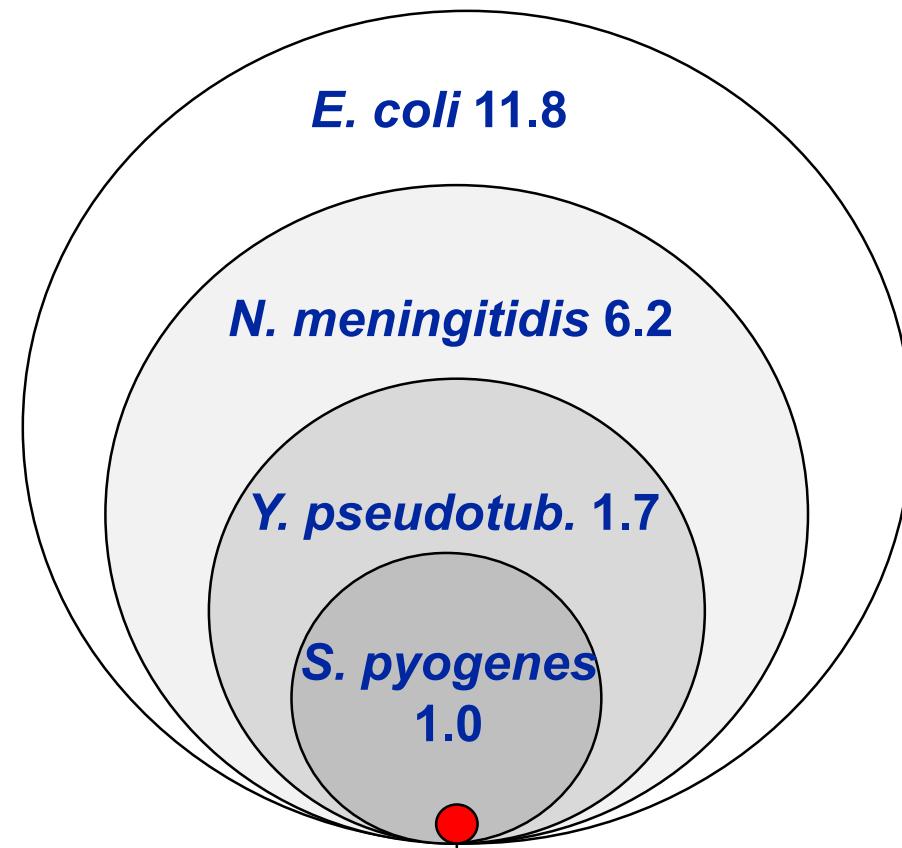
## ***M. tuberculosis* complex (MTBC) -> monomorphic bacterial population**

**However, are there also  
tubercule bacilli that are  
more diverse ?**

**ancestor of *M. tub.* is unknown**

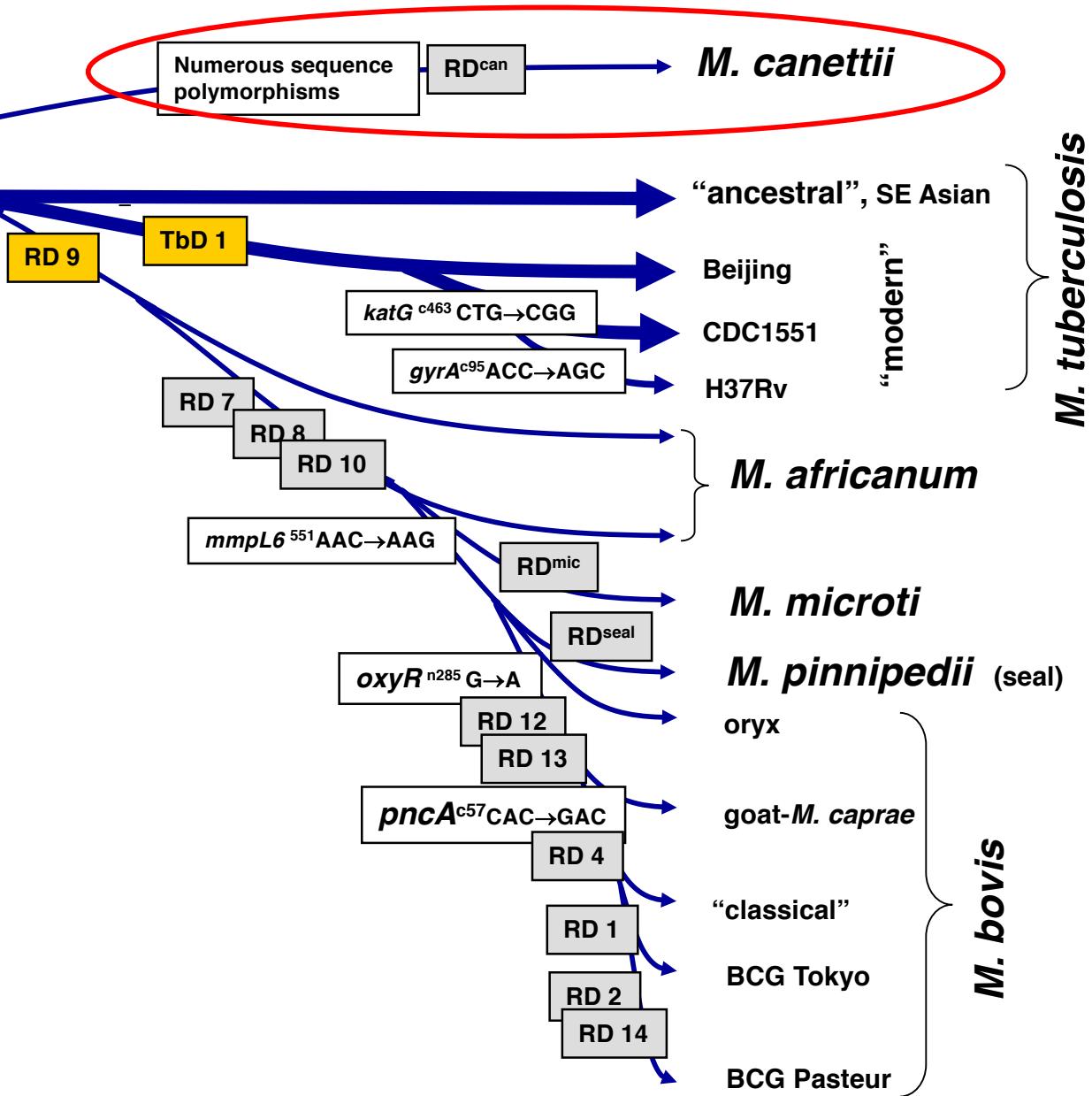
Sreevatasan et al., PNAS, 1997  
Fleischman et al., J Bact., 2002  
Comas et al., 2010

**Synonymous nt variation**

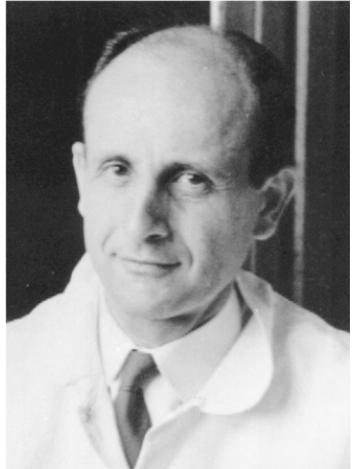


***M. tuberculosis* complex  
0.01-0.03%**

**Common ancestor of the tubercle bacilli**



# *Mycobacterium canetti* and smooth tubercle bacilli



Georges Canetti

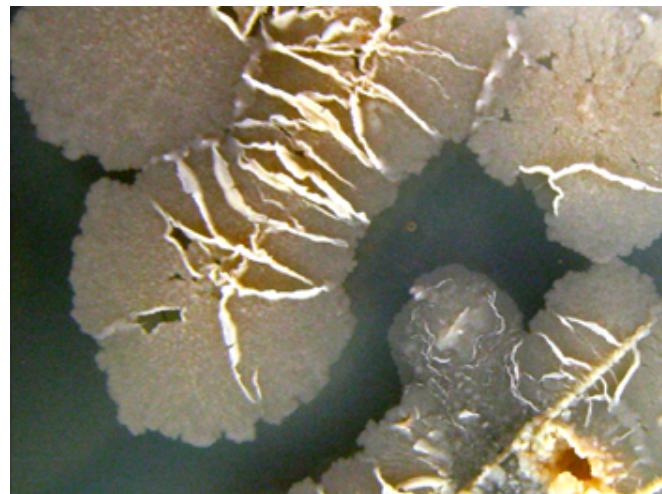
- less than 100 clinical isolates worldwide
  - mainly from TB patients from Djibouti, East Africa

INTERNATIONAL JOURNAL OF SYSTEMATIC BACTERIOLOGY, Oct. 1997, p. 1236–1245  
0020-7713/97/\$04.00+0  
Copyright © 1997, International Union of Microbiological Societies

Vol. 47, No. 4

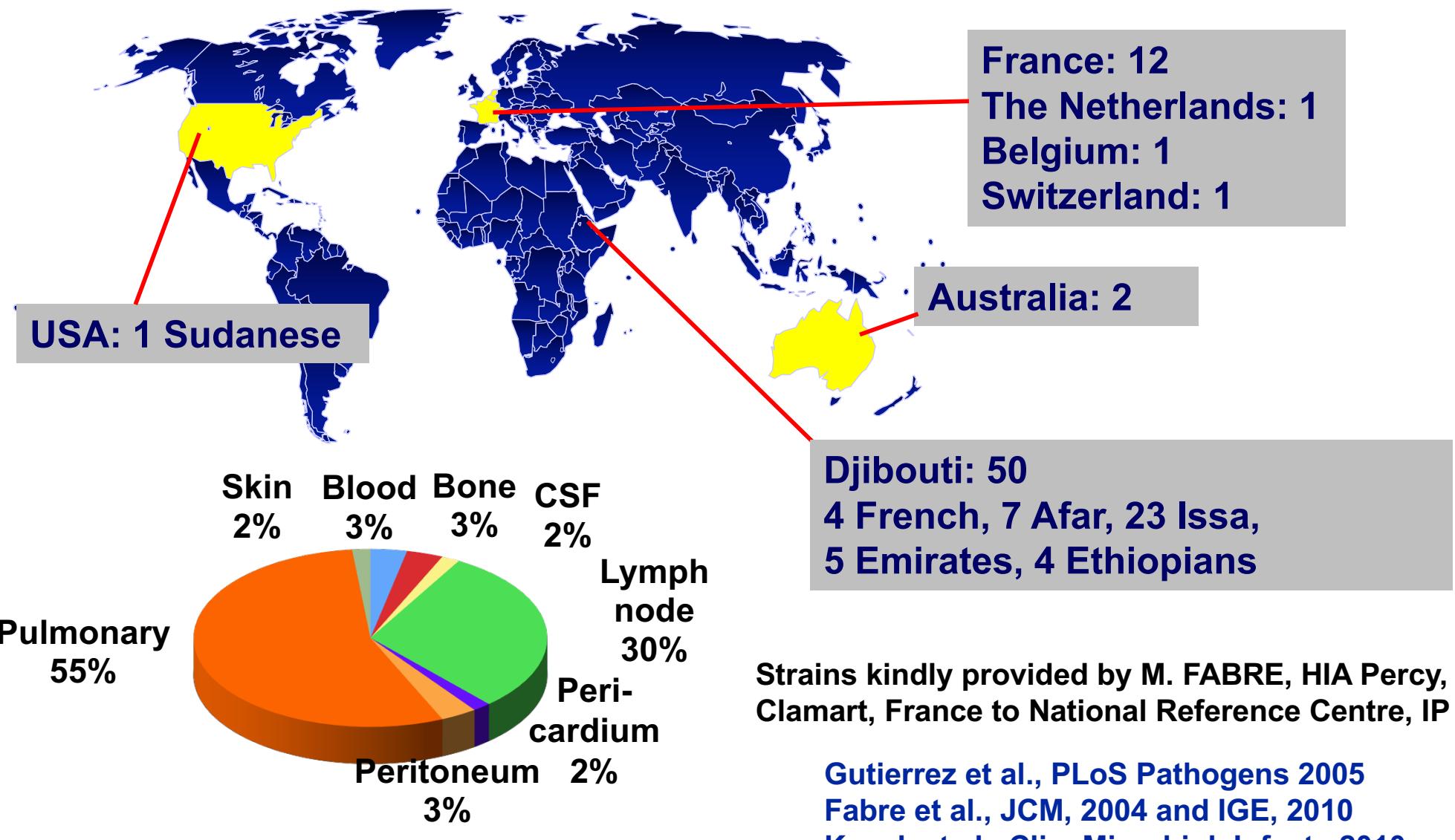
## A Novel Pathogenic Taxon of the *Mycobacterium tuberculosis* Complex, *Canetti*: Characterization of an Exceptional Isolate from Africa

DICK VAN SOOLINGEN,<sup>1,\*</sup> THEO HOOGENBOEZEM,<sup>2</sup> PETRA E. W. DE HAAS,<sup>1</sup> PETER W. M. HERMANS,<sup>2</sup> MARIANNE A. KOEDAM,<sup>3</sup> KOOS S. TEPPEMA,<sup>3</sup> PATRICK J. BRENNAN,<sup>4</sup> GURDYAL S. BESRA,<sup>4</sup> FRANCOISE PORTAEELS,<sup>5</sup> JANETTA TOP,<sup>6</sup> LEO M. SCHOUMLS,<sup>6</sup> AND JAN D. A. VAN EMBDEN<sup>6</sup>

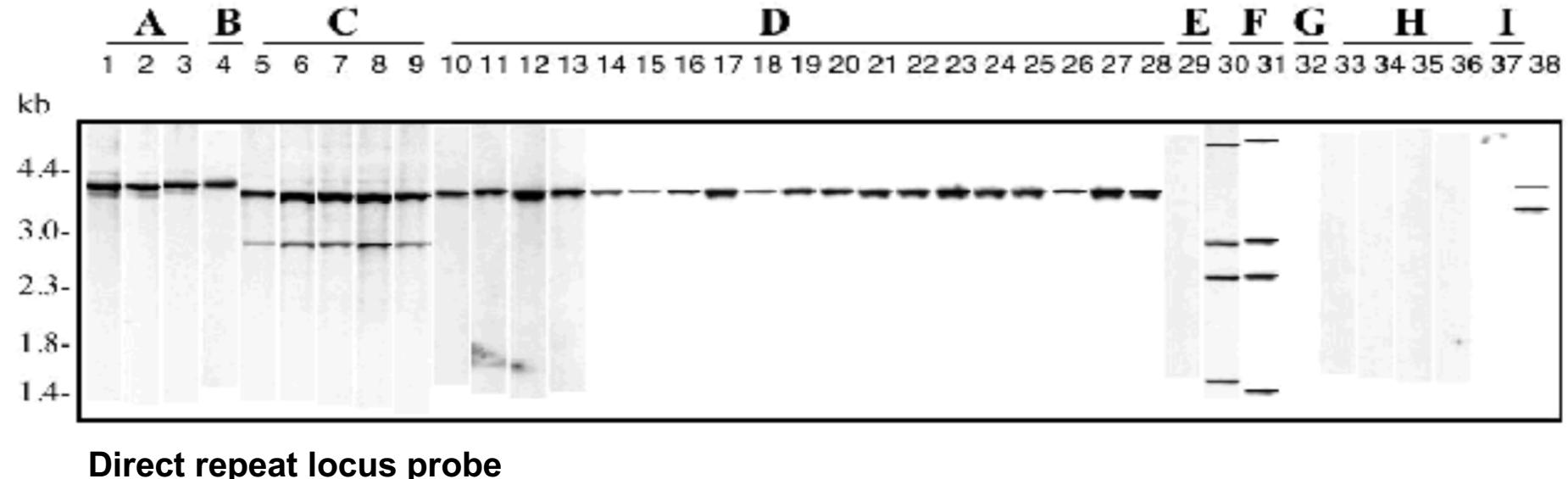


**Pfyffer et al., EID; Fabre et al., 2004, Gutierrez et al, PLoS Pathog; Koeck et al., CMI, 2010**

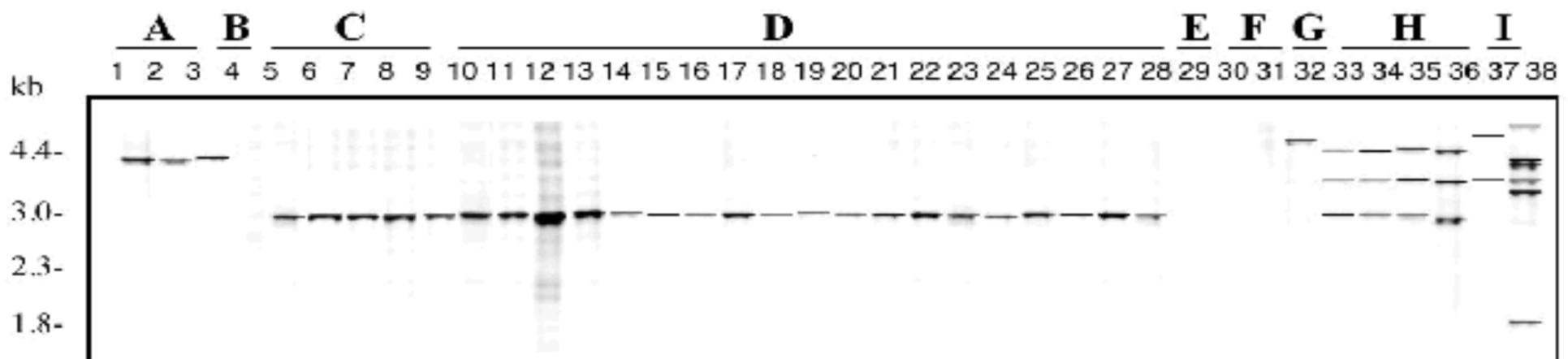
# Origin of clinical isolates of smooth tubercle bacilli



## Molecular typing of smooth tubercle bacilli defined lineages (IS6110, IS1081, RD12<sup>Can</sup>, ISMyca1)



Direct repeat locus probe

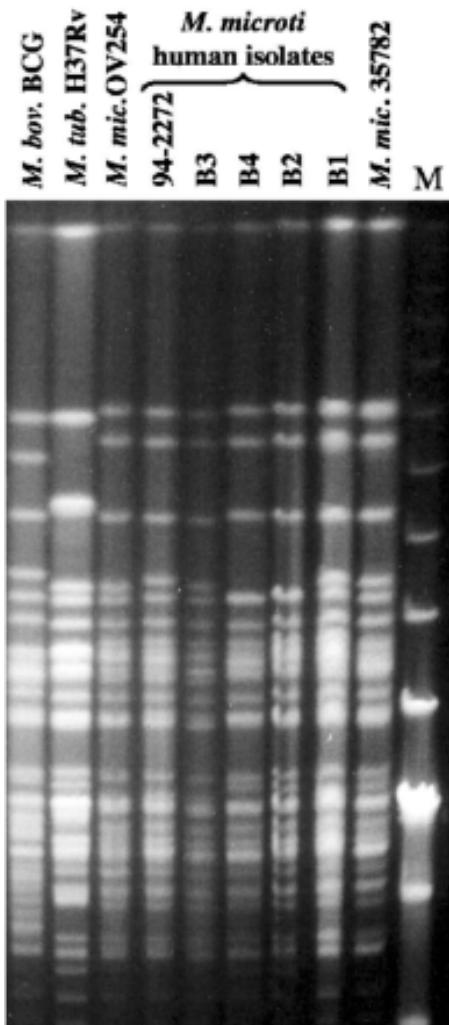


IS1081 probe

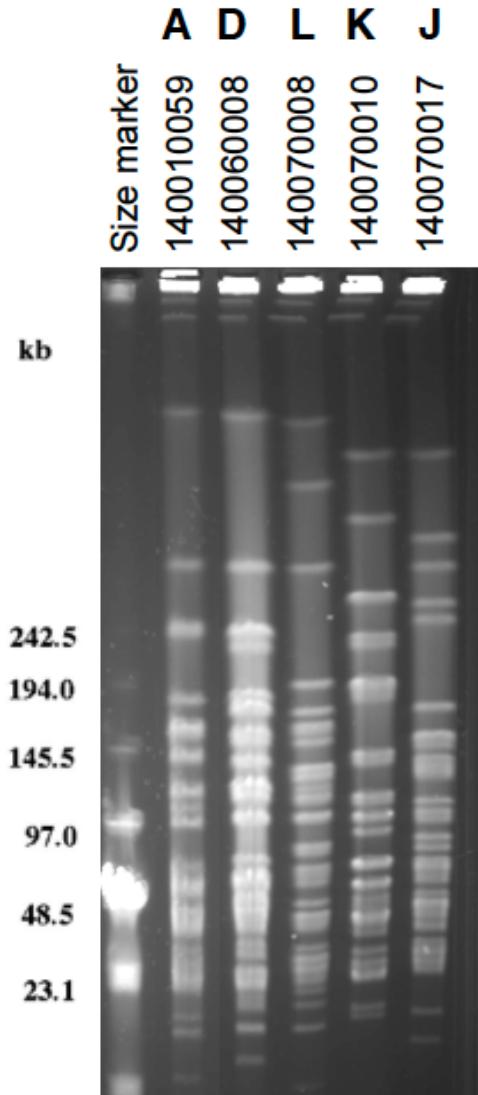
Gutierrez et al., PLoS Pathog, 2005;

# Genomic diversity of strains selected for sequencing

Classical *Mtb* complex



*M. proto-tuberculosis*



D 140060008= pulmonary TB  
L 140070008= lymph node TB  
K 140070010= cutaneous TB  
J 140070017= lymph node TB

→ collaboration with team of  
C. Medigue, Génoscope Evry  
and P. Supply (Pasteur-Lille)

A 140010059= pulmonary TB

*M. canettii* reference strain  
→ collaboration with team of  
J. Parkhill, Sanger Institute

**Genome sizes of smooth strains are larger than in *Mtb***

	<i>M. canettii</i>	STB-A	4,482 kb
	STB-D		4,448 kb
	STB-L		4,437 kb
	STB-K		4,525 kb
	STB-J		4,510 kb
<hr/>			
MTBC	<i>M. tub.</i> H37Rv		4,411 kb
	<i>M. africanum</i>		4,389 kb
	<i>M. bovis</i>		4,345 kb
<hr/>			
Non-tub. mycob.	<i>M. marinum</i>		6,637 kb
	<i>M. kansasii</i>		6.400 kb

- *Smooth TB* genomes are ~ 15 to 100 kbp larger than MTBC genomes
- ~ 2 Mb smaller than non-tuberculous *M. marinum* & *M. kansasii*
- consistent with ancestral status, and still close relationship with *M. tub.*

Supply *et al.*, *Nat. Genetics*, 2013

**Quantitative SNP analysis shows up to 25 times more SNPs than within classical *M. tuberculosis* complex**

<i>M. tuberculosis</i>	H37Rv	STB-D	STB-A	STB-L	STB-J
STB-D		16,168			
STB-A	19,673		9,525		
STB-L	23,629	22,049		22,657	
STB-J	45,897	46,077	44,767	47,145	
STB-K	61,228	57,501	56,723	55,829	65,744
<i>M. bovis</i>	2,348 SNPs				

Supply *et al.*, *Nat. Genetics*, 2013

# Characteristics of *M. canettii* strains - summary based on whole genome sequencing project of 9 strains

nature  
genetics

OPEN

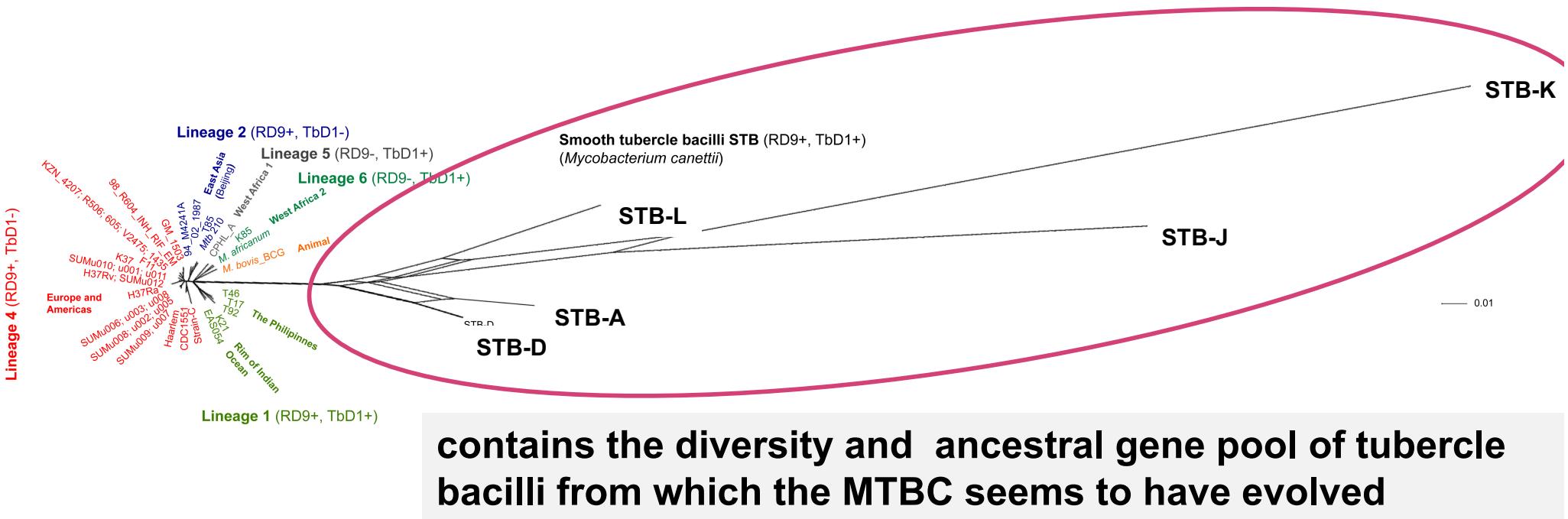
## Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*

Philip Supply<sup>1-4</sup>, Michael Marceau<sup>1-4</sup>, Sophie Mangenot<sup>5</sup>, David Roche<sup>5,6</sup>, Carine Rouanet<sup>1-4</sup>, Varun Khanna<sup>7</sup>, Laleh Majlessi<sup>8,9</sup>, Alexis Criscuolo<sup>10</sup>, Julien Tap<sup>10</sup>, Alexandre Pawlik<sup>7</sup>, Laurence Fiette<sup>11,12</sup>, Mickael Orgeur<sup>7</sup>, Michel Fabre<sup>13</sup>, Cécile Parmentier<sup>7</sup>, Wafa Frigui<sup>7</sup>, Roxane Simeone<sup>7</sup>, Eva C Boritsch<sup>7</sup>, Anne-Sophie Debrie<sup>1-4</sup>, Eve Willery<sup>1-4</sup>, Danielle Walker<sup>14</sup>, Michael A Quail<sup>14</sup>, Laurence Ma<sup>15</sup>, Christiane Bouchier<sup>15</sup>, Grégory Salvignol<sup>5,6</sup>, Fadel Sayes<sup>8,9</sup>, Alessandro Cascioferro<sup>7</sup>, Torsten Seemann<sup>16</sup>, Valérie Barbe<sup>5</sup>, Camille Locht<sup>1-4</sup>, Maria-Cristina Gutierrez<sup>1-4,17</sup>, Claude Leclerc<sup>8,9</sup>, Stephen D Bentley<sup>14</sup>, Timothy P Stinear<sup>18</sup>, Sylvain Brisse<sup>10</sup>, Claudine Médigue<sup>5,6</sup>, Julian Parkhill<sup>14</sup>, Stéphane Cruveiller<sup>5,6</sup> & Roland Brosch<sup>7</sup>

- up to 100 kb larger genomes
- lower virulence in mice
- up to 65 000 SNPs
- mosaic genome structures

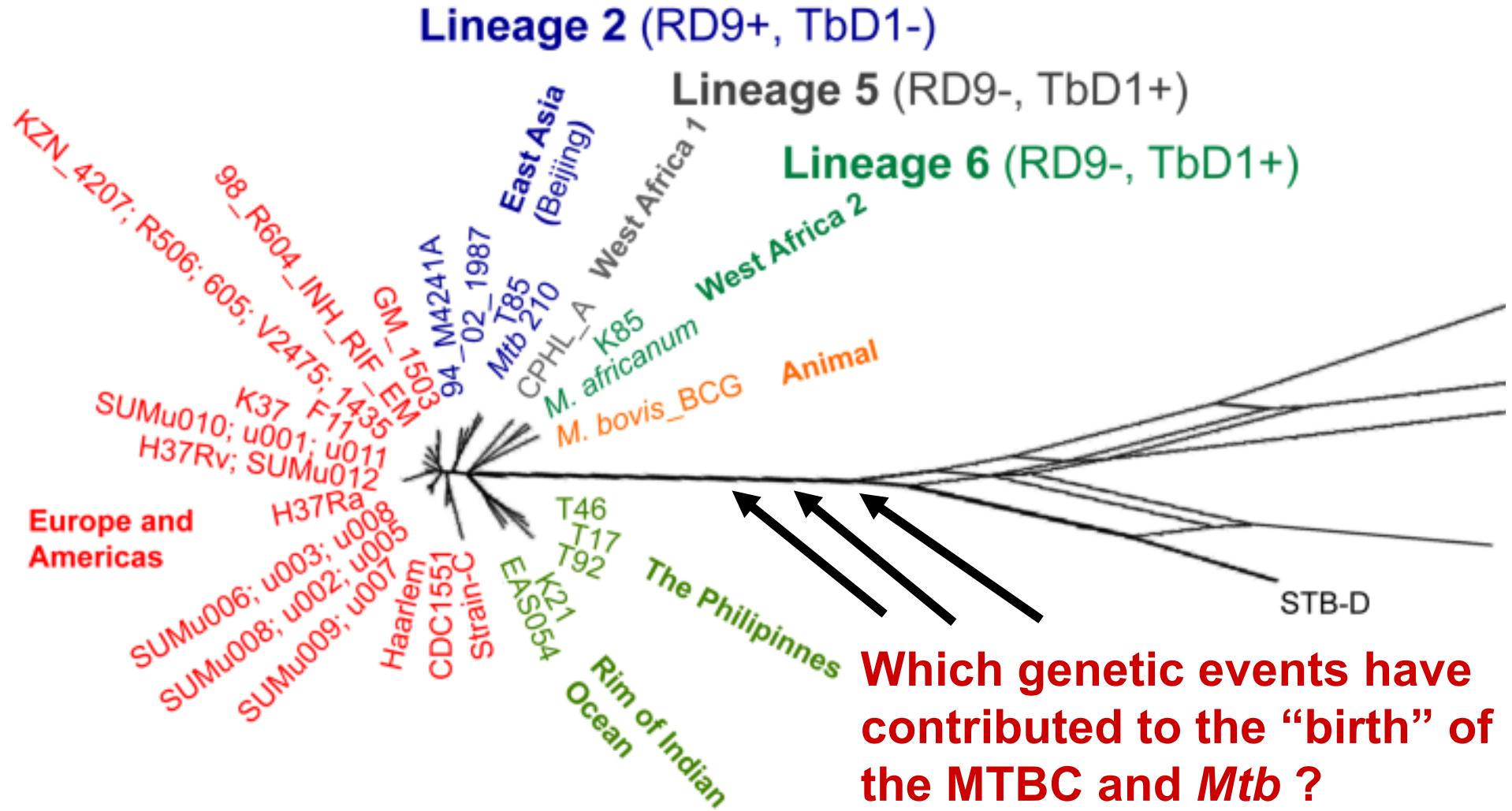
Supply et al., 2013, Nat Genet

**NeighborNet alignments of whole-genome SNP data display the much larger genetic diversity of smooth strains**



Network phylogeny inferred among the five smooth tubercle bacilli isolates subjected to complete genome sequence analysis and 39 selected genome sequences from members of the classical *M. tuberculosis* complex (MTBC) by NeighborNet analysis, based on pairwise alignments of whole-genome SNP data.

## Lineage 4 (RD9+, TbD1-)

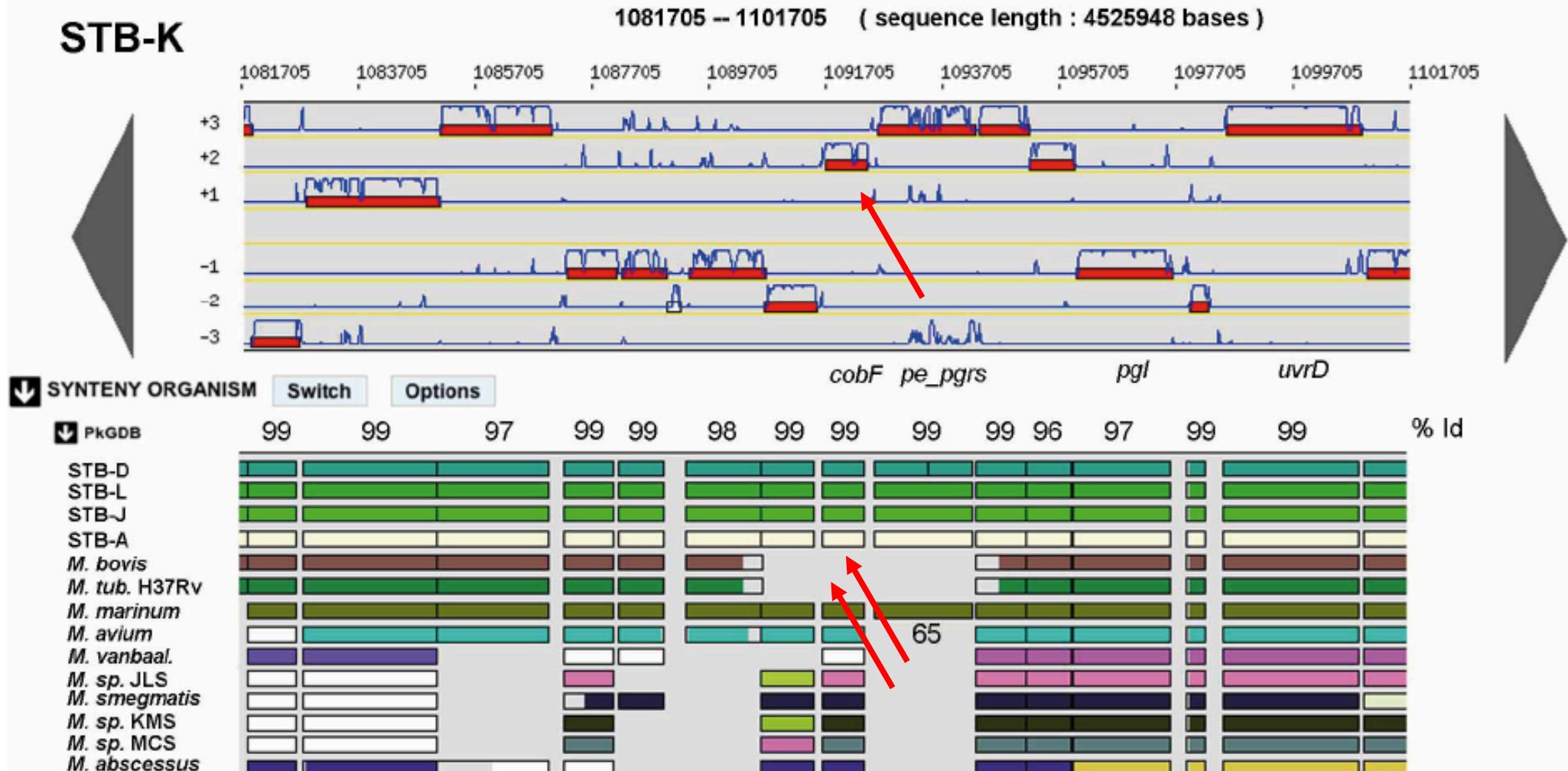


**Which genetic events have contributed to the “birth” of the MTBC and *Mtb* ?**

## Lineage 1 (RD9+, TbD1+)

# **cobF is absent from all members of the MTBC**

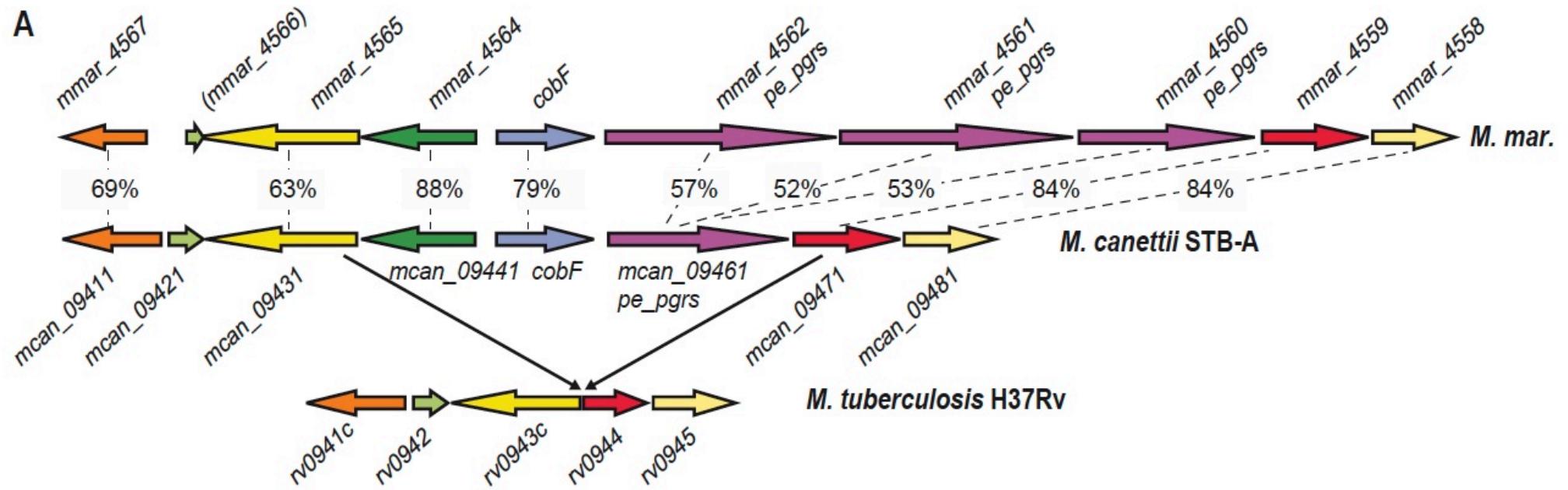
**STB-K**



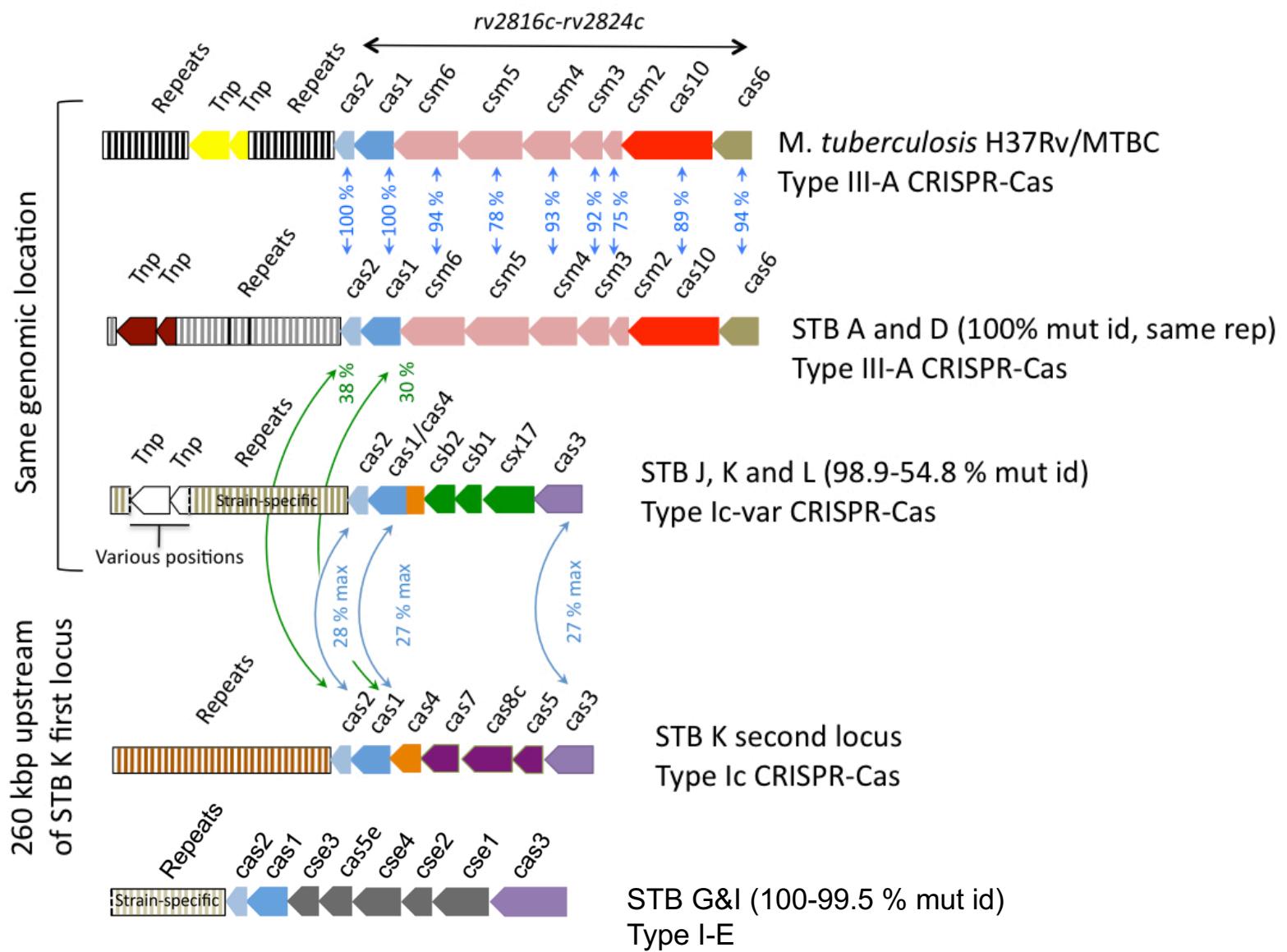
precorrin-6A synthase

MaGe <https://www.genoscope.cns.fr/agc/microscope/mage/viewer.php>

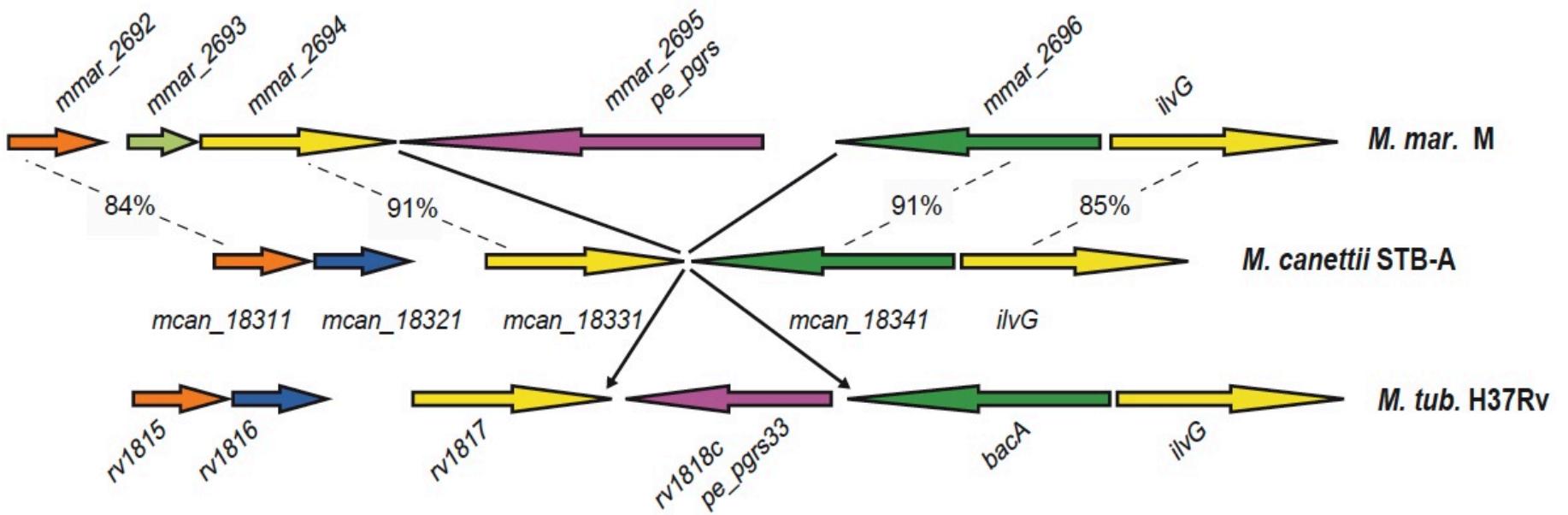
# *cobF* is absent from all members of the MTBC



# *M. canettii* strains harbour distinct CRISPR regions

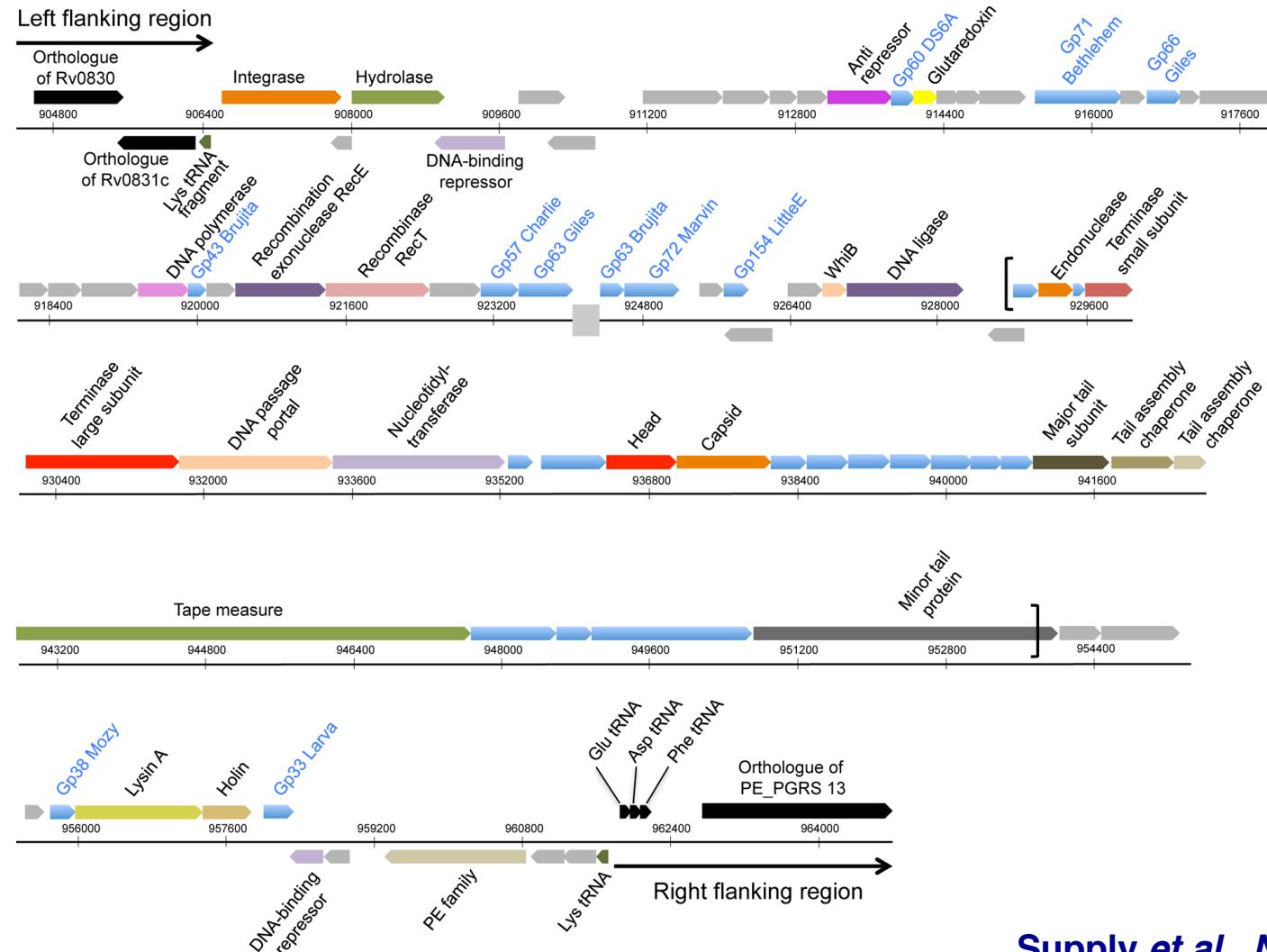


# Some PE-PGRS genes are unique for MTBC:



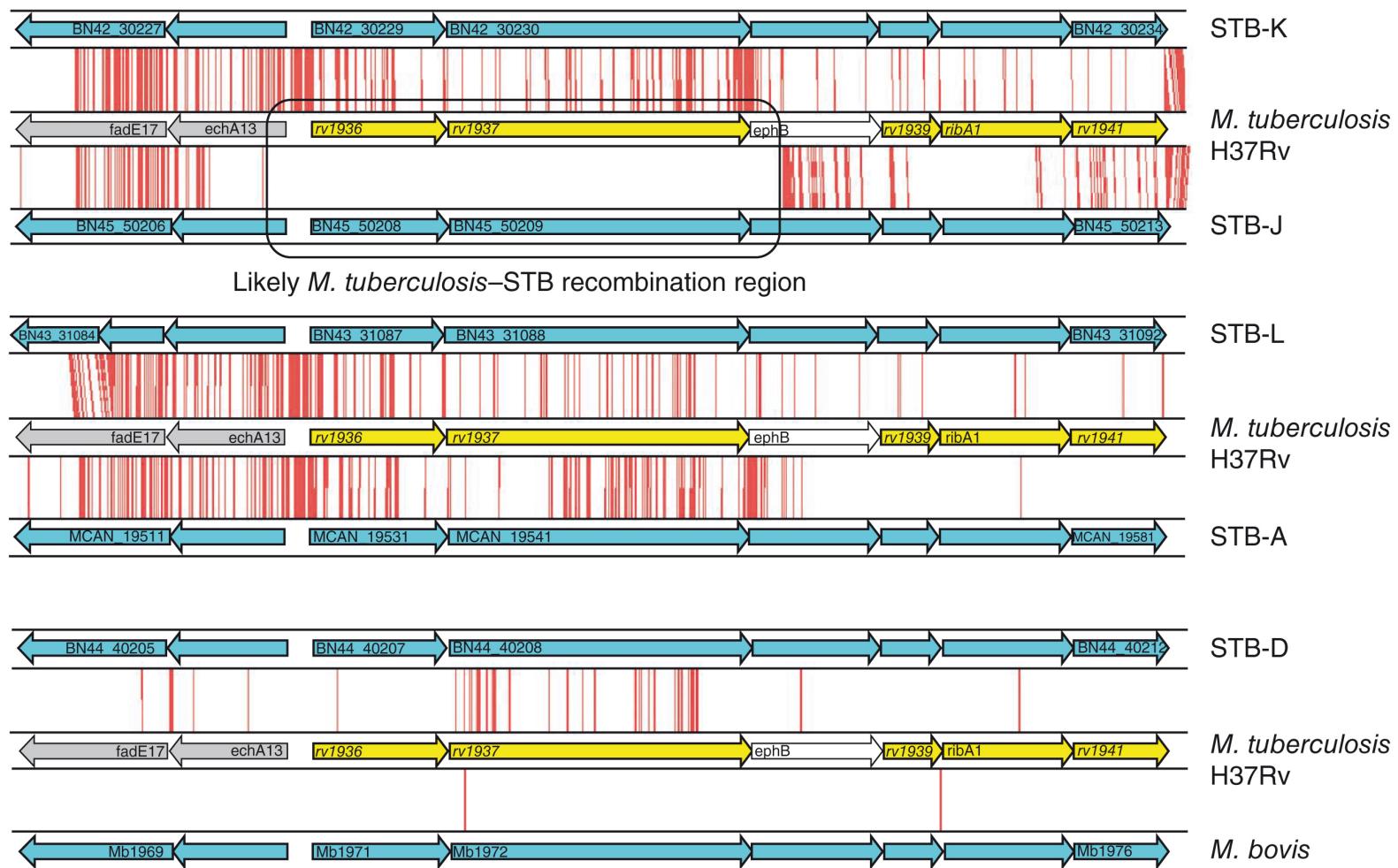
Boritsch *et al.*, Mol Microbiol, 2014

# Identification of a putative 55 kb phage inserted into the Lys tRNA gene of *M. canettii* strain I



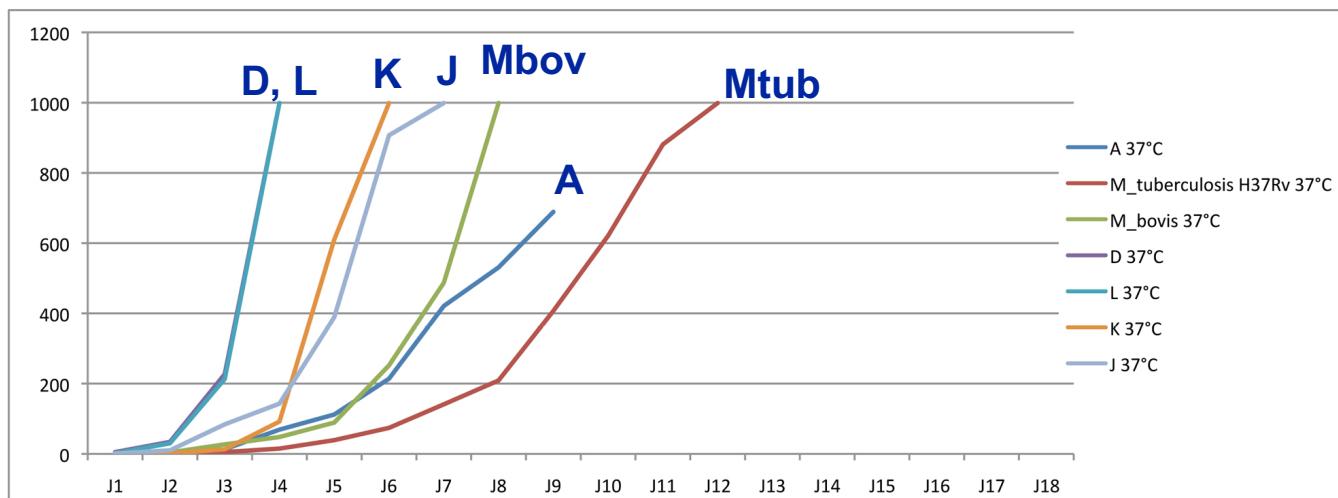
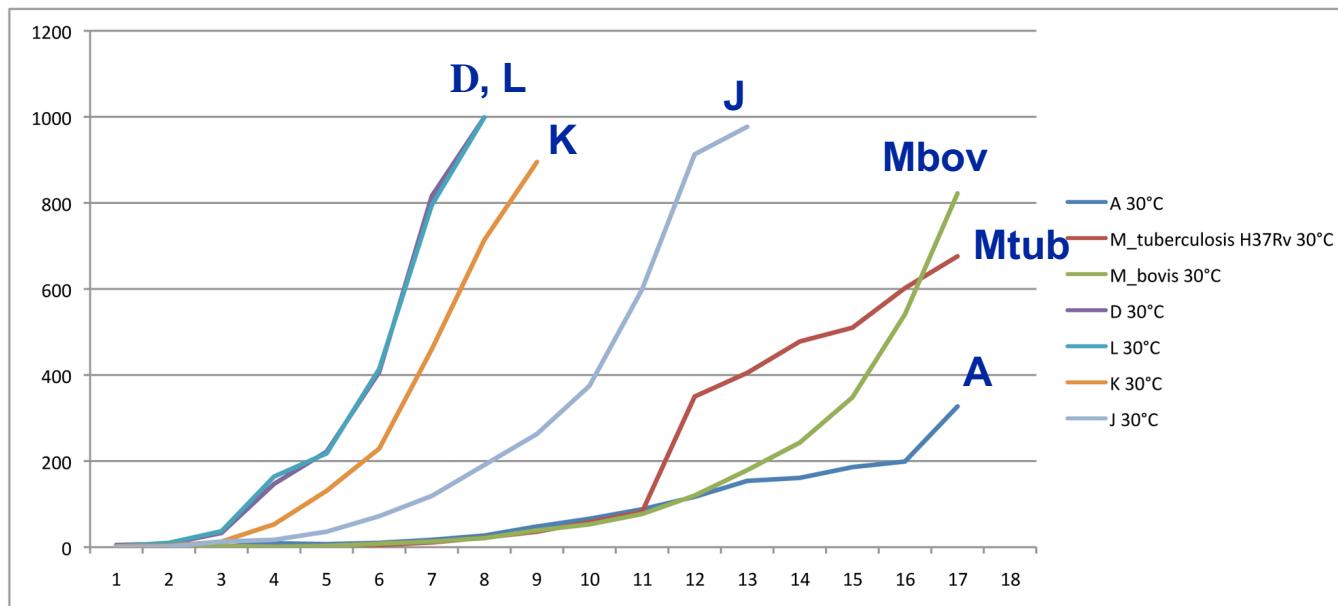
Supply et al., Nat. Genetics, 2013

# Extensive recombination among *M. canettii* as well as Mcan-MTBC

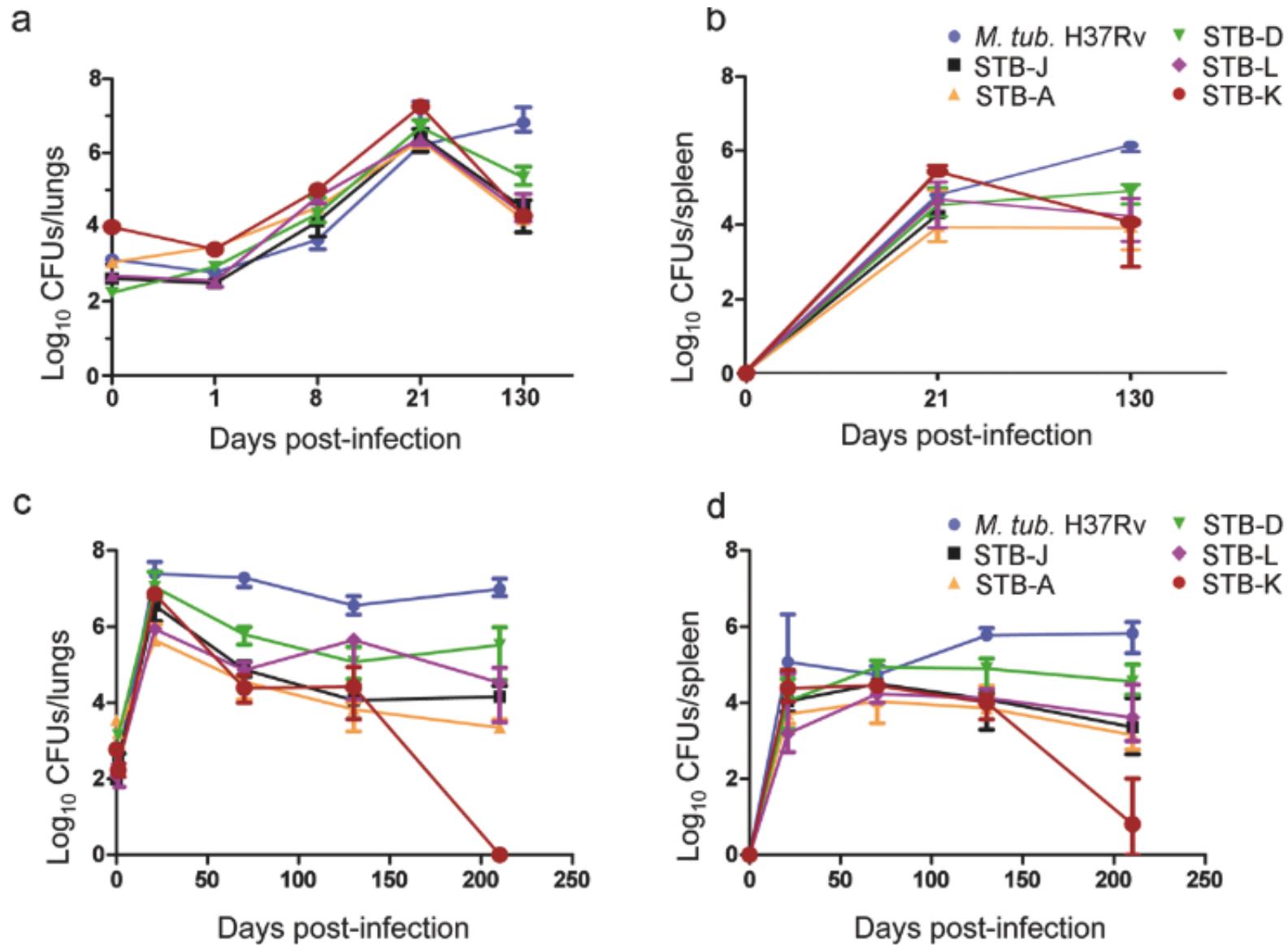


Supply *et al.*, *Nat. Genetics*, 2013;  
Boritsch *et al.*, *Mol Microbiol*, 2014

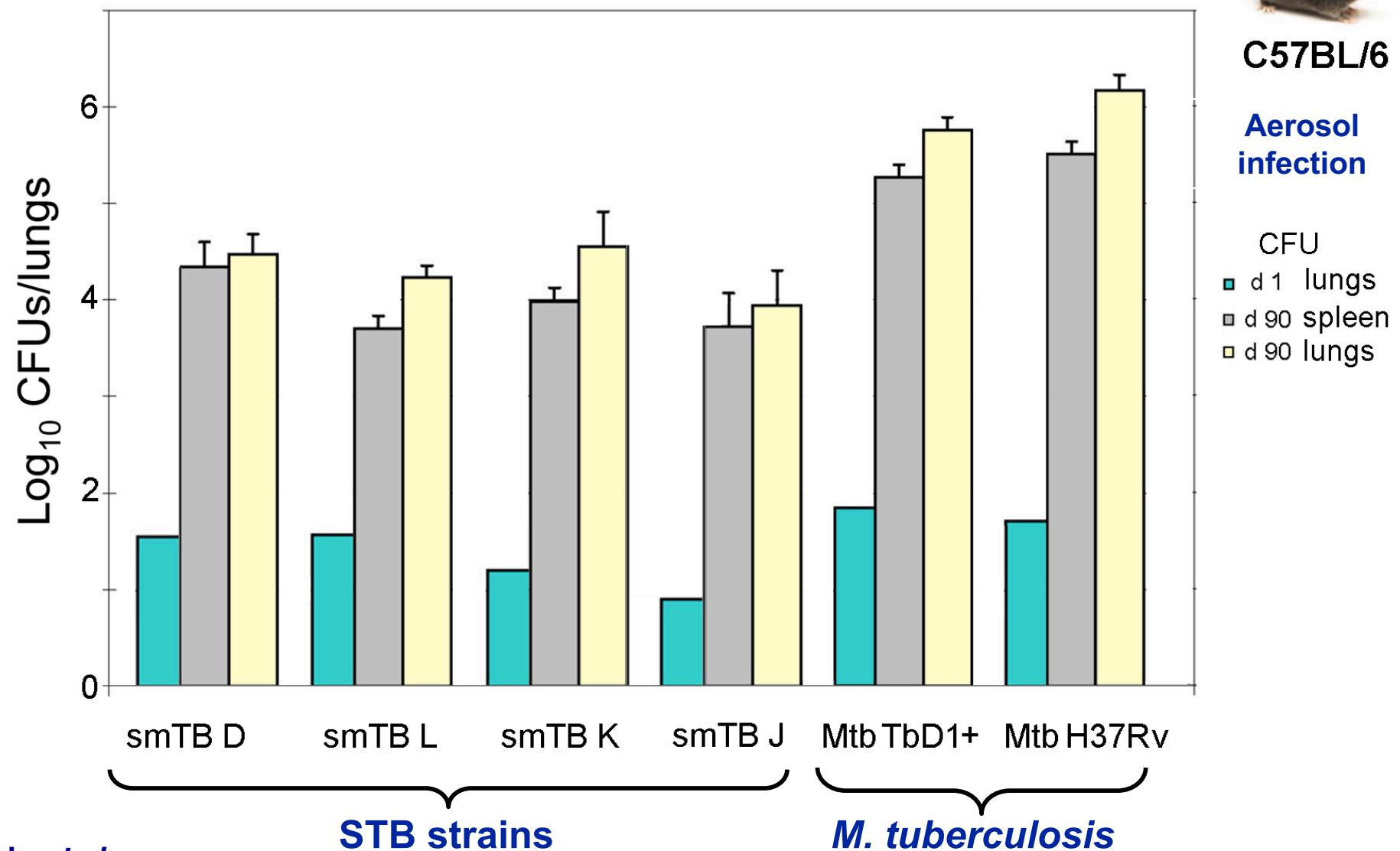
# Smooth TB bacilli grow faster at 30° C and 37° C in BACTEC



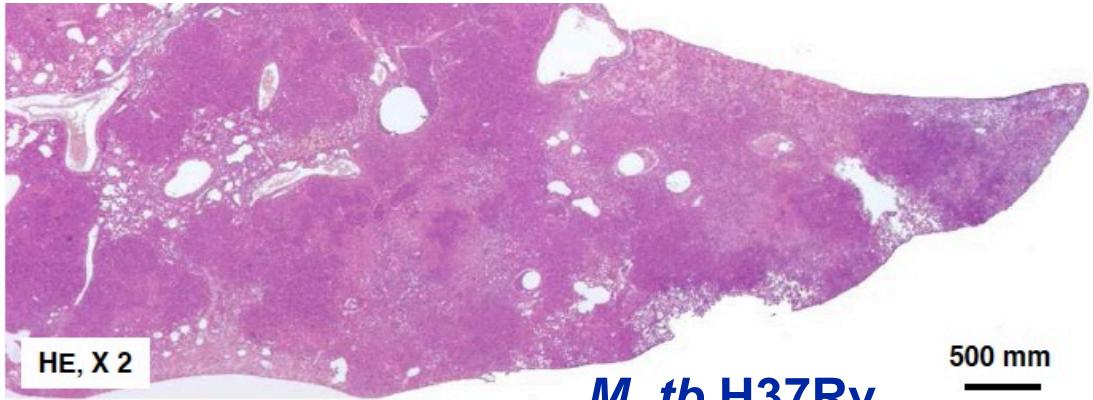
## Smooth tubercle bacilli are less virulent & less persistent in Balb/C mice



## *Smooth tubercle bacilli* are also less virulent in C57BL/6 mice



Supply et al.,  
*Nat. Genetics*, 2013



**Histology confirms :**  
STB strains are less virulent  
in the C57BL/6 mouse model

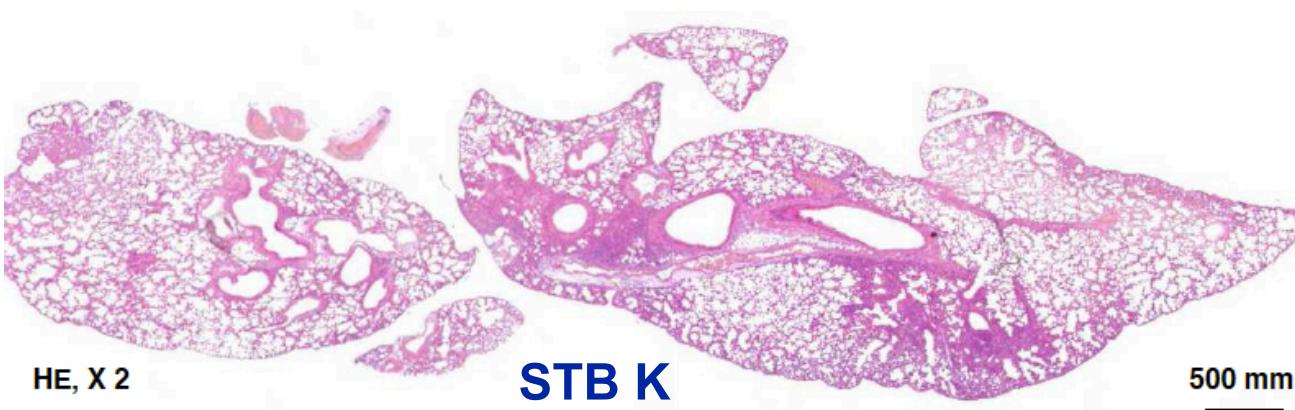
*M. tb* H37Rv



HE, X 2

STB D

500 mm



HE, X 2

STB K

500 mm

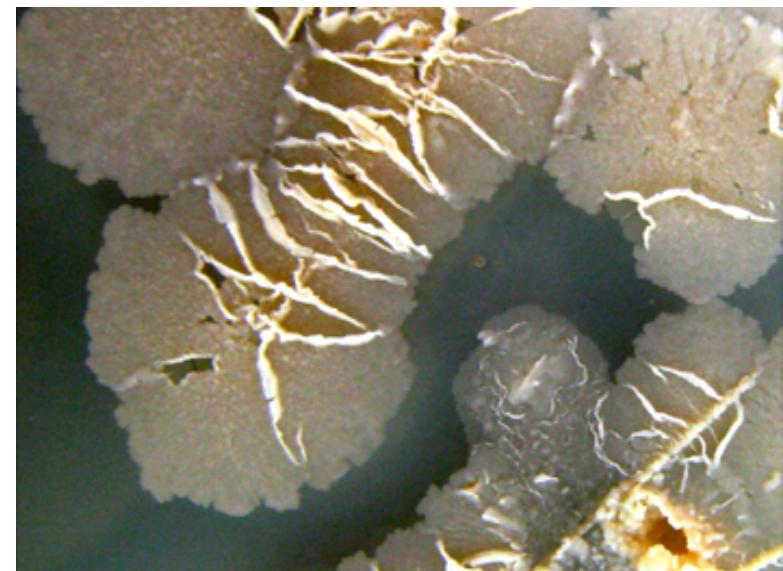
Supply et al.,  
*Nat. Genetics*, 2013

Why are *M. canettii* strains show a smooth morphotype ?

Has this anything to do with virulence & pathogenicity

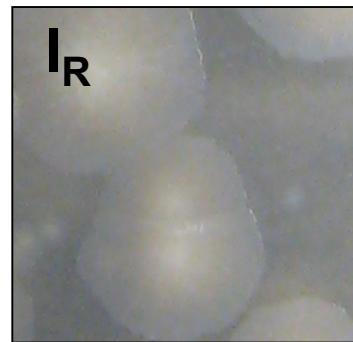
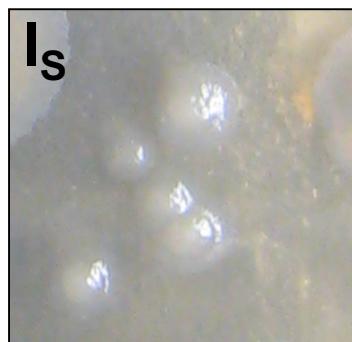
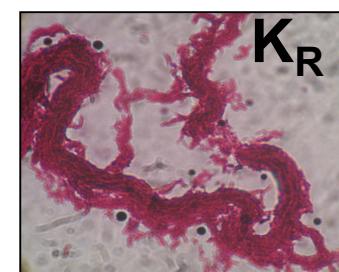
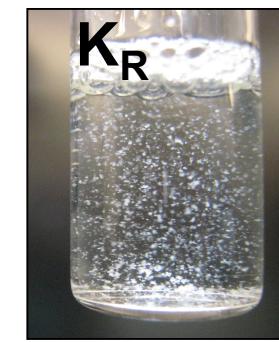
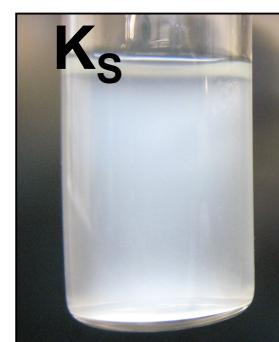
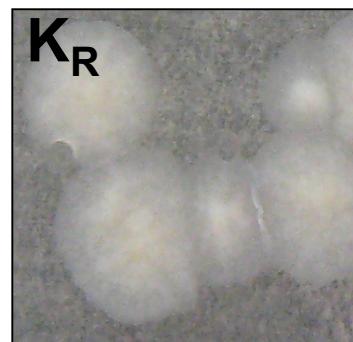
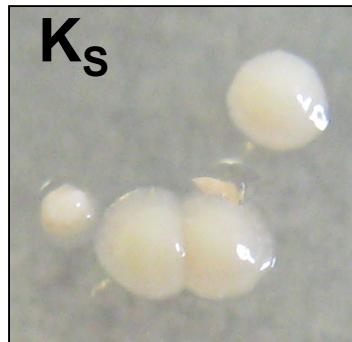


*M. canettii*



*M. tuberculosis*

## Isolation of spontaneous rough mutants of *M. canettii* strains K & I



Spontaneous rough *M. canettii* variants showed consistent fixed morphotype → genetic change ?

Genome sequencing of **smooth** and **spontaneous rough**  
***M. canettii*** variants

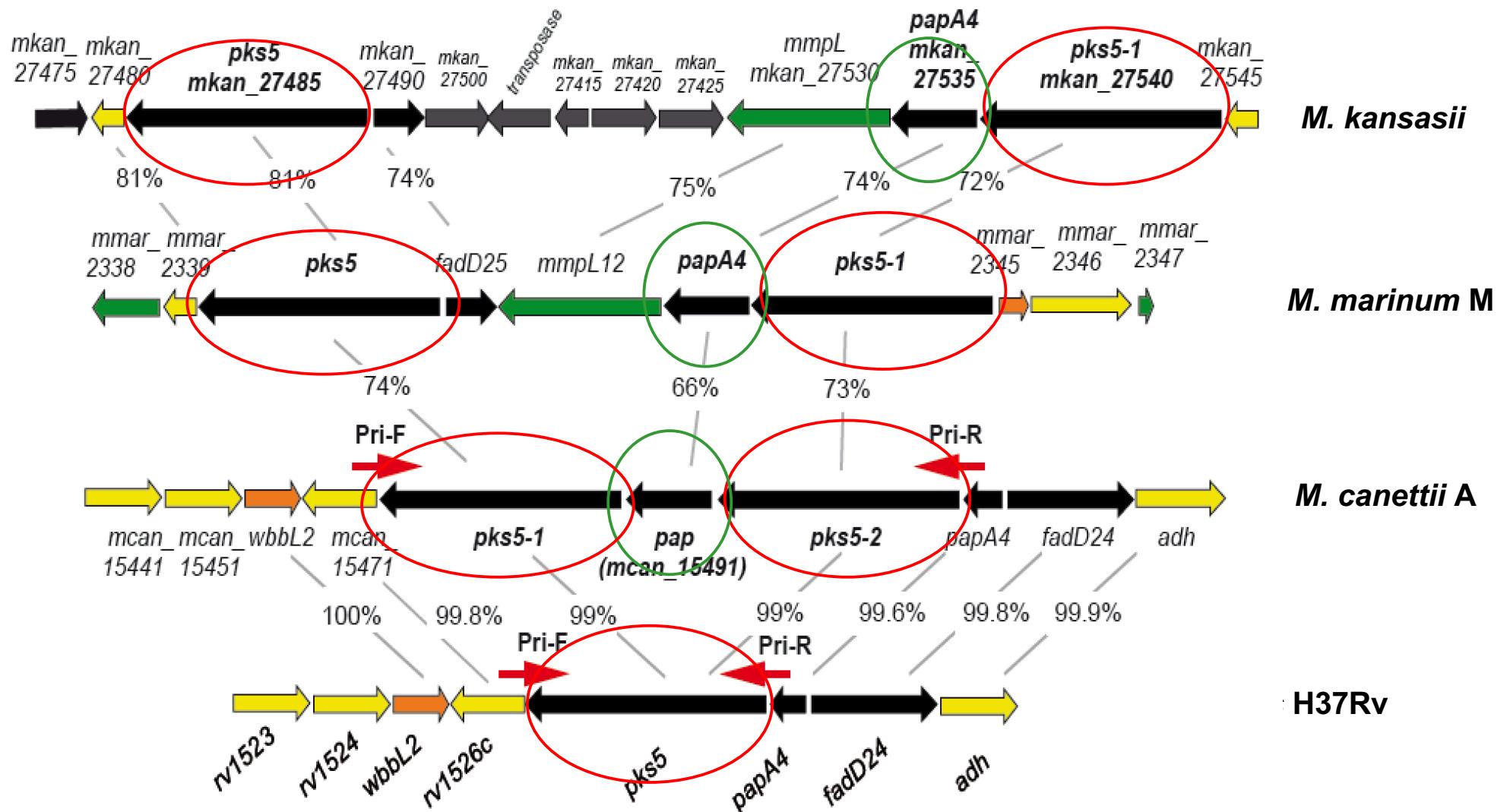
→ identified SNPs and read coverage depth differences in ***pks5***

***pks5*, encodes for a type I polyketide synthase**

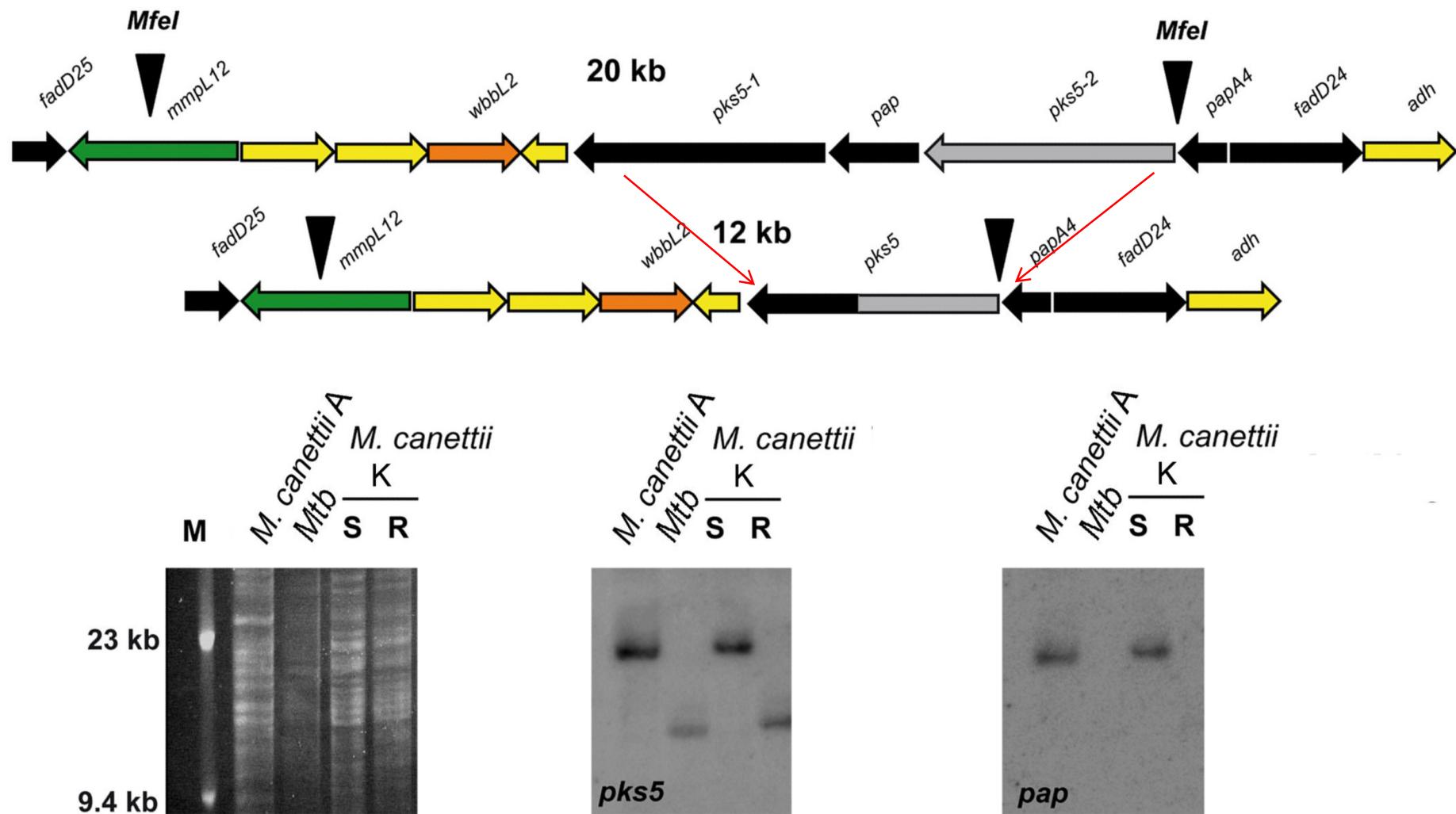
***In M. marinum and M. kansasii Pks5 & Pks5.1 polyketide synthases***  
→ synthesis of polymethyl-branched fatty acids that are modified by glycosyltransferases, methyltransferases to produce  
**Lipooligosaccharide (LOS)**

Rombouts, Y. et al , JBC, 2011; Ren et al., Mol Microbiol 2007; Minnikin, et al. InTech, 2015

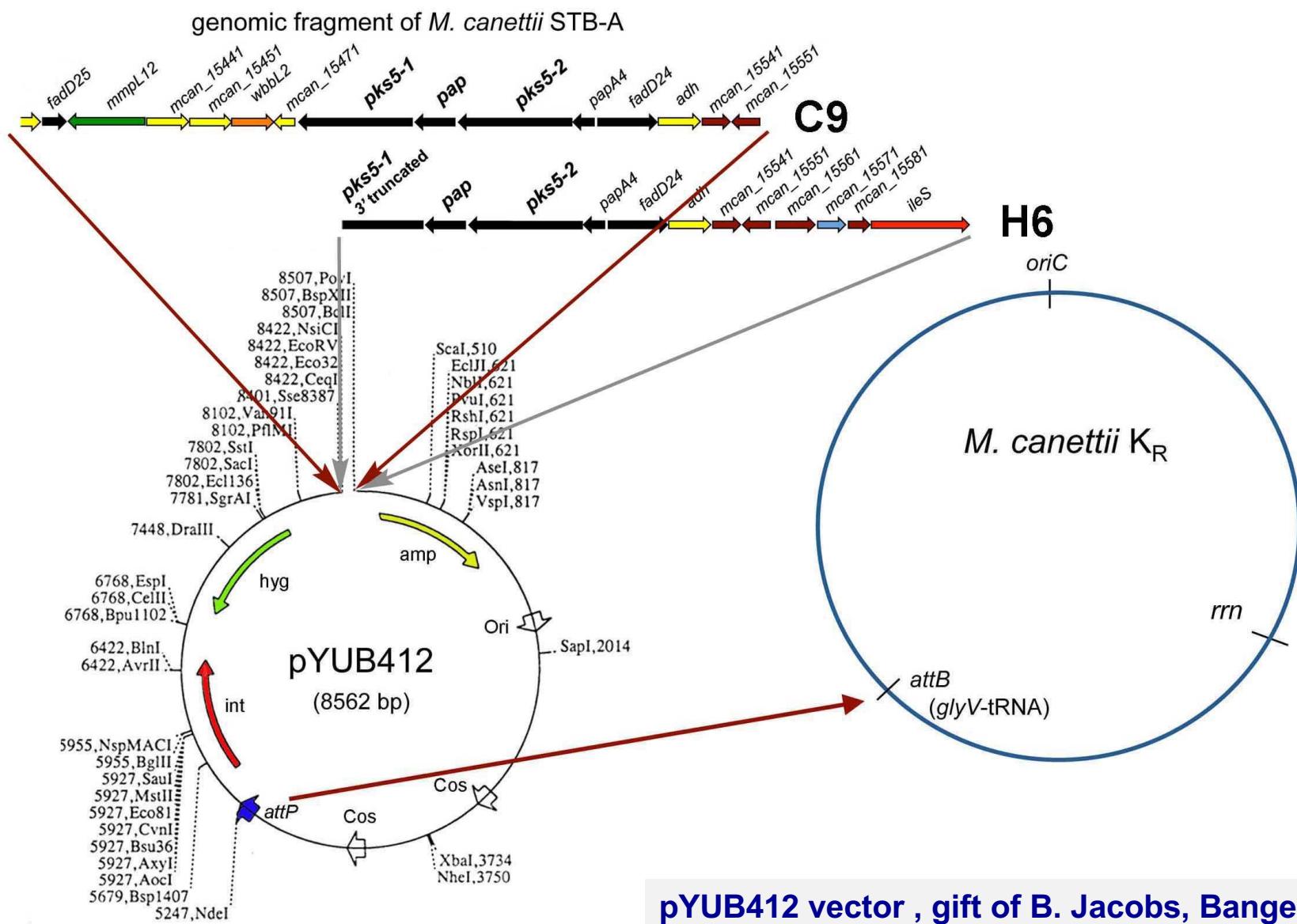
# Genetic organisation of the *pks5* locus in different mycobacteria



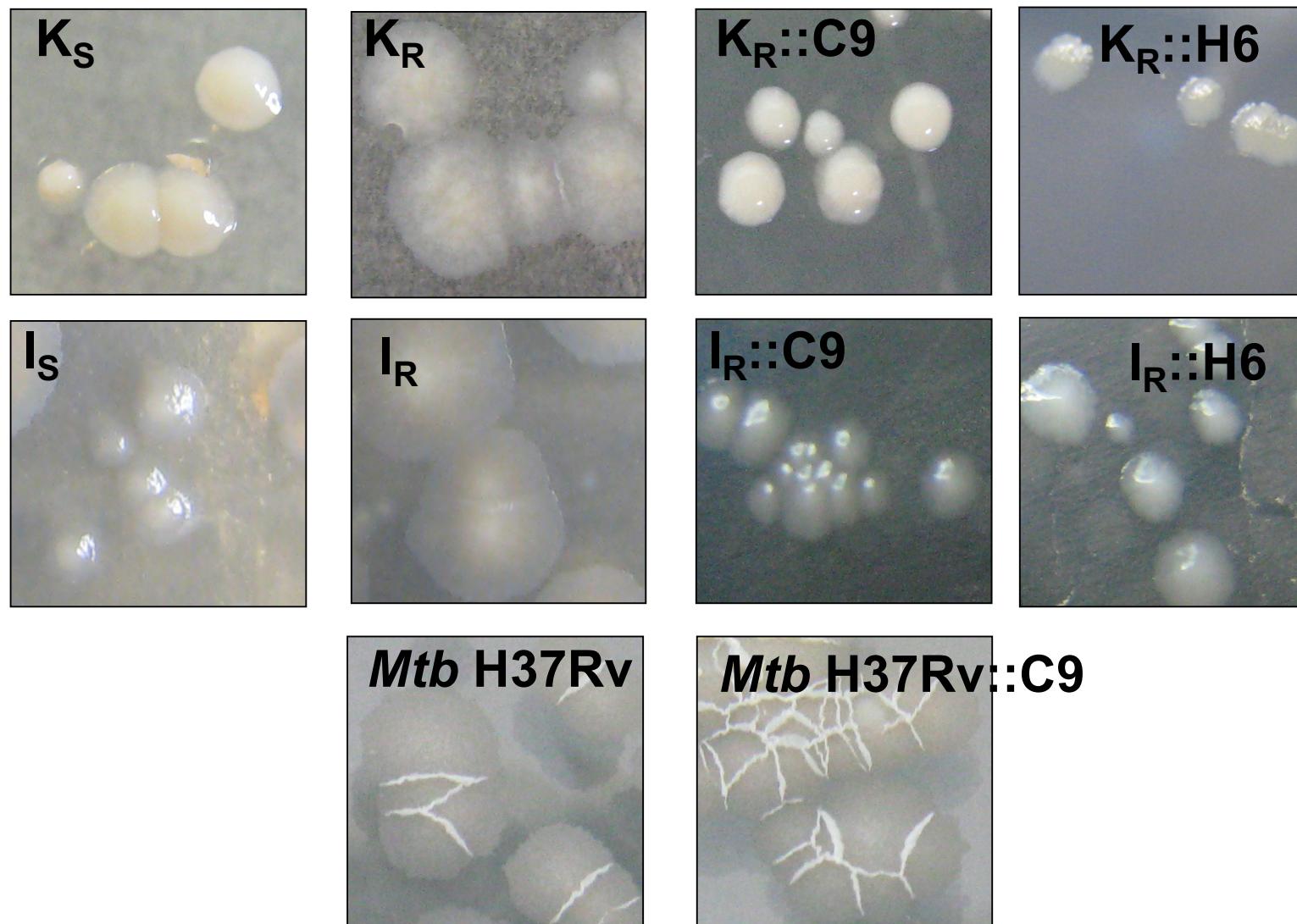
# Spontaneous rough *M. canettii* K variant has recombined *pks5* & lost acyltransferase encoding gene *pap*



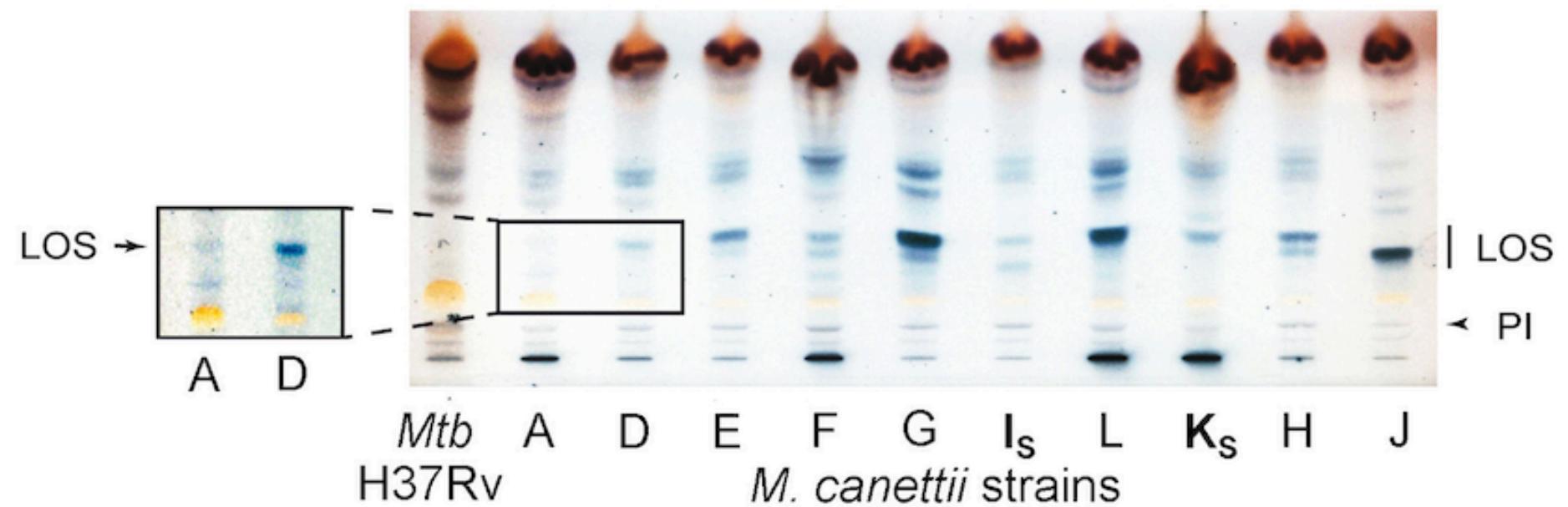
# Construction of complementation vector library using *M. canetti* A



Transformation with cosmids C9 and H6 restores smooth morphotypes in *M. canettii* K<sub>R</sub> and I<sub>R</sub> strains, but not in *Mtb* H37Rv

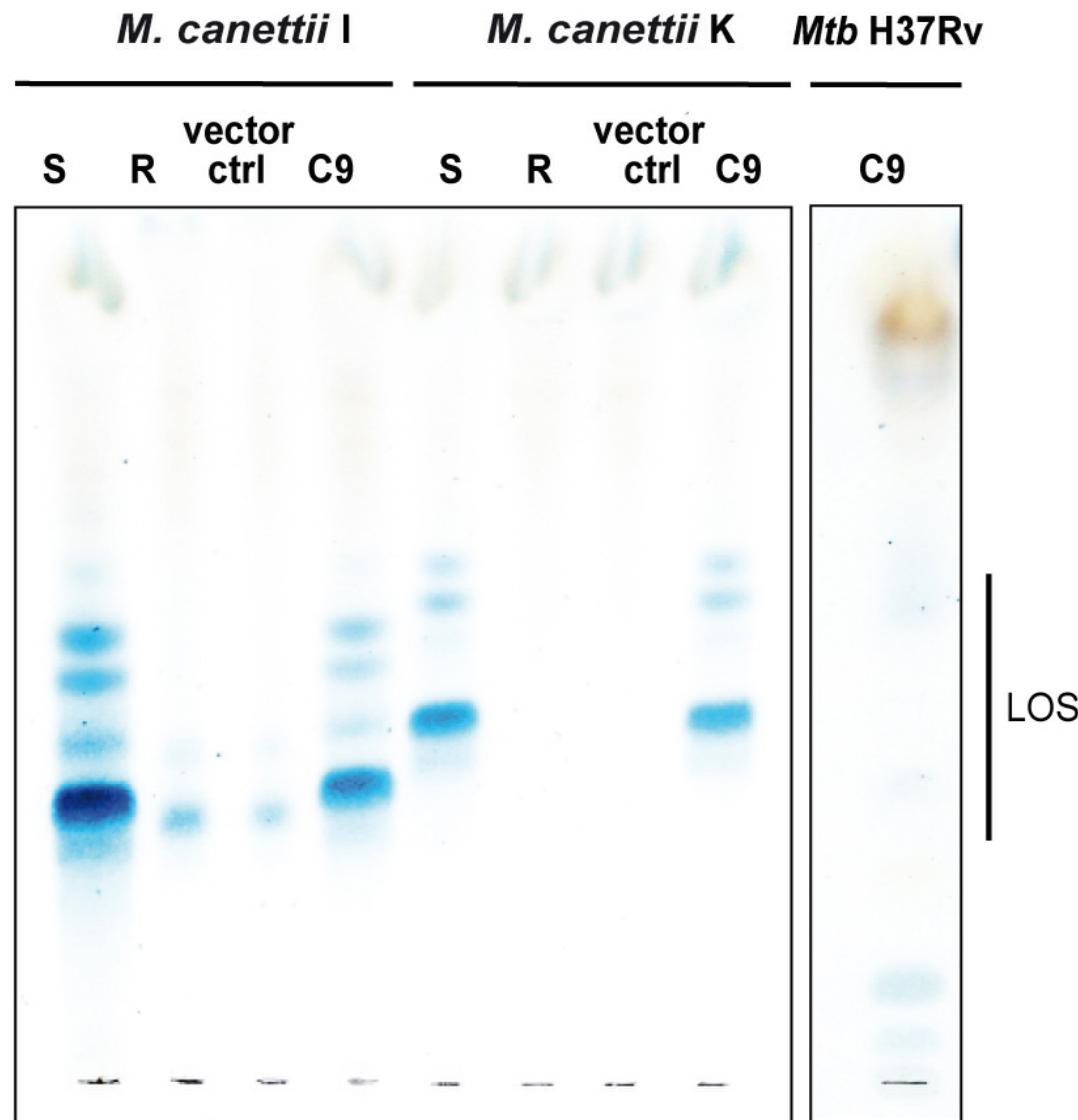


## *M. canettii* strains produce Lipooligosaccharide (LOS)



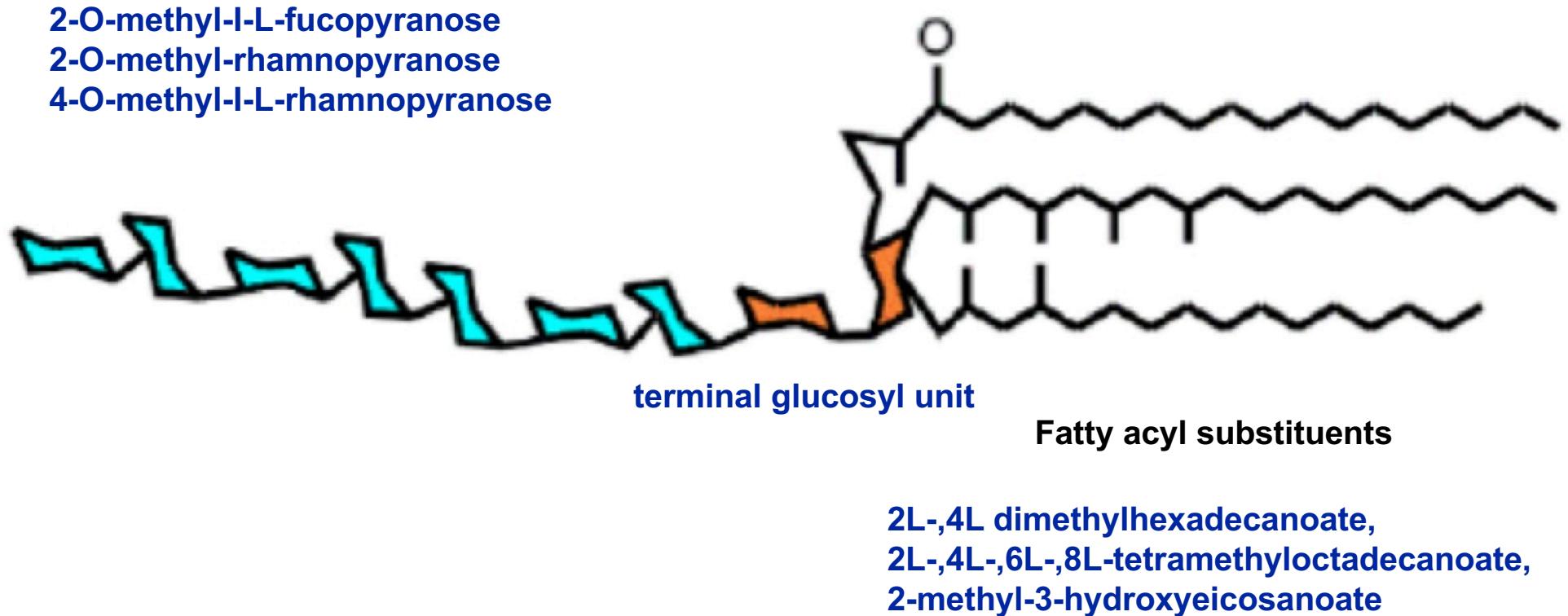
Daffé et al., Biochemistry 1991; van Soolingen, IJSM 1997; Boritsch et al., Nat Microbiol., 2016

# Transformation with cosmid C9 restores lipooligosaccharide synthesis *M. canettii* K<sub>R</sub> and I<sub>R</sub> strains, but not in *Mtb* H37Rv

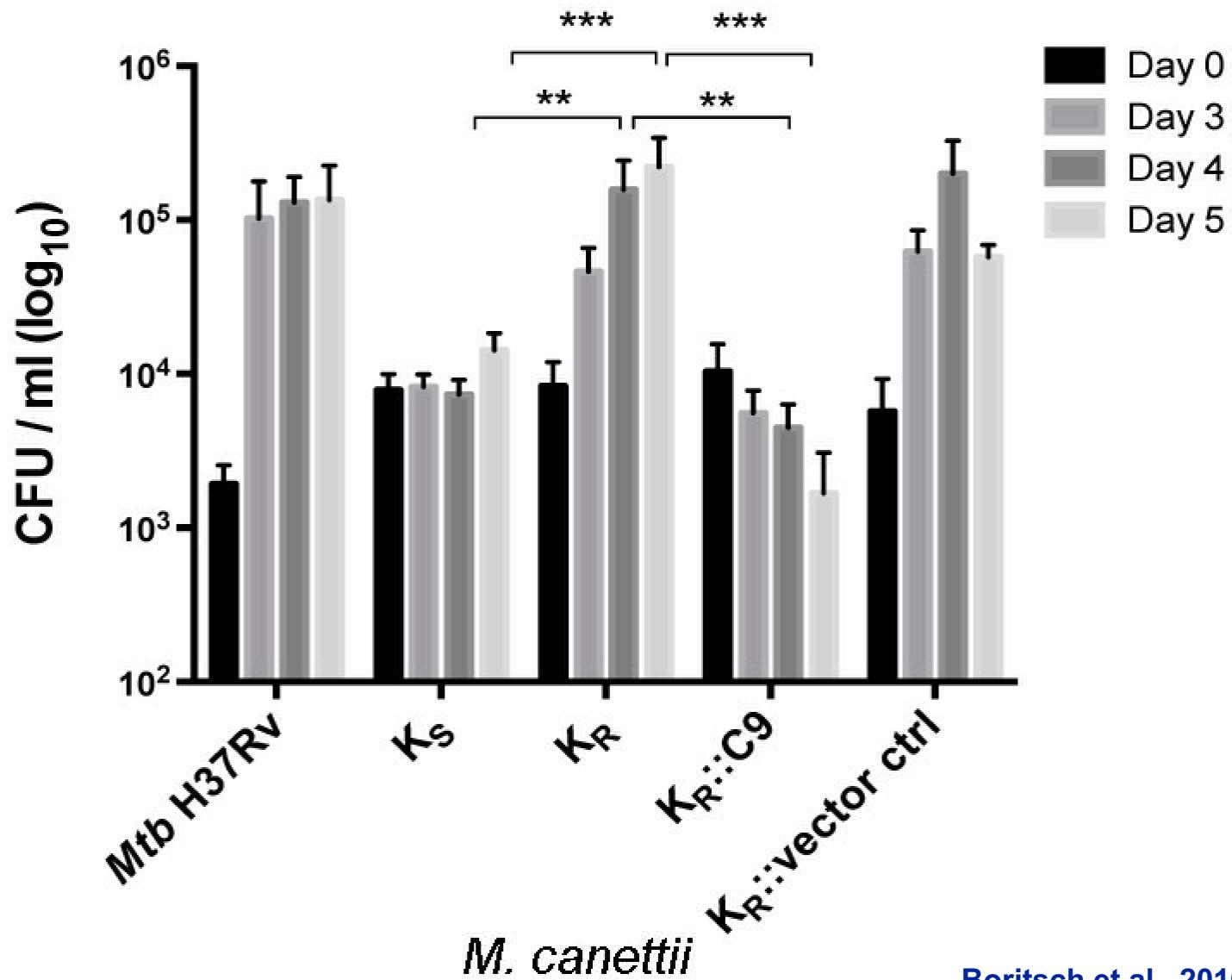


Boritsch et al., 2016, Nat Microbiol

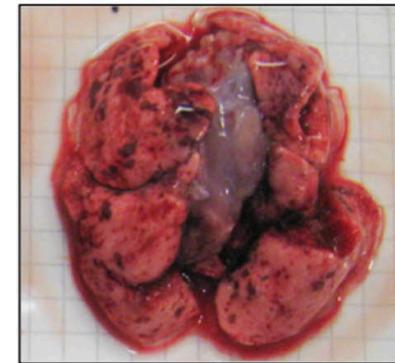
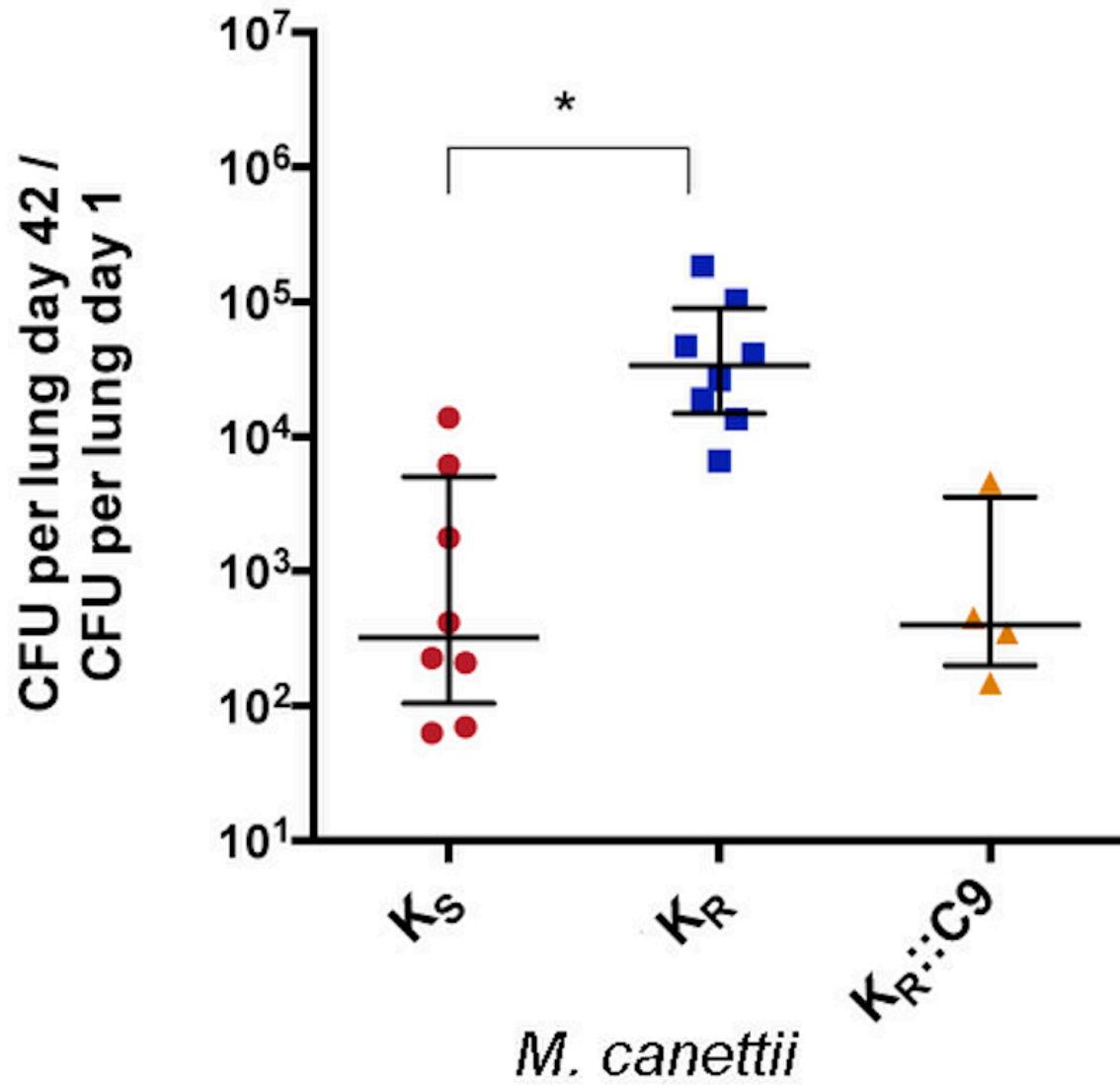
# Structure of lipooligosaccharide (LOS) of *M. canettii*



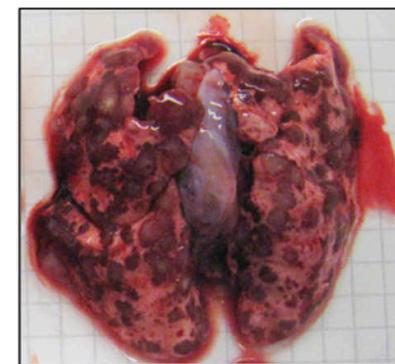
# *M. canettii* K<sub>R</sub> strains show increased virulence in THP-1 cells



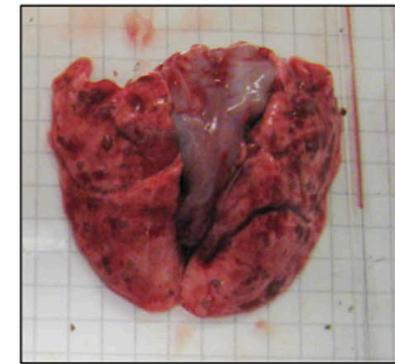
## *M. canettii* K<sub>R</sub> strains show increased virulence in guinea pig lungs



K<sub>S</sub>

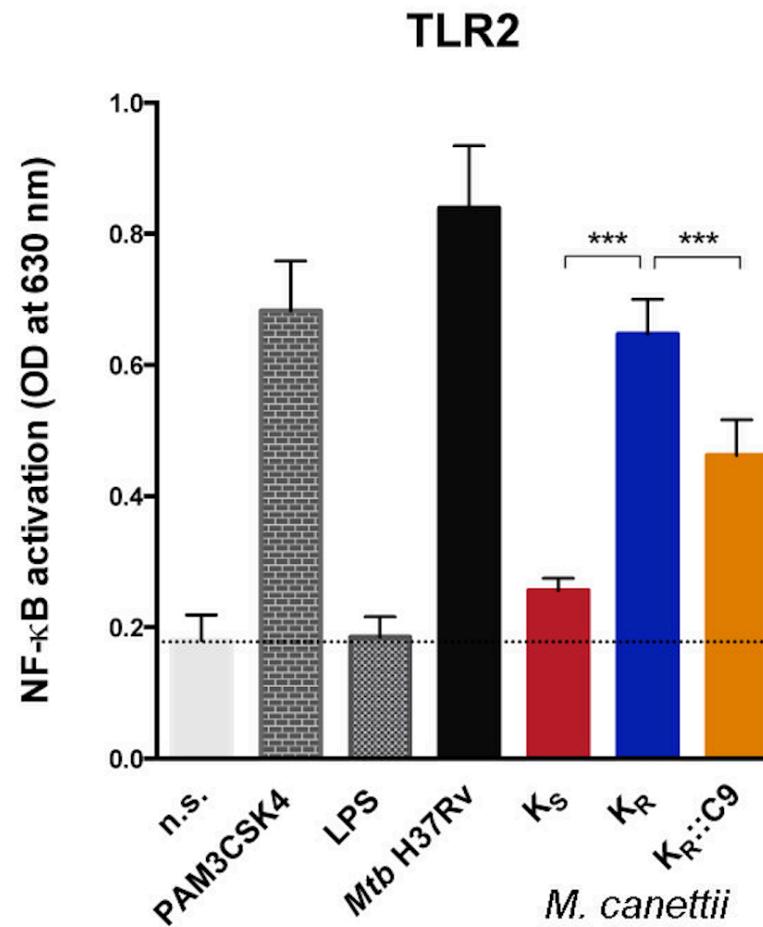


K<sub>R</sub>



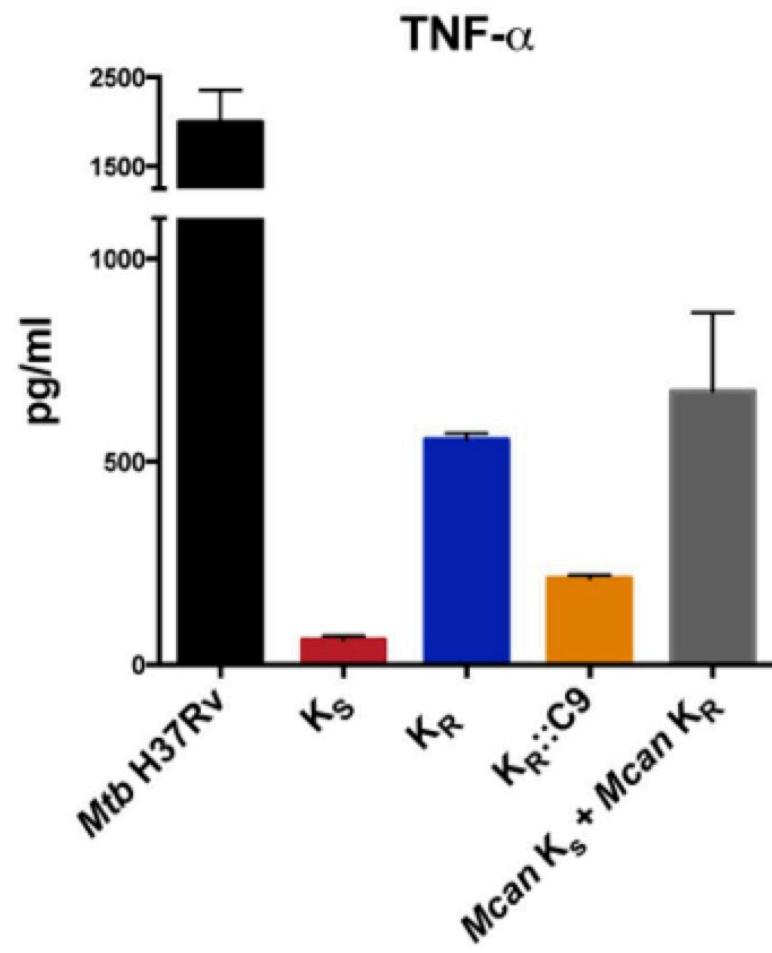
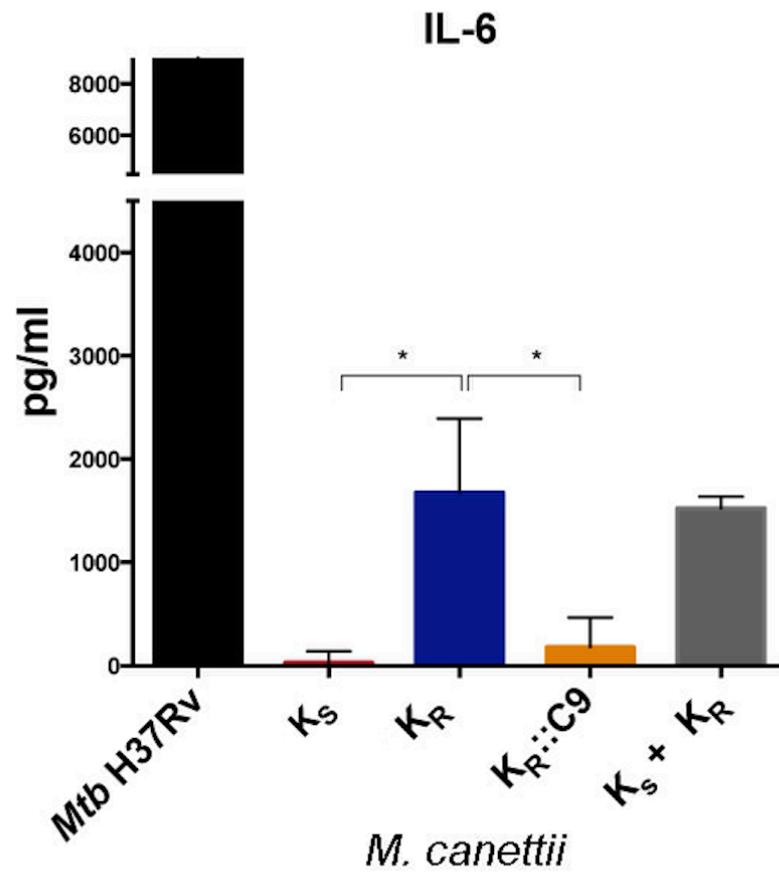
K<sub>R</sub>::C<sub>9</sub>

## Differential interaction of S and R morphotypes with TLR2



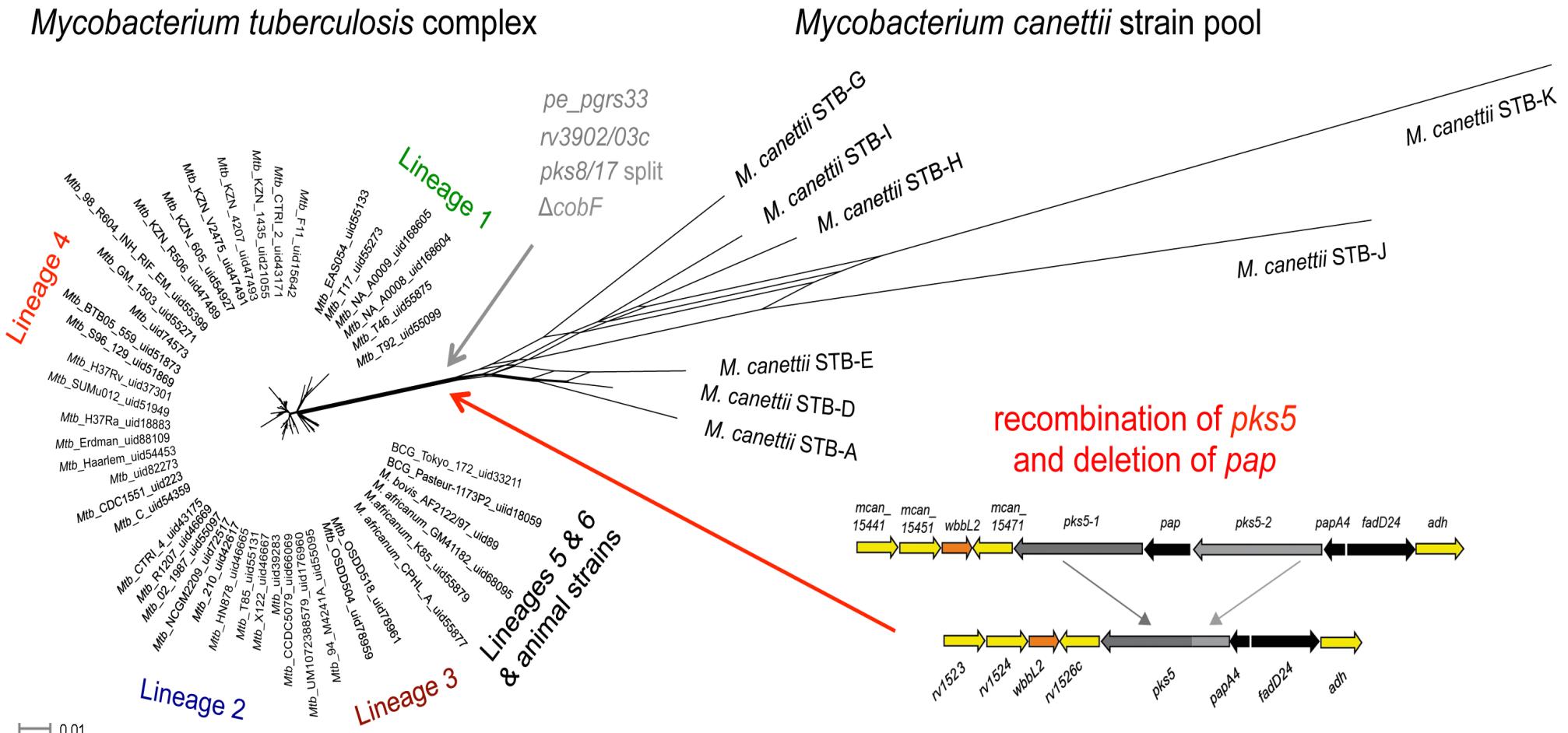
Human TLR2- or TLR4-transfected HEK293 cells were incubated with S and R morphotypes at an MOI of 1. At 24 h pi, levels of secreted embryonic alkaline phosphatase reporter gene under the control of NF- $\kappa$ B were measured with a spectrophotometer.

## *M. canettii* K<sub>R</sub> strains induce increased cytokine release than K<sub>S</sub>

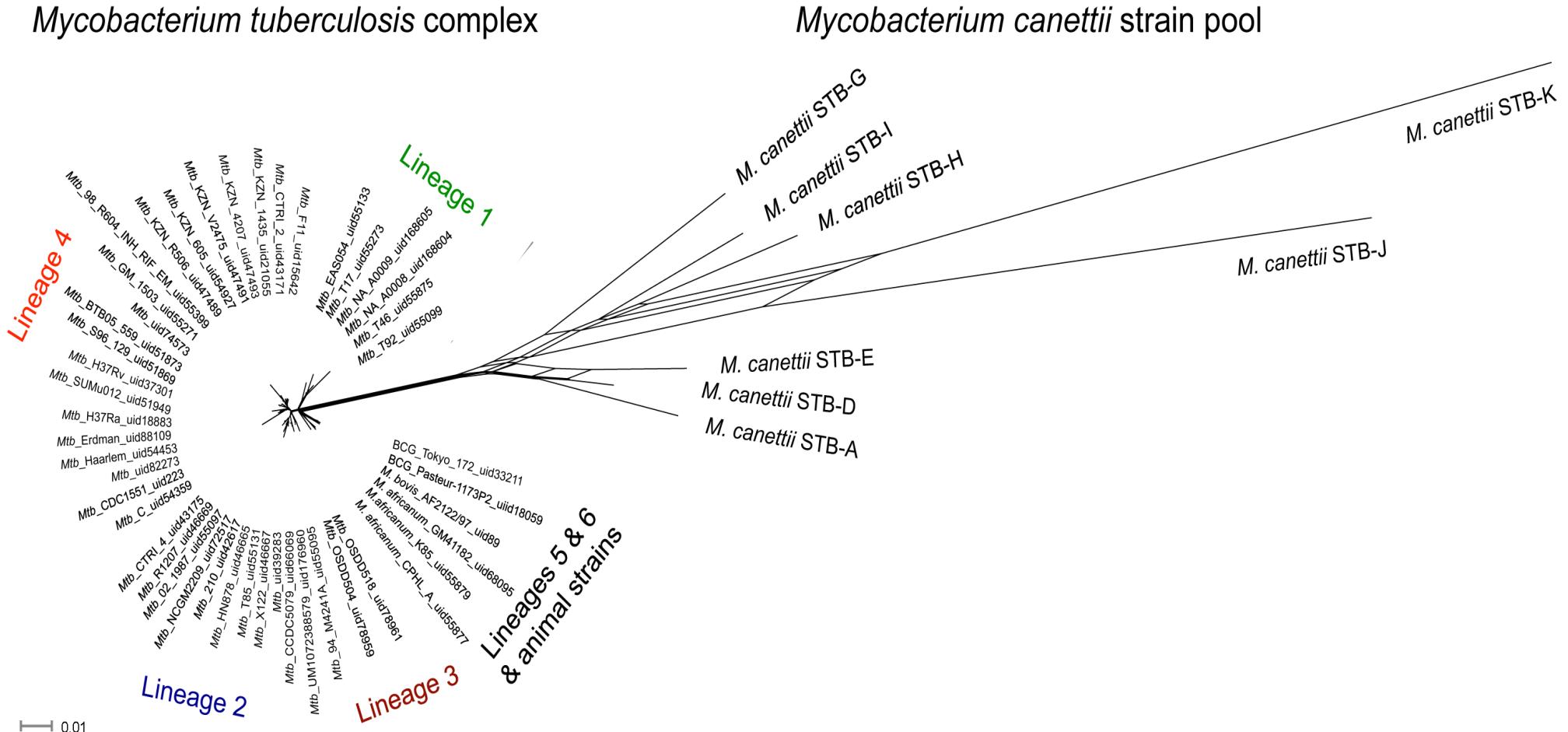


IL-6 production in BM-DCs upon infection with different *M. canettii* morphotypes.

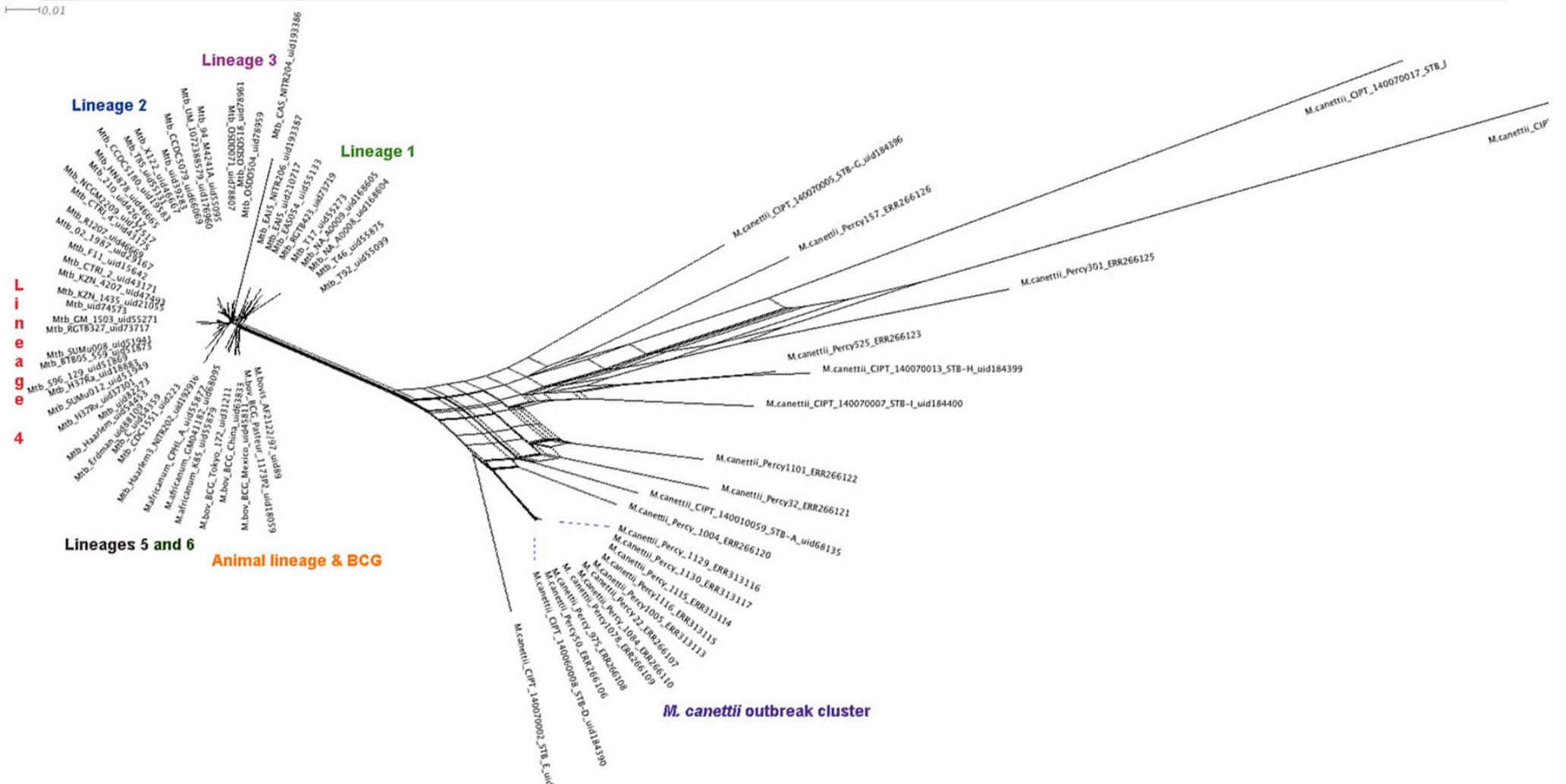
# Recombination of *pks5* genes → key event in evolution of tubercle bacilli



# *M. canettii* strains show non-clonal, mosaic genomic structure



# *M. canettii* strains show non-clonal, mosaic genomic structure



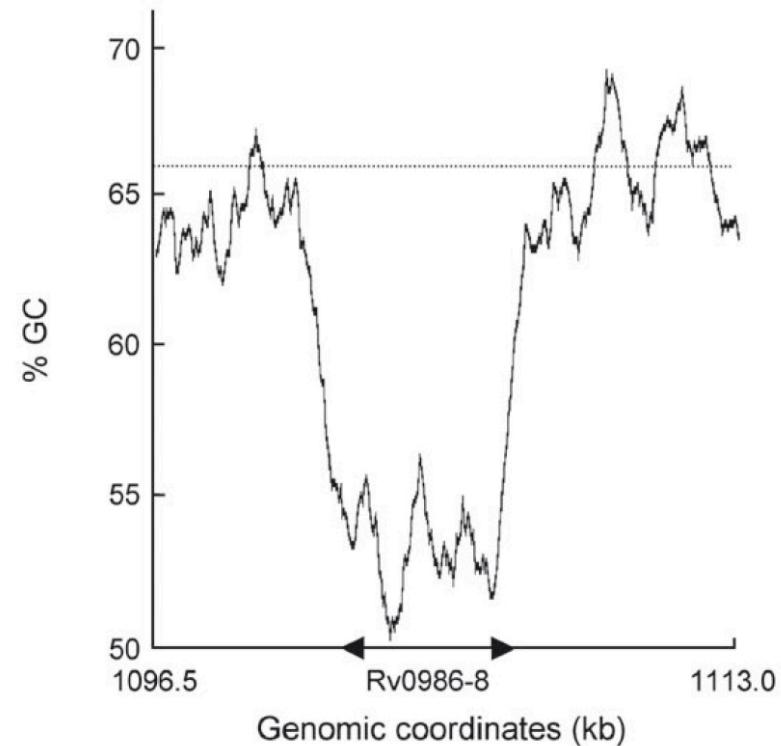
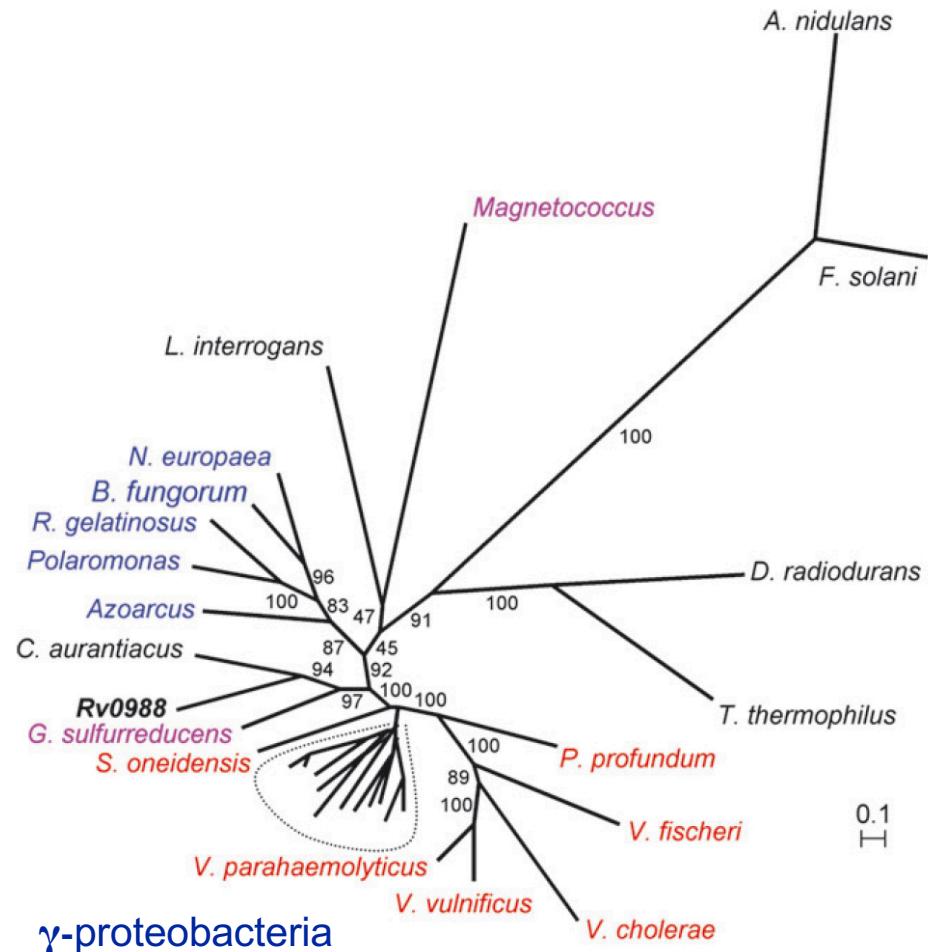
Blouin et al., 2014, *Emerg Inf Dis.*

Boritsch et al. 2014, *Mol Microbiol.*

*M. canettii* → a model to investigate how inter-strain horizontal gene transfer (HGT) is organized in tubercle bacilli

- important aspect – as theoretical possibility of transfer of resistance mutations exists
- MTBC members show ancient traces of HGT

Ancient traces of horizontal gene transfer (HGT) are visible in *M. tuberculosis* and other tubercle bacilli, e.g. *Rv0986-87-88*



Rosas-Magallanes, Mol. Biol. Evol. 2006

Question: Is DNA transfer still active in Mtb & other tubercle bacilli?

→ In *M. smegmatis*, HGT is accomplished by DCT

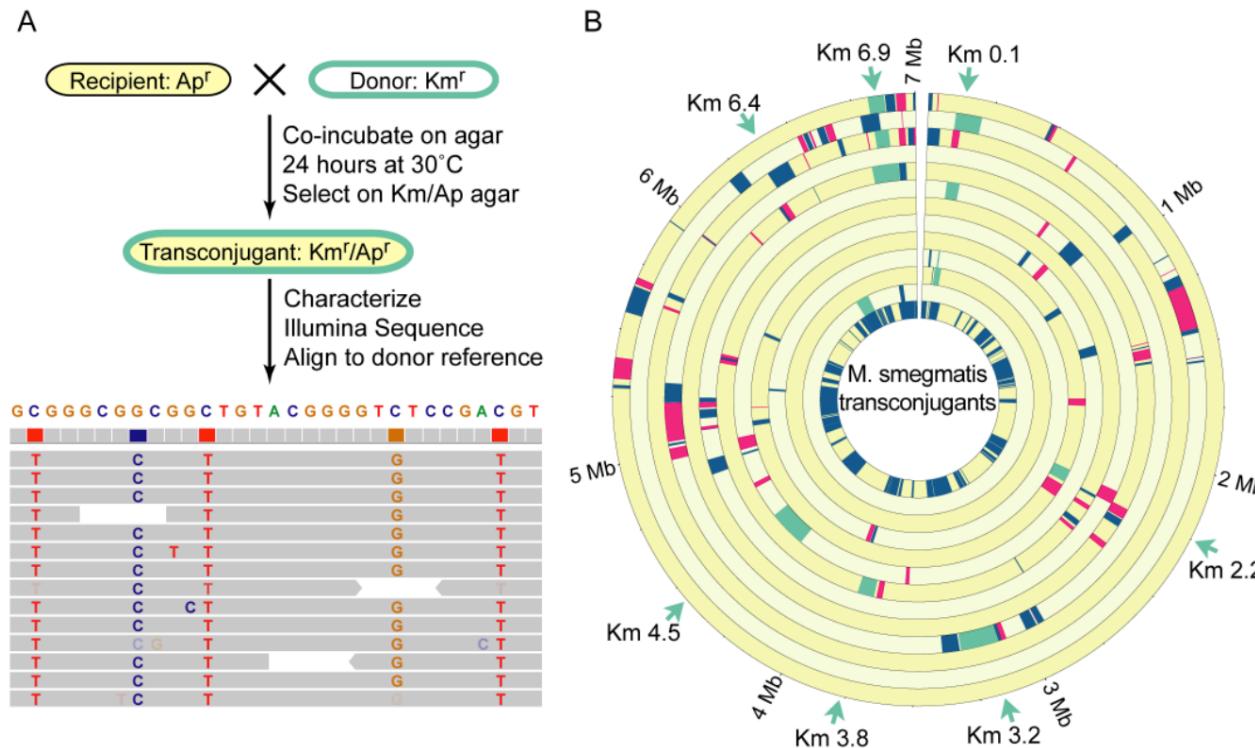
OPEN  ACCESS Freely available online

PLOS | BIOLOGY

# Distributive Conjugal Transfer in Mycobacteria Generates Progeny with Meiotic-Like Genome-Wide Mosaicism, Allowing Mapping of a Mating Identity Locus

Todd A. Gray<sup>1,2\*</sup>, Janet A. Krywy<sup>2</sup>, Jessica Harold<sup>1</sup>, Michael J. Palumbo<sup>1</sup>, Keith M. Derbyshire<sup>1,2\*</sup>

**1** Division of Genetics, Wadsworth Center, New York State Department of Health, Albany, New York, United States of America, **2** Department of Biomedical Sciences, University at Albany, Albany, New York, United States of America



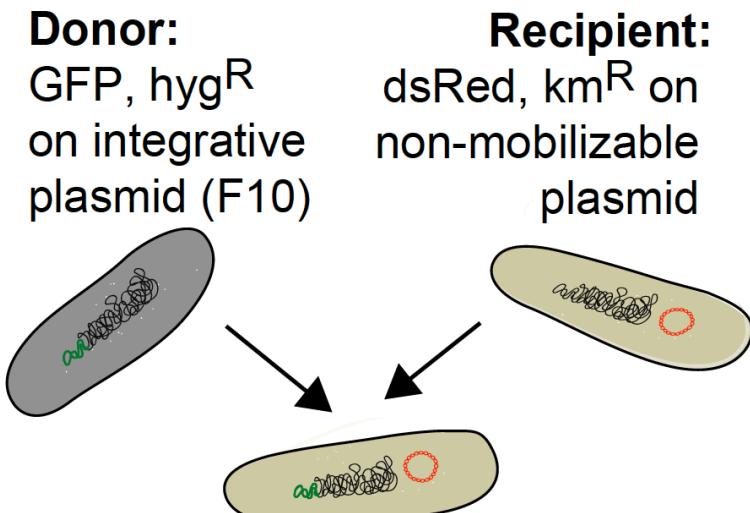
# Filter-exposure assay adapted for *M. canettii* strains



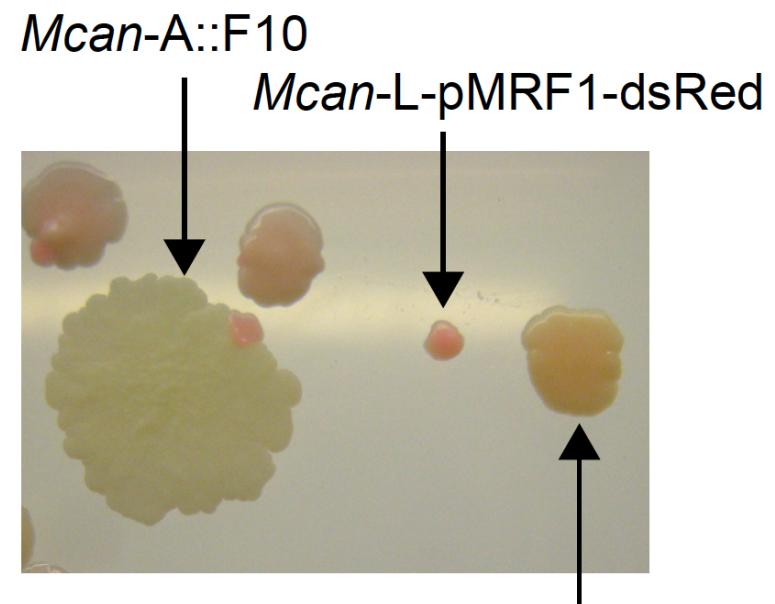
## Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria

Eva C. Boritsch<sup>a</sup>, Varun Khanna<sup>b</sup>, Alexandre Pawlik<sup>a</sup>, Nadine Honoré<sup>a</sup>, Victor H. Navas<sup>c</sup>, Laurence Ma<sup>d</sup>, Christiane Bouchier<sup>d</sup>, Torsten Seemann<sup>e</sup>, Philip Supply<sup>f</sup>, Timothy P. Stinear<sup>g</sup>, and Roland Brosch<sup>a,1</sup>

<sup>a</sup>Institut Pasteur (IP), Unit for Integrated Mycobacterial Pathogenomics, 75015 Paris, France; <sup>b</sup>IP, Hub Bioinformatique et Biostatistique, C3BI, Unité de Services de la Recherche, UMR 3756, IP/CNRS, 75015 Paris, France; <sup>c</sup>IP, Lymphocyte Cell Biology Unit, 75015 Paris, France; <sup>d</sup>IP, PF1-Plate-Forme Génomique, 75015 Paris, France; <sup>e</sup>Victorian Life Sciences Computation Initiative, University of Melbourne, Carlton, VIC 3053, Australia; <sup>f</sup>INSERM U1019, CNRS UMR 8204, Center for Infection and Immunity, Institut Pasteur de Lille, Université de Lille, 59000 Lille, France; and <sup>g</sup>Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, VIC 3000, Australia



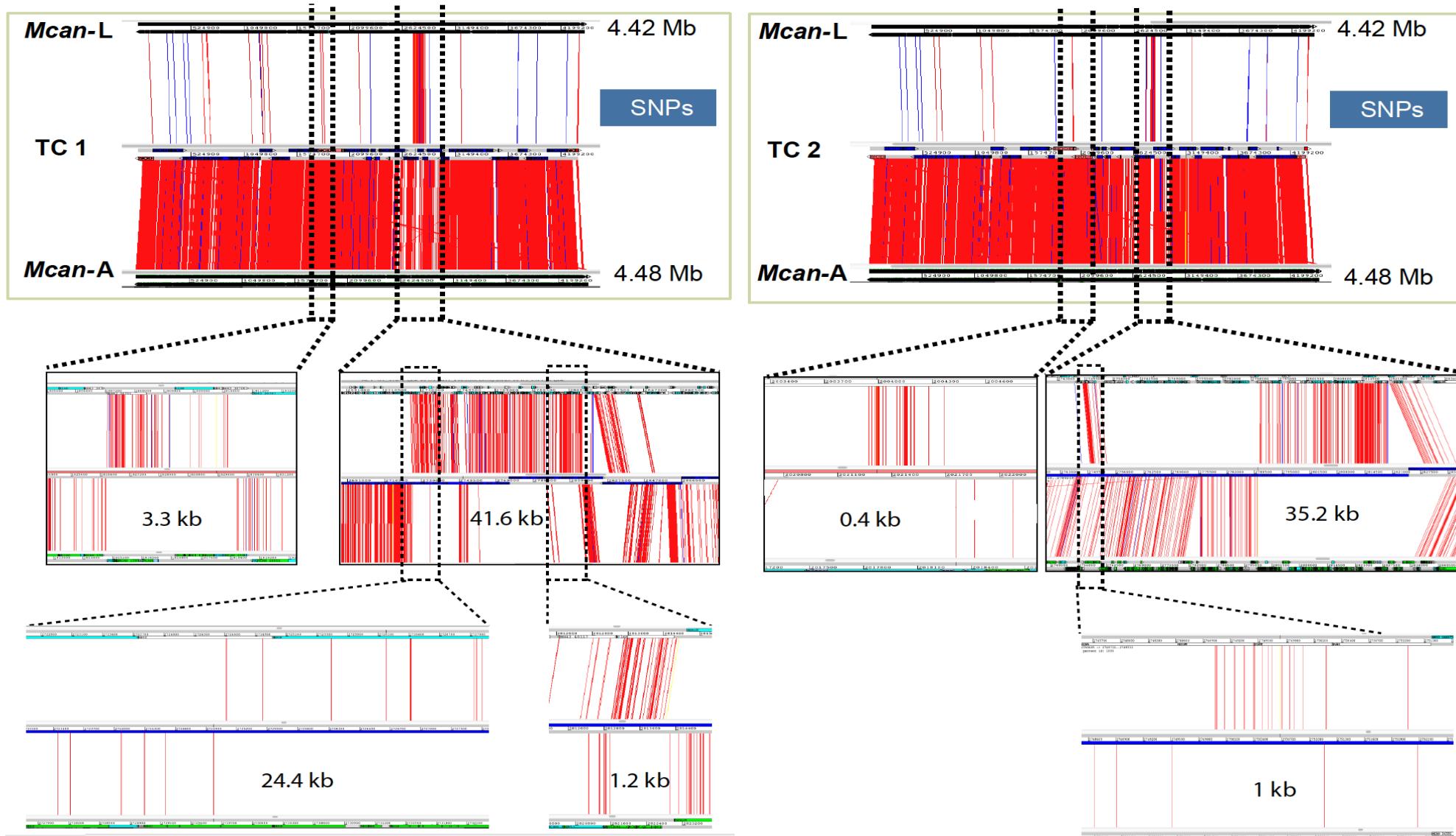
**Transconjugant (TC):**  
GFP, dsRed, hyg<sup>R</sup> and km<sup>R</sup>



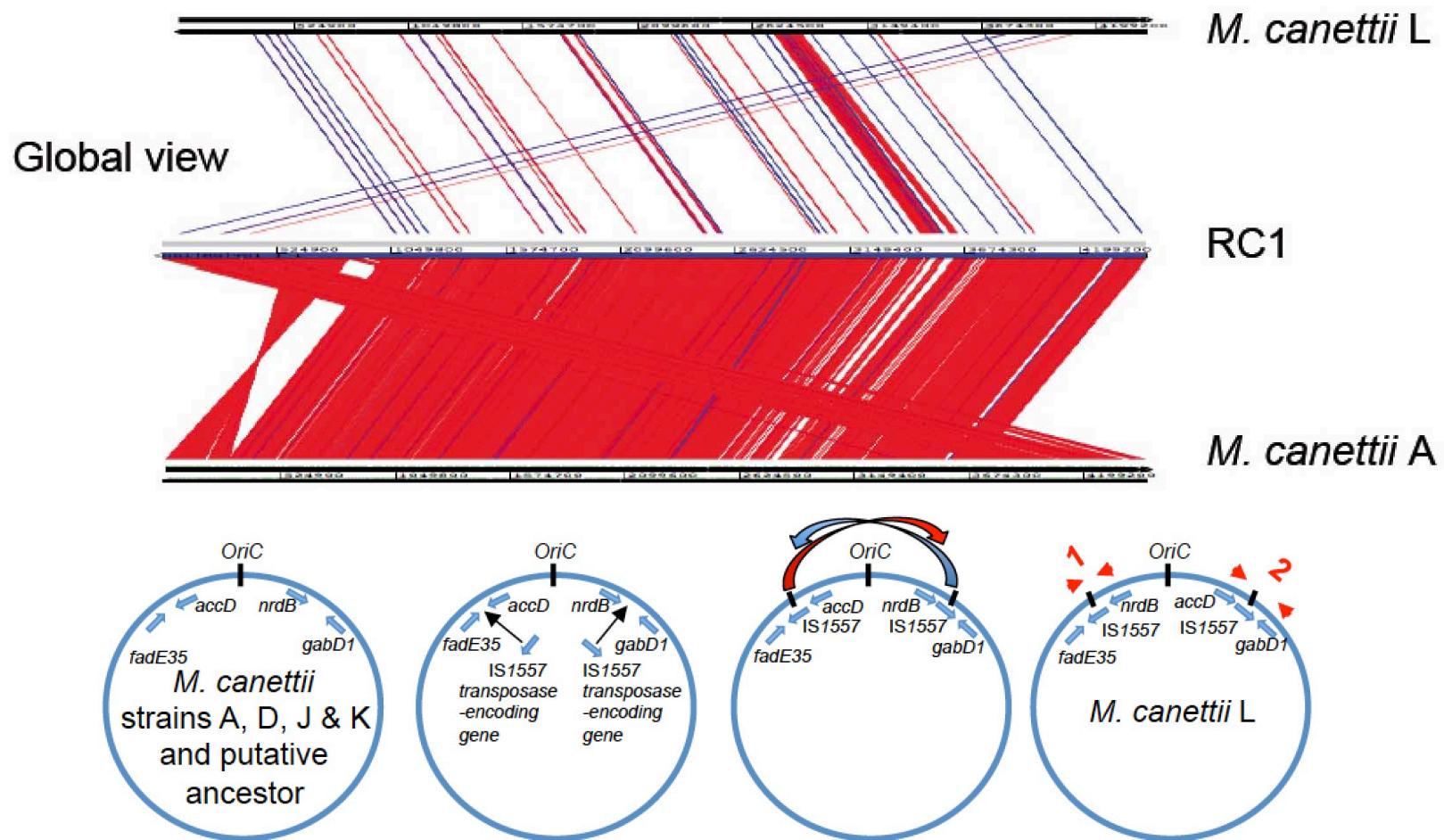
**Transconjugant**

Boritsch et al., 2016, PNAS

# Identification of transferred DNA segments in transconjugants

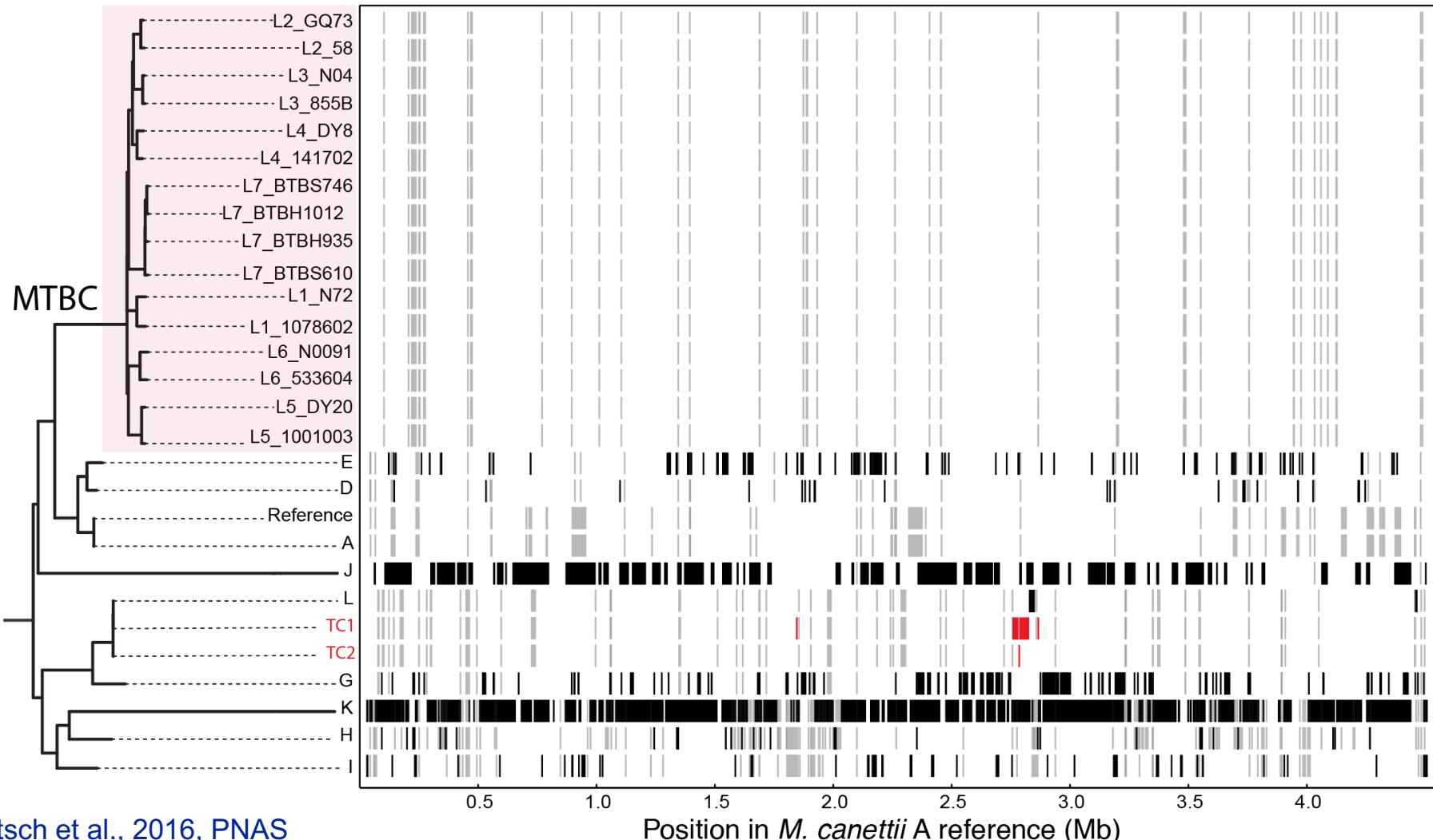


## PacBio-sequence confirmation identified inversion of 435 kb in strain L

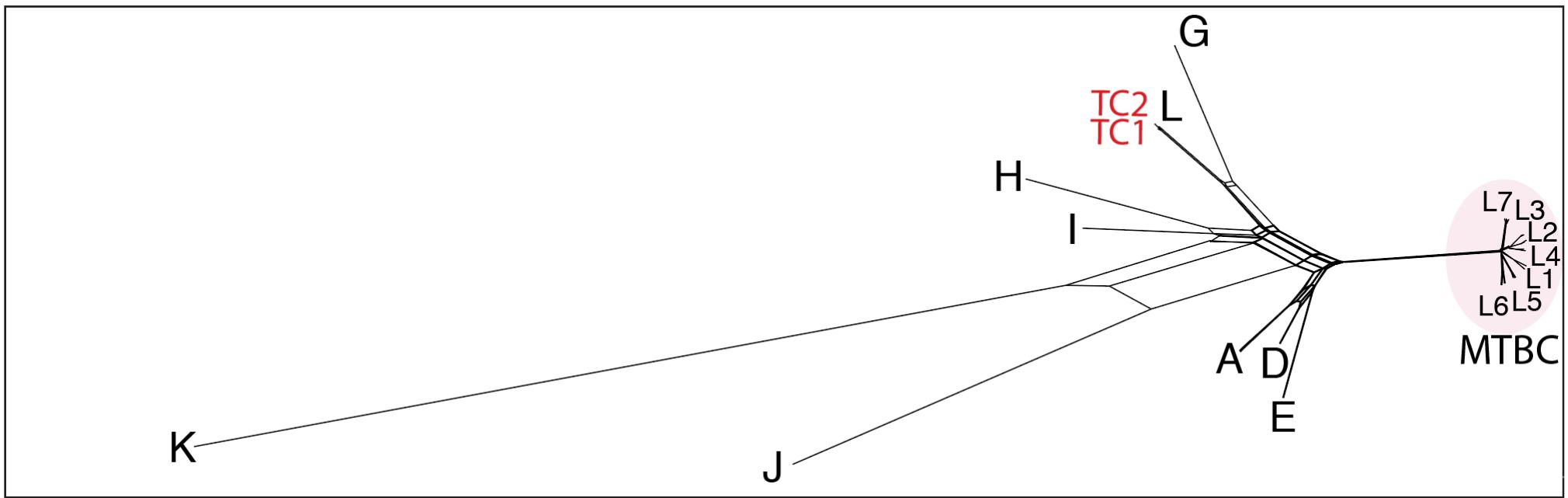


**Fig. S5A:** Model for occurrence of putative genomic inversion of a 435 kb-sized genomic segment spanning the origin of replication in strain *M. canettii* L, as observed by large sequence read genome sequencing (PacBio) of recombinants RC1 and RC2 (Fig. S4). Results suggest that the inversion is already present in wildtype strain *M. canettii* L (STB-L), unlike previously reported in GenBank entry F0203508. 1, 2, fragments to be amplified as shown in Fig. S5B.

# ClonalFrameML analysis of recombination shows recombination regions in *M. canettii* strains (only)



# NeighbourNet analysis of nine *M. canettii* genomes, 2 transconjugants and 16 MTBC (lineage 1-7) genomes



Only 2 recombinants shown  
Similar experiments with Mtb strains yielded 0 recombinants

# Results are different from published analyses based on *in silico* analyses

Research

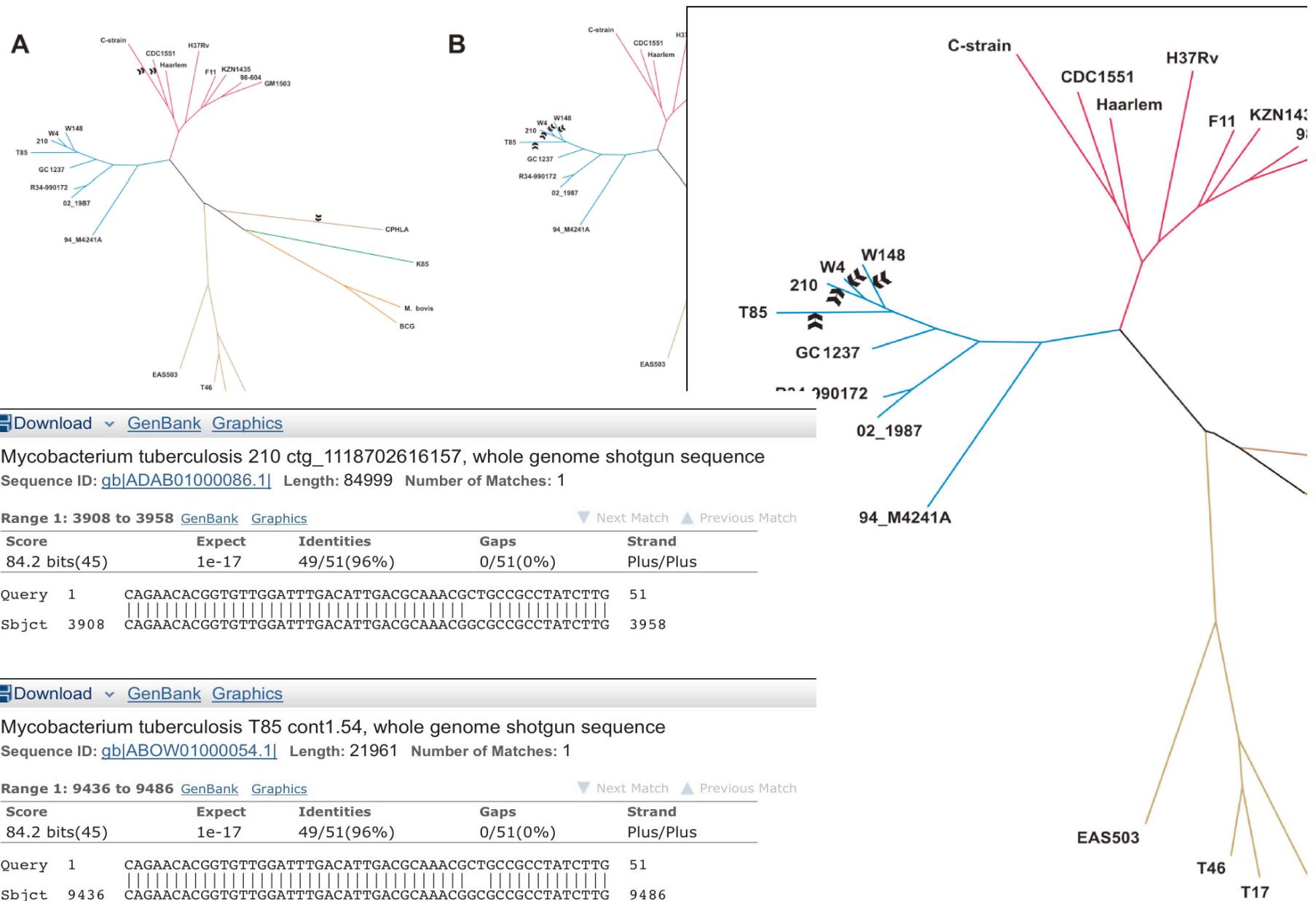
## After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection

Amine Namouchi,<sup>1,7</sup> Xavier Didelot,<sup>2</sup> Ulrike Schöck,<sup>3</sup> Brigitte Gicquel,<sup>1,6</sup> and Eduardo P.C. Rocha<sup>4,5,6</sup>

<sup>1</sup>Unité de Génétique Mycobactérienne, Institut Pasteur, 75015, Paris, France; <sup>2</sup>Department of Statistics, Oxford University, OX1 3TG Oxford, United Kingdom; <sup>3</sup>GATC Biotech AG, D-78467 Konstanz, Germany; <sup>4</sup>Institut Pasteur, Microbial Evolutionary Genomics, 75015, Paris, France; <sup>5</sup>CNRS, UMR3525, 75015, Paris, France

Many of the most virulent bacterial pathogens show low genetic diversity and sexual isolation. Accordingly, *Mycobacterium tuberculosis*, the deadliest human pathogen, is thought to be clonal and evolve by genetic drift. Yet, its genome shows few of the concomitant signs of genome degradation. We analyzed 24 genomes and found an excess of genetic diversity in regions encoding key adaptive functions including the type VII secretion system and the ancient horizontally transferred virulence-related regions. Four different approaches showed evident signs of recombination in *M. tuberculosis*. Recombination tracts add a high density of polymorphisms, and many are thus predicted to arise from outside the clade. Some of these tracts match *Mycobacterium canettii* sequences. Recombination introduced an excess of non-synonymous diversity in general and even more in genes expected to be under positive or diversifying selection, e.g., cell wall component genes. Mutations leading to non-synonymous SNPs are effectively

mainly based on Clonal frame analysis of minimum of two incongruent closely spaced SNPs

**Figure S4. E**

3418350; D:  
of ticks in arr

## Conclusions and perspectives I

- Comparative genomics showed that *M. bovis* is not the ancestor of *M. tuberculosis*
- Tuberculosis is not a zoonosis
- The clonal *M. tuberculosis* complex (MTBC) contains 7 human strain lineages (L1-L7) with L2-L3-L4 being the most recently evolved ones (bottleneck TbD1 deletion) and several animal-adapted strains that branch next to L6 strains.
- some strains geographically highly restricted (e.g. L7), whereas other can be found worldwide (animal-adapted strains)

## Conclusions and perspectives II

- Smooth morphotype of *M. canettii* due to production of LOS  
→ important for an environmental organism ?
- Production of LOS was lost during evolution towards MTBC  
→ gain in virulence
- Transfer of multiple, large DNA regions between *M. canettii*  
→ resembling Distributive Conjugal Transfer seen in *M. smegmatis*
- *M. tuberculosis* and other clonal MTBC strains (L1-7 & animal L8) seem to have lost the ability of DNA transfer
- important message for predicting transfer of drug resistance & fitness mutations (not observed)

# The UPMI team



Laurence  
**LANGLAIS**



Laleh  
**MAJLESSI**



Roxane  
**SIMEONE**



Wafa  
**FRIGUI**



Alexandre  
**PAWLIK**



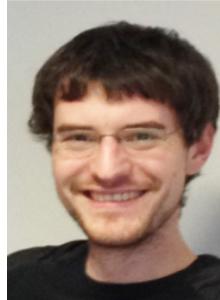
Ludovic  
**TAILLEUX**



Roland  
**BROSCH**



Fadel  
**SAYES**



Mickael  
**ORGEUR**



Jan  
**MADACKI**



Florence  
**BROSSIER** (50%)

Contrat d'interface pour  
hospitaliers



Alexandre  
**GIRAUD GATINEAU**



Nadine  
**HONORE**



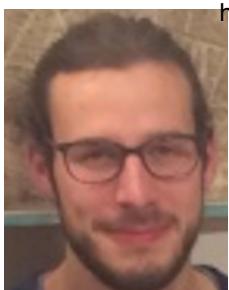
Alessandro  
**CASCIOPERO**



Fabien  
**LE CHEVALIER**



Eva  
**BORITSCH**



Matthias  
**GRÖSCHEL**



Christophe  
**QUEVAL**



Louis  
**ATES**

# Collaborations and funding

IP Pasteur  
Sylvain BRISSE  
Jost ENNINGA  
Claude LECLERC  
Pedro ALZARI  
Varun KHANNA

PF1 Genomics  
Laurence MA  
Christiane BOUCHIER

Daria BOTTAI  
Univ. Pisa

Tim STINEAR  
Torsten SEEMANN  
Univ Melbourne

Philip SUPPLY  
Priscille BRODIN  
IP Lille  
Stephane CRUVEILLER  
Claudine MEDIGE  
Genoscope Evry  
Julian PARKHILL  
Sanger Institute  
Jean Louis HERRMANN  
Univ. Versailles/St Quentin

Laurent KREMER  
Univ. Montpellier  
Wilbert BITTER  
Univ. Amsterdam



Wladimir MALAGA  
Mamadou DAFFE  
Christoph GUILHOT  
IPBS Toulouse  
Jesús GONZALO-ASENSIOA  
Carlos MARTIN  
Univ. Zaragoza



## FP7 consortia

NOVSEC-TB  
NEWTBVAC  
MM4TB  
PreDICT-TB



Horizon 2020  
partner in  
TBVAC-2020

