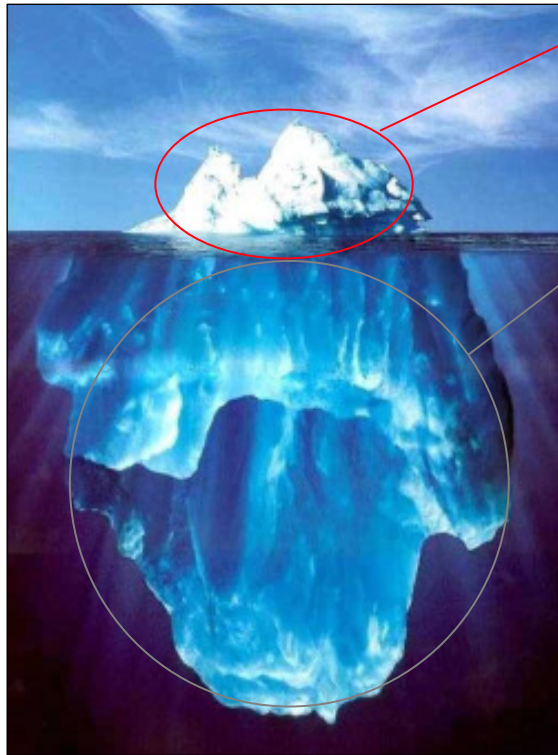


Evolution and diversity in *Mycobacterium tuberculosis*

Brigitte Gicquel

**Mycobacterial Genetics Unit
Institut Pasteur, Paris**

1.7 million deaths
15 million TB cases



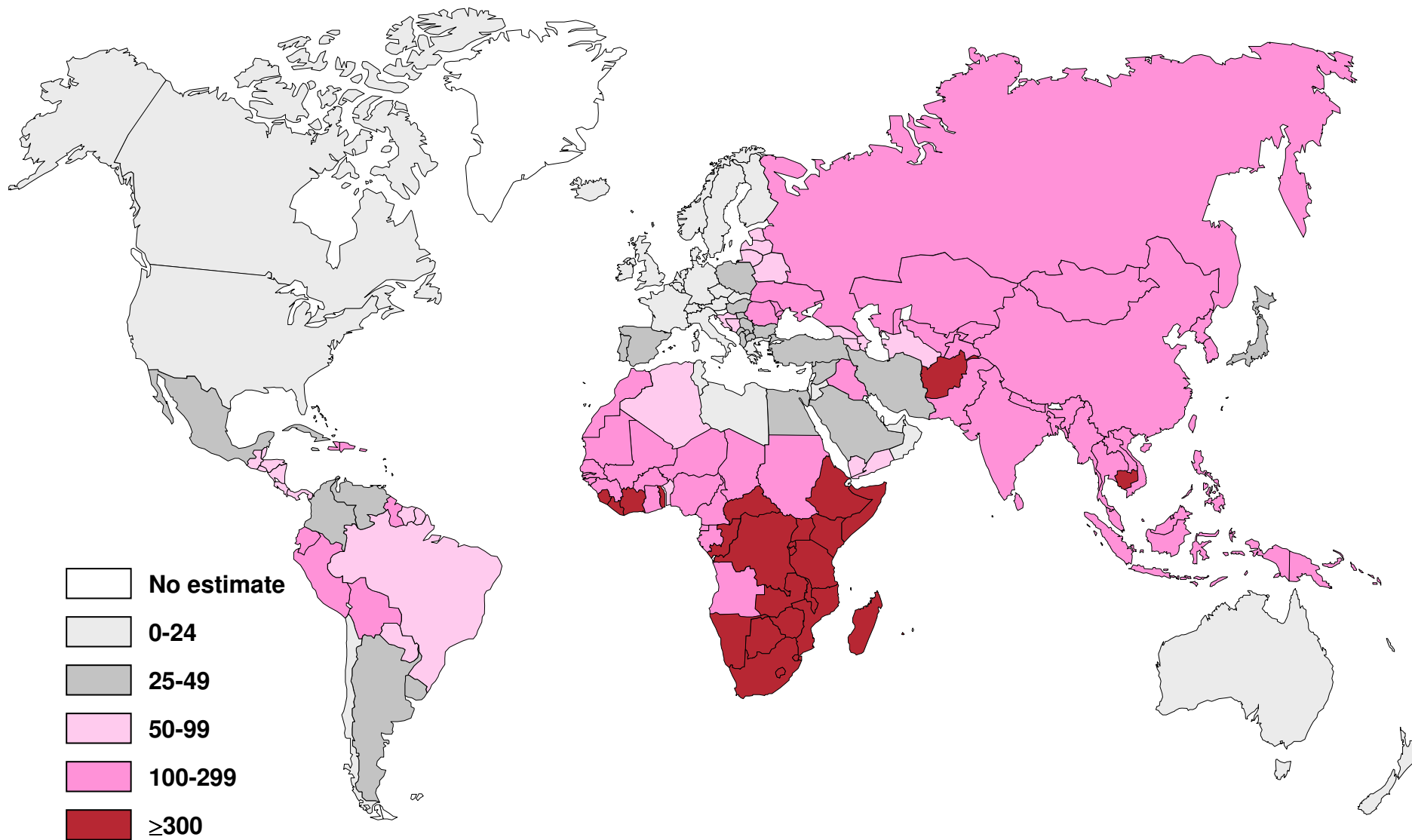
1.9 billion PPD+ individuals

WHO projections by 2020

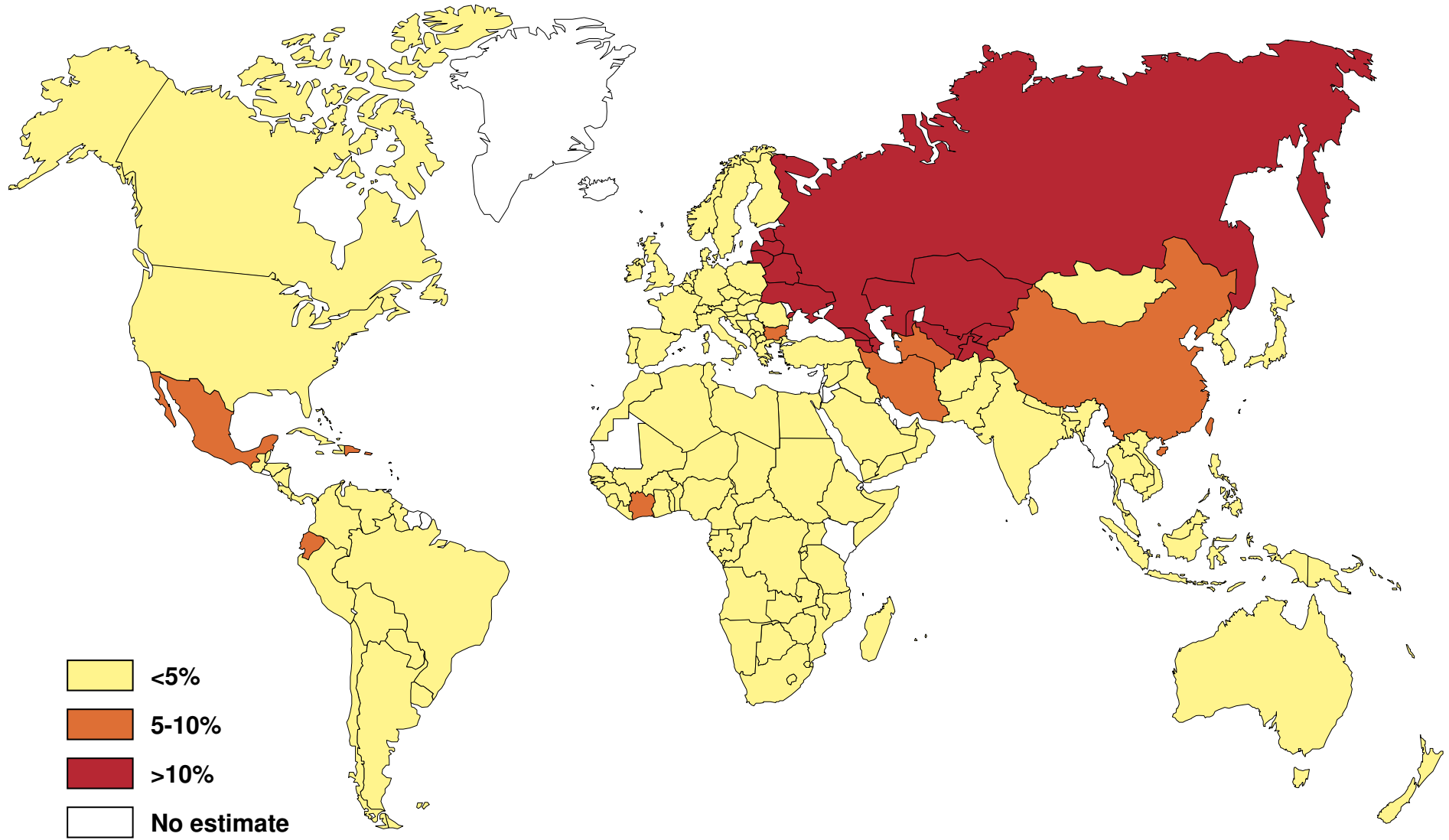
- 1 billion people newly infected
- 200 million people sick
- 35 million deaths

Estimated TB incidence rate

(per 100,000 population - 2004)



% MDR among reported TB cases

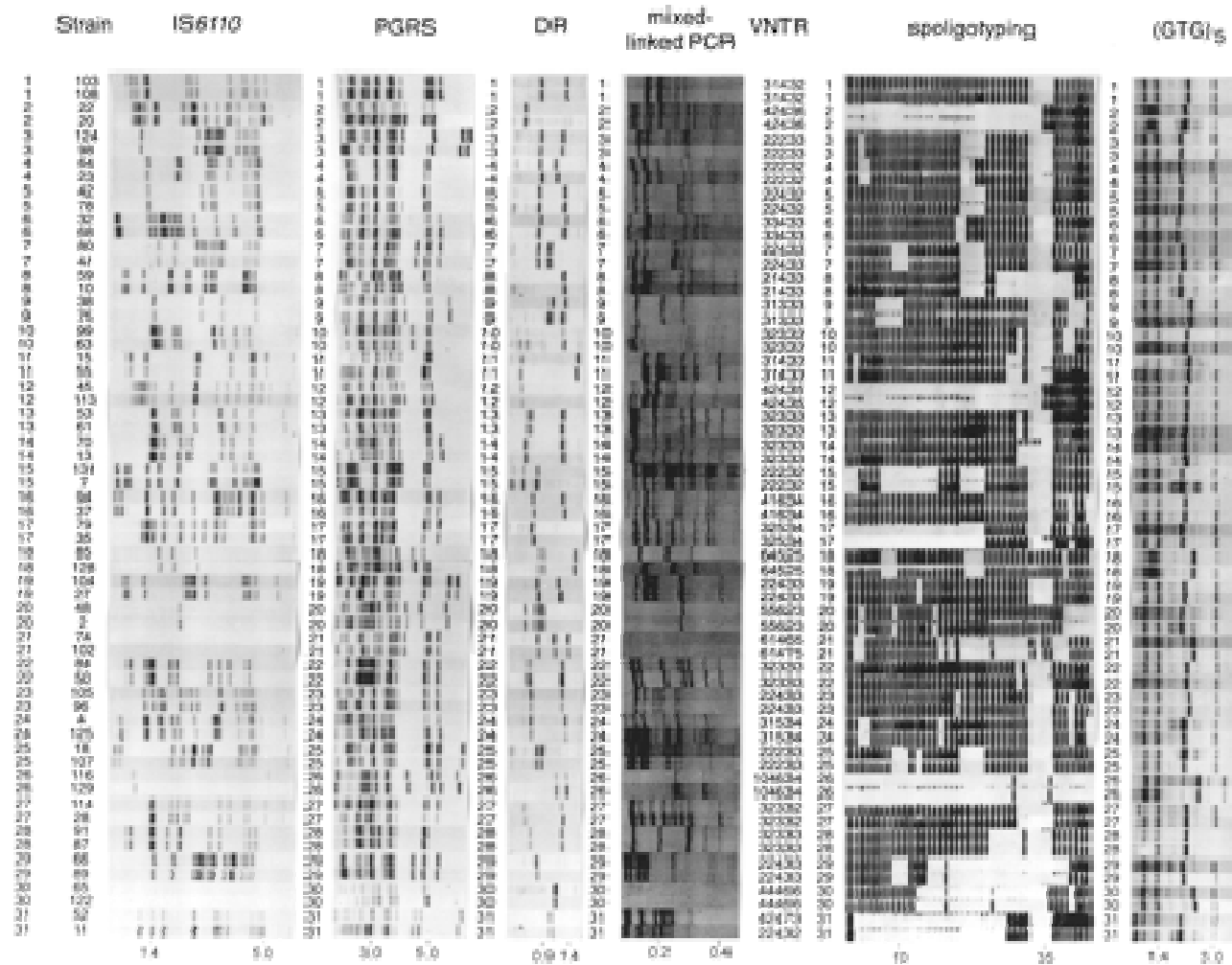


Levels of allelic polymorphism recorded in genes of pathogenic bacteria

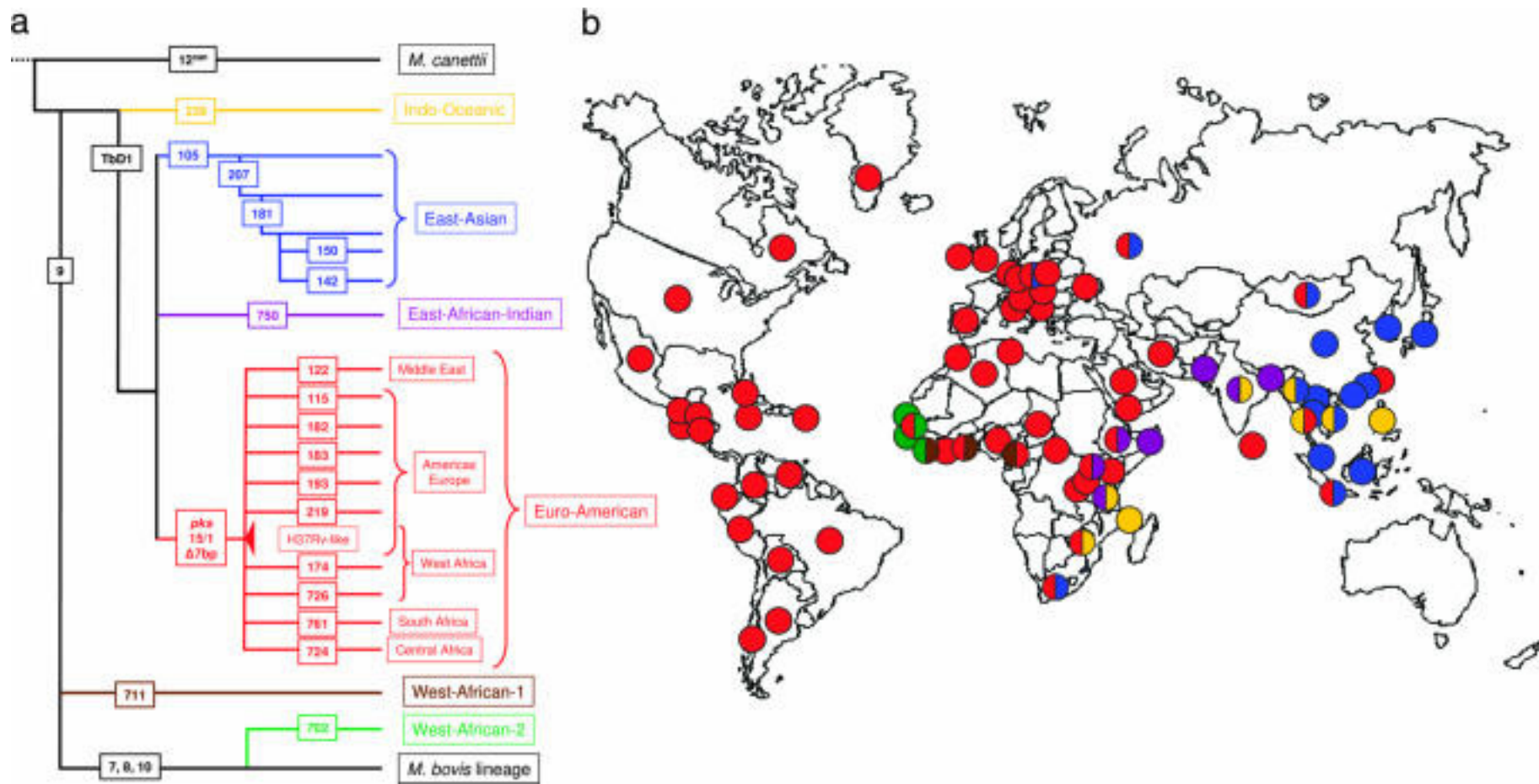
Organism	No. Of genes or gene segments	D _s	Relative Variation
<i>M. Tuberculosis complex</i>	26	<0.01	1
<i>Shigella sonnei</i>	2	0.01	1
<i>Vibrio cholerae</i>	1	0.41	41
<i>Streptococcus pyogenes</i>	3	1.02	102
<i>Neisseria meningitidis</i>	4	6.18	618
<i>M. avium-intracellulare</i>	1	10.10	1,010
<i>Escherichia coli</i>	11	11.77	1,117
<i>Borrelia burgdorferi</i>	3	≈20	≈2,000
<i>Salmonella enterica</i>	5	42.08	4,208

Sreevatsan S, et al.
Proc Natl Acad Sci U S A. 1997. 94:9869-74

Diversity of *Mycobacterium tuberculosis* clinical strains



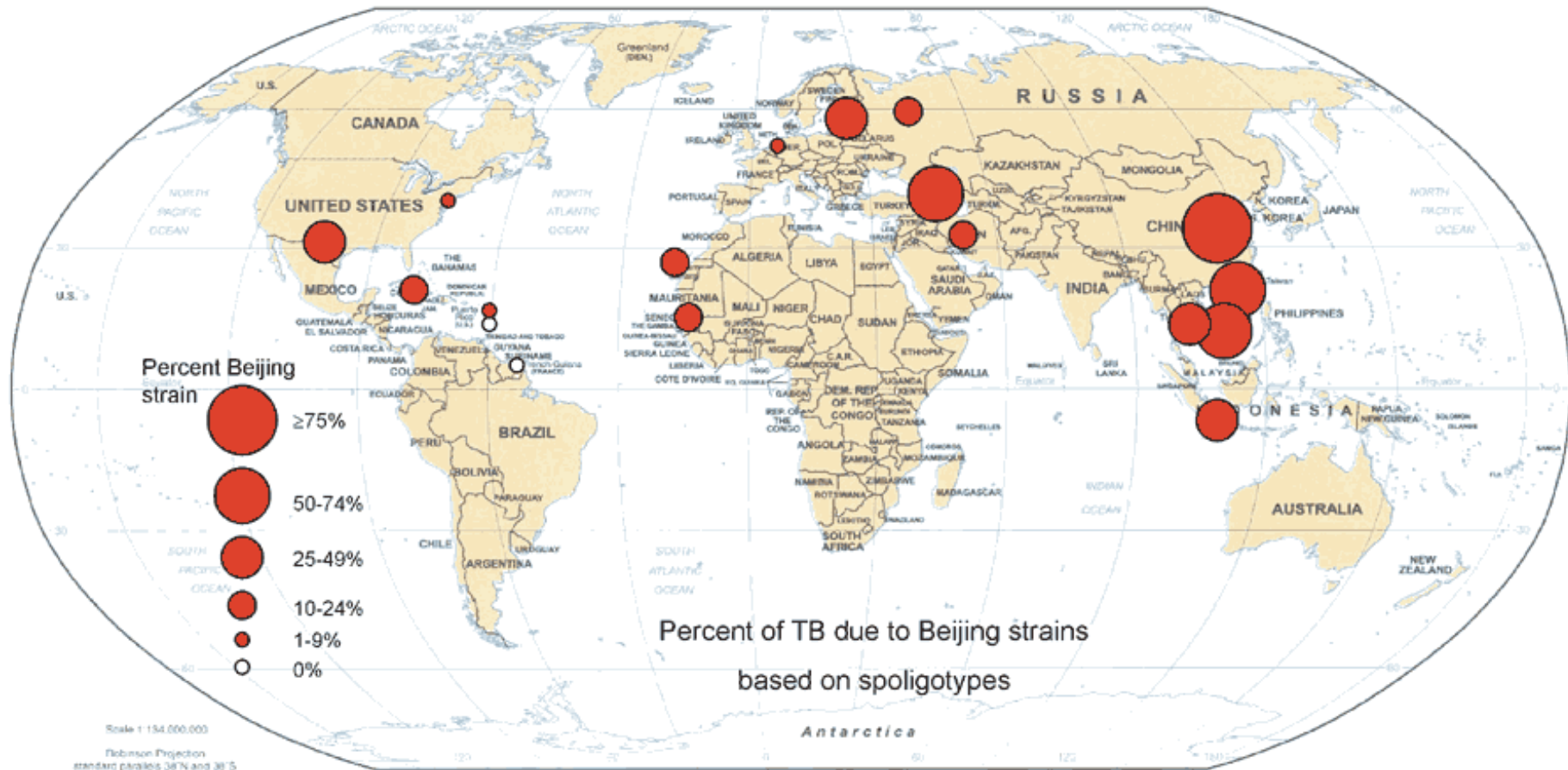
From Kremer K et al. 1999 J. Clin. Microbiol. 37:2607-18.



Gagneux S et al. Proc Natl Acad Sci U S A. 2006. 103:2869-73.

Worldwide Occurrence of Beijing/W Strains of *Mycobacterium tuberculosis*: A Systematic Review

Judith R. Glynn,* Jennifer Whiteley,* Pablo J. Bifani,† Kristin Kremer,‡ and Dick van Soolingen‡



*London School of Hygiene and Tropical Medicine, London, UK; †Institut Pasteur de Lille, France; and ‡National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Frequency of spontaneous *rpoB* mutations

	H37Rv 2.2+/-3.1	MTT1KO 33.9+/-12.1	MTT2KO 3.3+/-4.1	MTT3KO 3.2+/-3.	MTT4KO 3.1+/-3.0
	mc2/155 8.3+/-4.6	MST1KO 98.7+/-24.2	MST2KO 14.3+/-5.7	MST3KO 15.3+/-5.4	MST4KO 397.0+/-82.4
MT(X)	1.0	15.5	1.5	1.5	1.4
MS(X)	1.0	12	1.7	1.9	48.1

Dos Vultos et al. J. Bacteriol. 2006, 188:3159-3161

DNA repair enzymes in *M. tuberculosis*

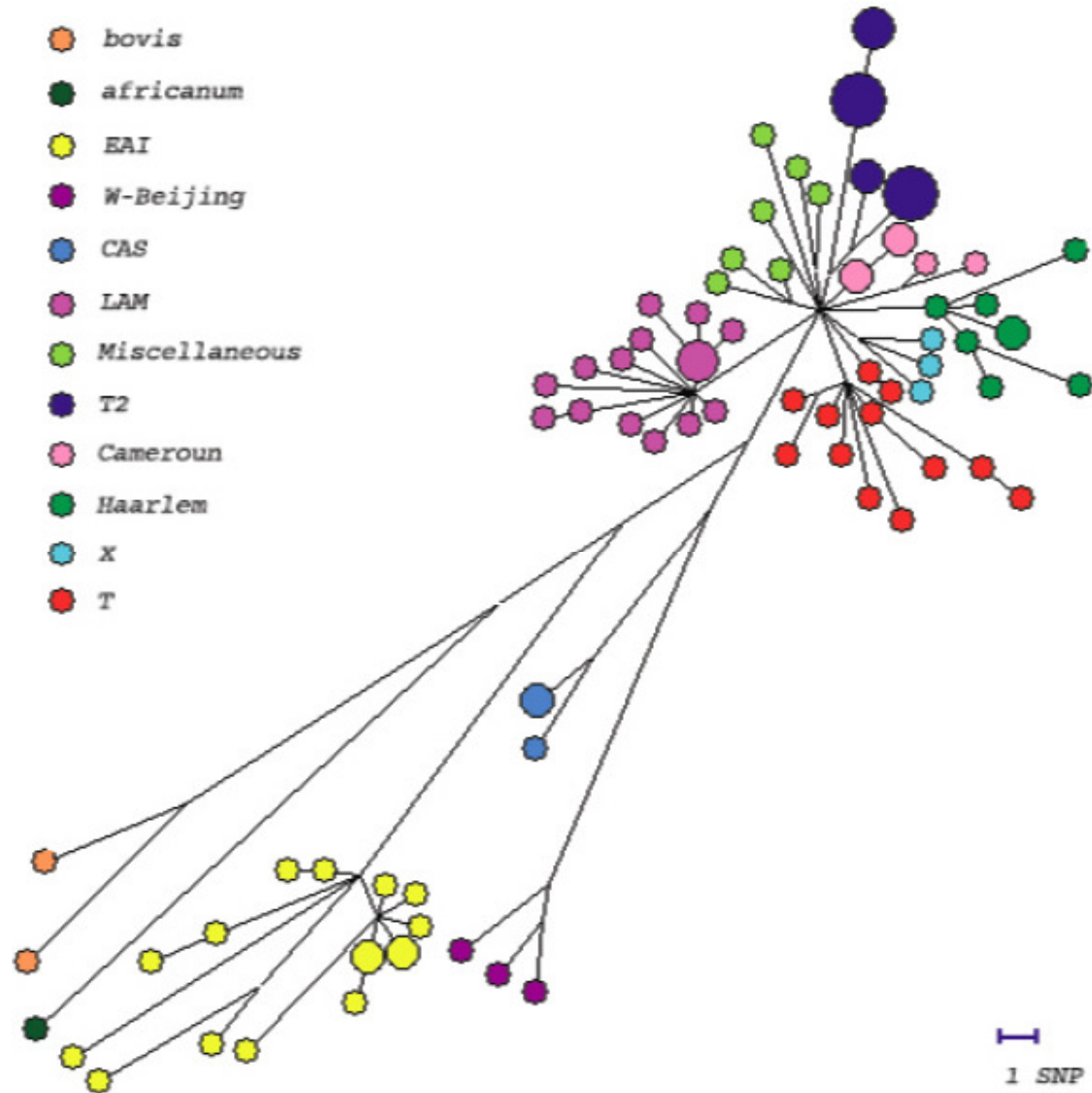
- Damage reversion: *ada, ogt*
- Damage excision:
 - Base excision repair: *tagA, alkA, mpg, ung, mutY, nth, mutM, nei, xthA, end, Rv0944, Rv2464c*
 - Nucleotide excision repair: *uvrA, uvrB, uvrC, uvrD, recA, lexA, mfd, ercc3, recX*
- Recombination: *recB, recC, recD, recQ, recN, recA, radA, recF, recO, recR, ssB, ruvA, ruvB, ruvC, recG, sbcD, topA, gyrA, gyrB, polA, ligA, uvrD, Rv3644c, ligA, ligB, ligC, ligD, dinX, dinP*
- Nucleotide pool maintenance: *mutT, dut, deoA*
- High fidelity DNA copy machine: *dnaQ, dnaZX, dnaN*
- Post-replicative mismatch correction

Dos Vultos *et al.* 2008 *PLoS ONE* 3(2):e1538

- **48 spoligotype-based diverse strains**
- **2 groups of clinical strains from two distinct geographic locations**
 - **Bangui, Central African Republic**
 - **Madagascar**
- **56 DNA repair, recombination & replication genes**
- **73 kbases per strain (1.66% H37rv genome)**
- **6.7 Mbases of sequence (1.5 H37Rv genome)**

3R gene diversity in *M. tuberculosis*

- 52 polymorphic genes
- 259 SNPs found
- 161 nsSNPs (62%)
- 131 singleton variations
- 3 stop codons
- 7 deletions
- 74 haplotypes



Highly polymorphic genes

- *recB, recC and recD*
- *polA*
- *alkA*
- *ligB, ligC and ligD*
- *dnaQ*

Natural Selection ∇ 3R SNP diversity?

- Ka/Ks

The ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks).

- $Ka/Ks = 1$ in a neutral model of evolution
- $Ka/Ks > 1$ in a positive selection model of evolution
- $Ka/Ks < 1$ in a negative model of evolution

Natural Selection ∇ 3R SNP diversity?

- 6 genes showed values of K_a/K_s considerably higher than 1 (*recB*, *dnaQ*, *uvrC*, *dinF*, *alkA*, *ogt*, *ligD*)



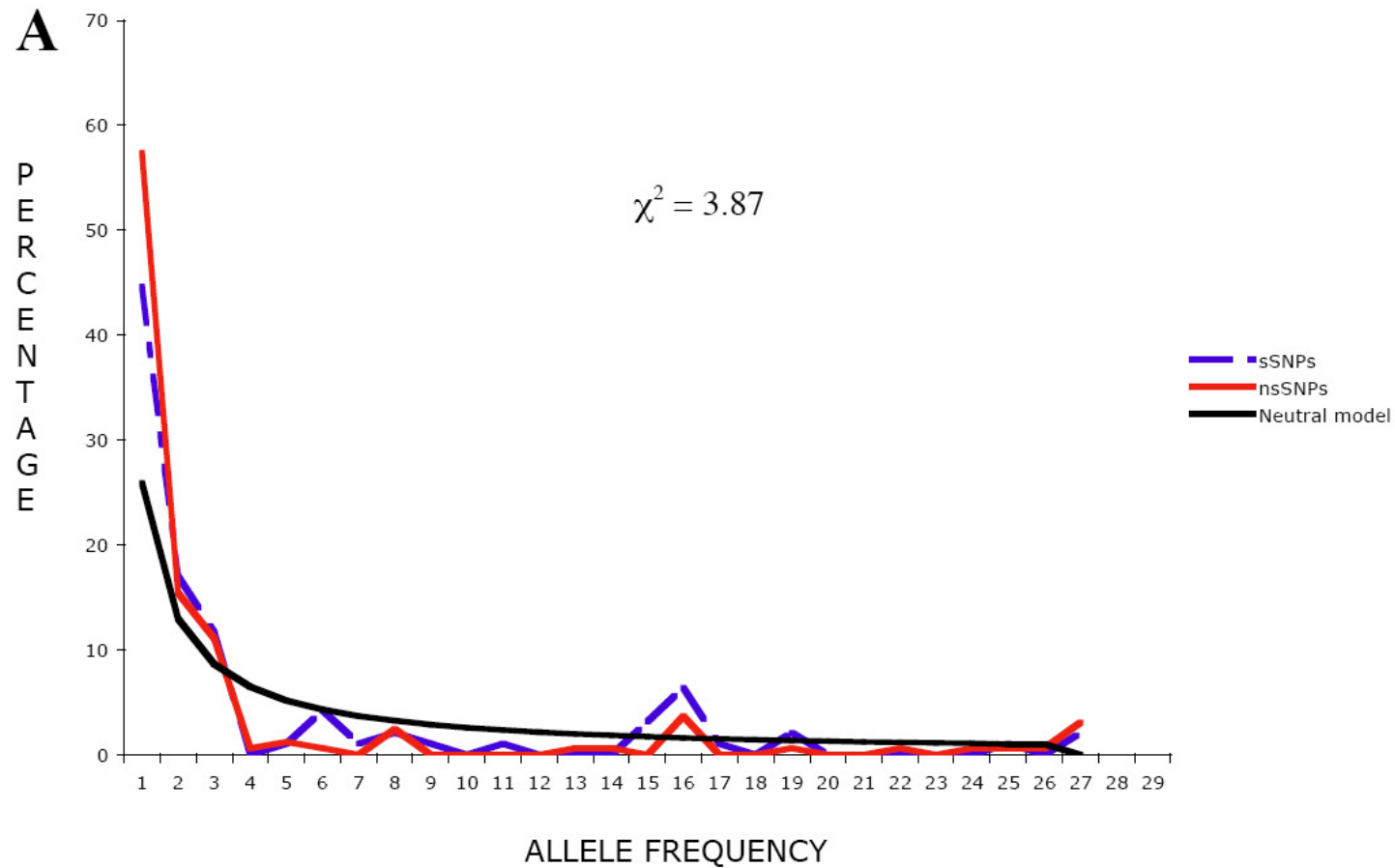
Suggestive of positive selection

- The majority showed values of K_a/K_s considerably lower than 1



**in agreement with the negative/purifying model
suggested by the SFS spectrum**

Natural Selection ∇ 3R SNP diversity?

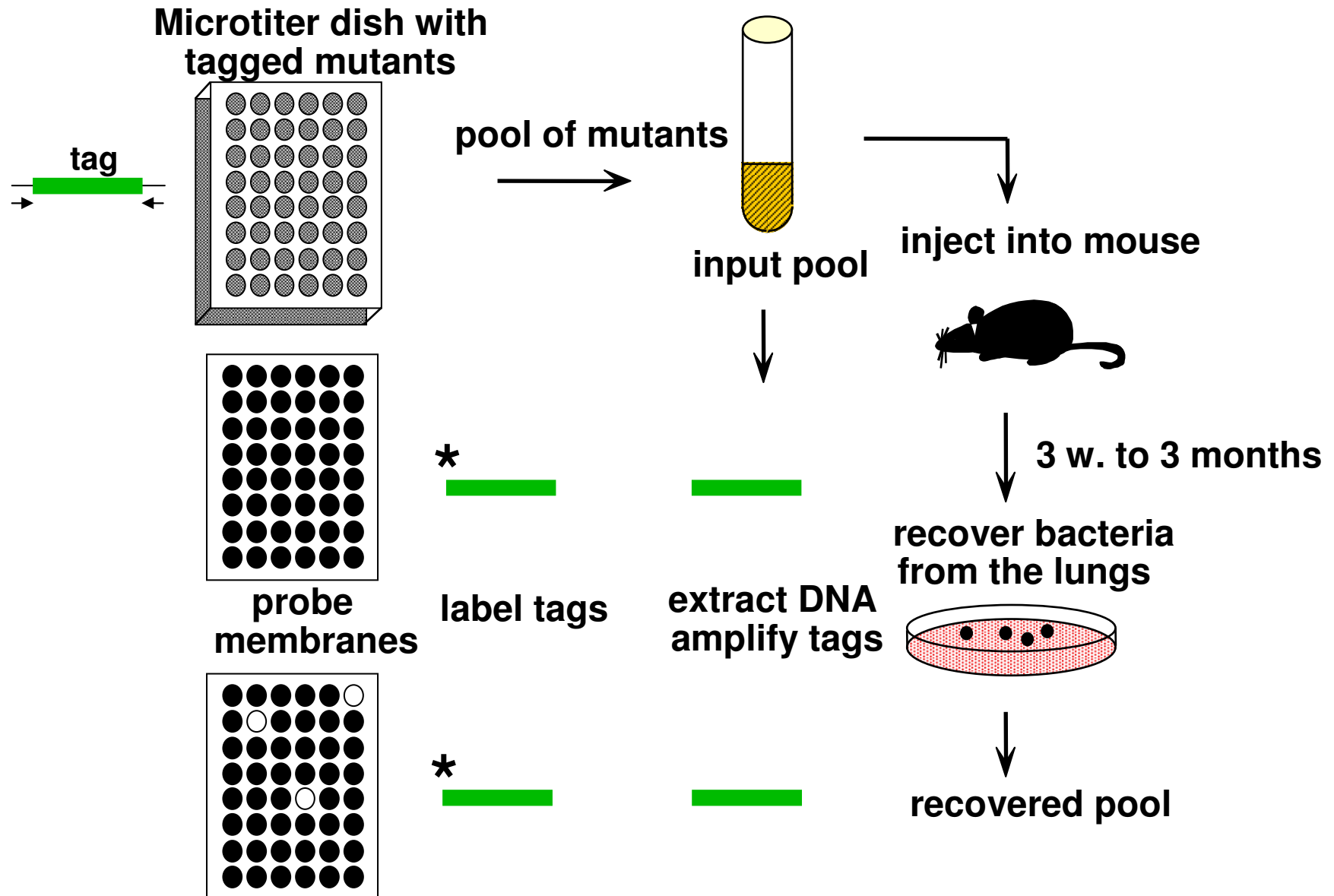


- In a neutral model sSNP and ns SNP patterns should be similar
- Our results are indicative of purifying/negative selection

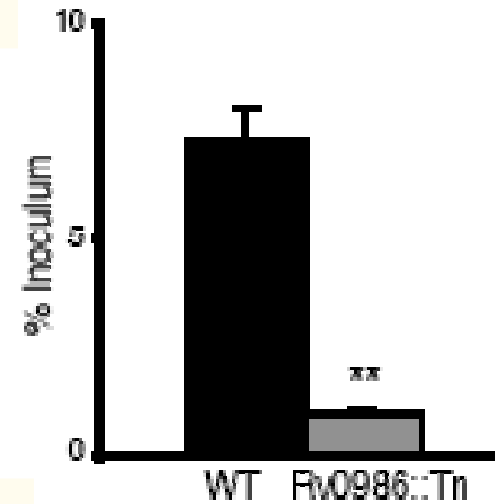
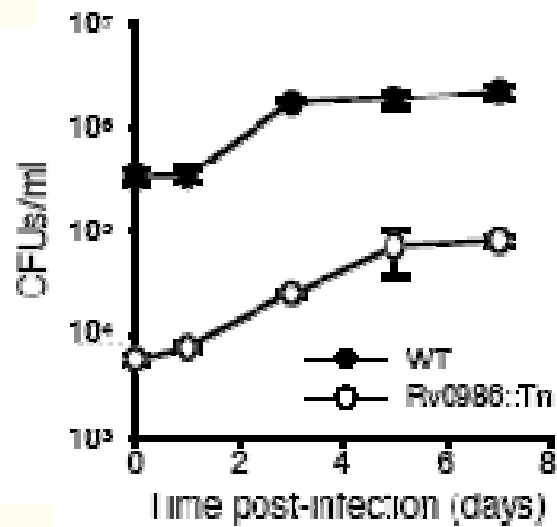
Conclusions

- *Mycobacterium tuberculosis* is more diverse than initially thought
- 3R SNP analysis seems to be a very robust analysis tool that might open new doors even at the level of localized epidemics
- It seems clear that 3R genes have and had an important role in the evolution of MTC strains
- The observation of changes in the observed patterns of equilibrium may be a sign of warning for strains with adaptive advantages

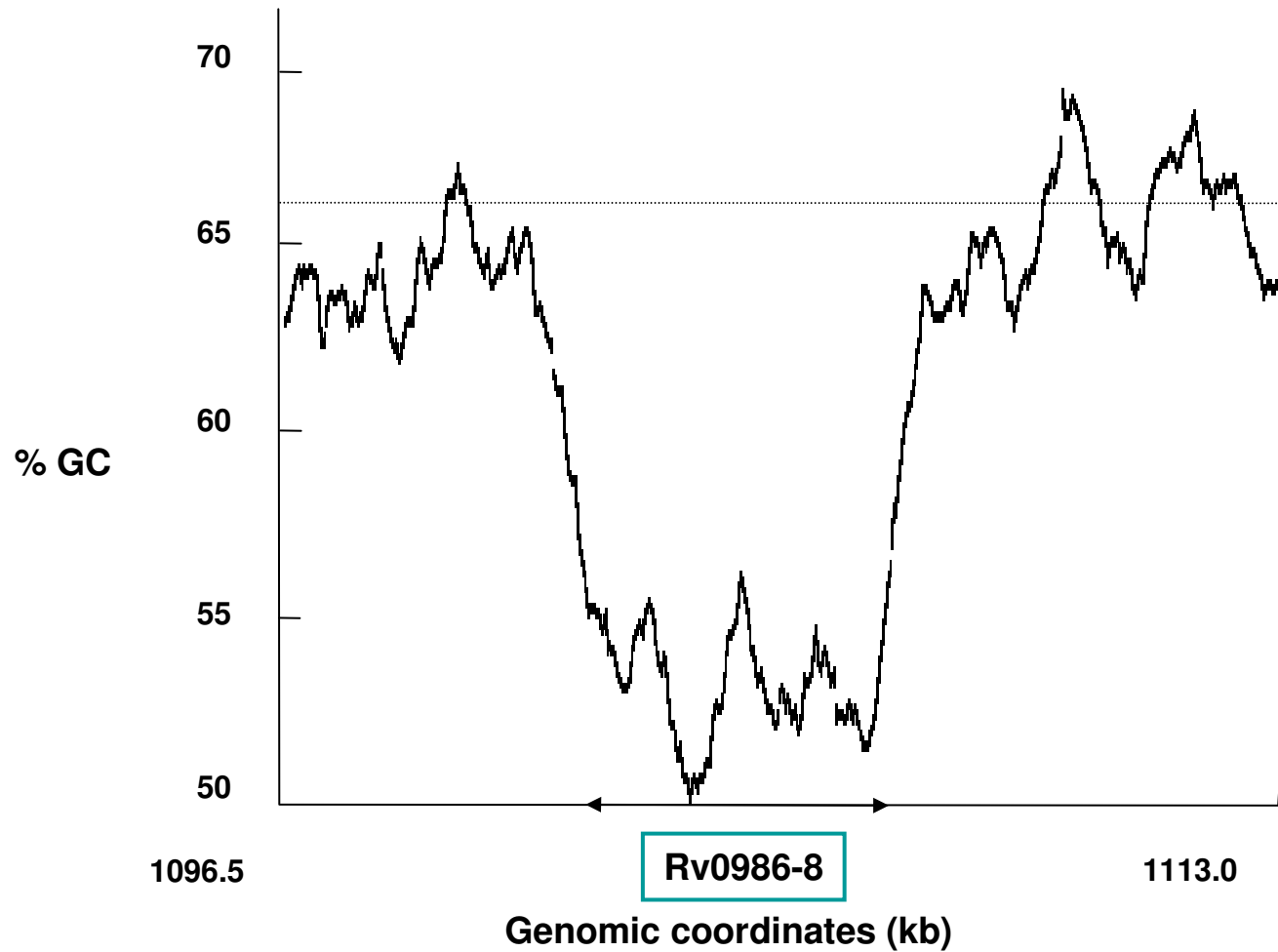
Identification of *M. tuberculosis* virulence genes by STM



Survival and binding in THP-1 cells

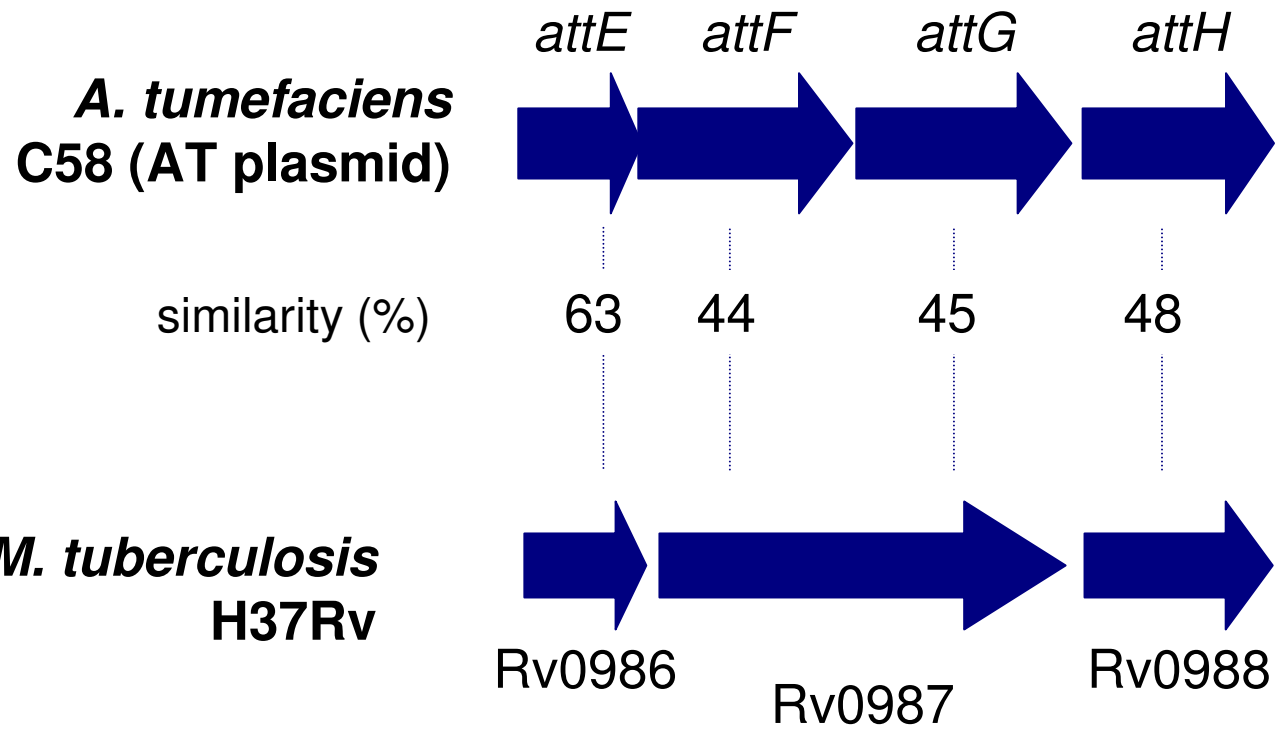


Less binding of Rv0986 mutant strain



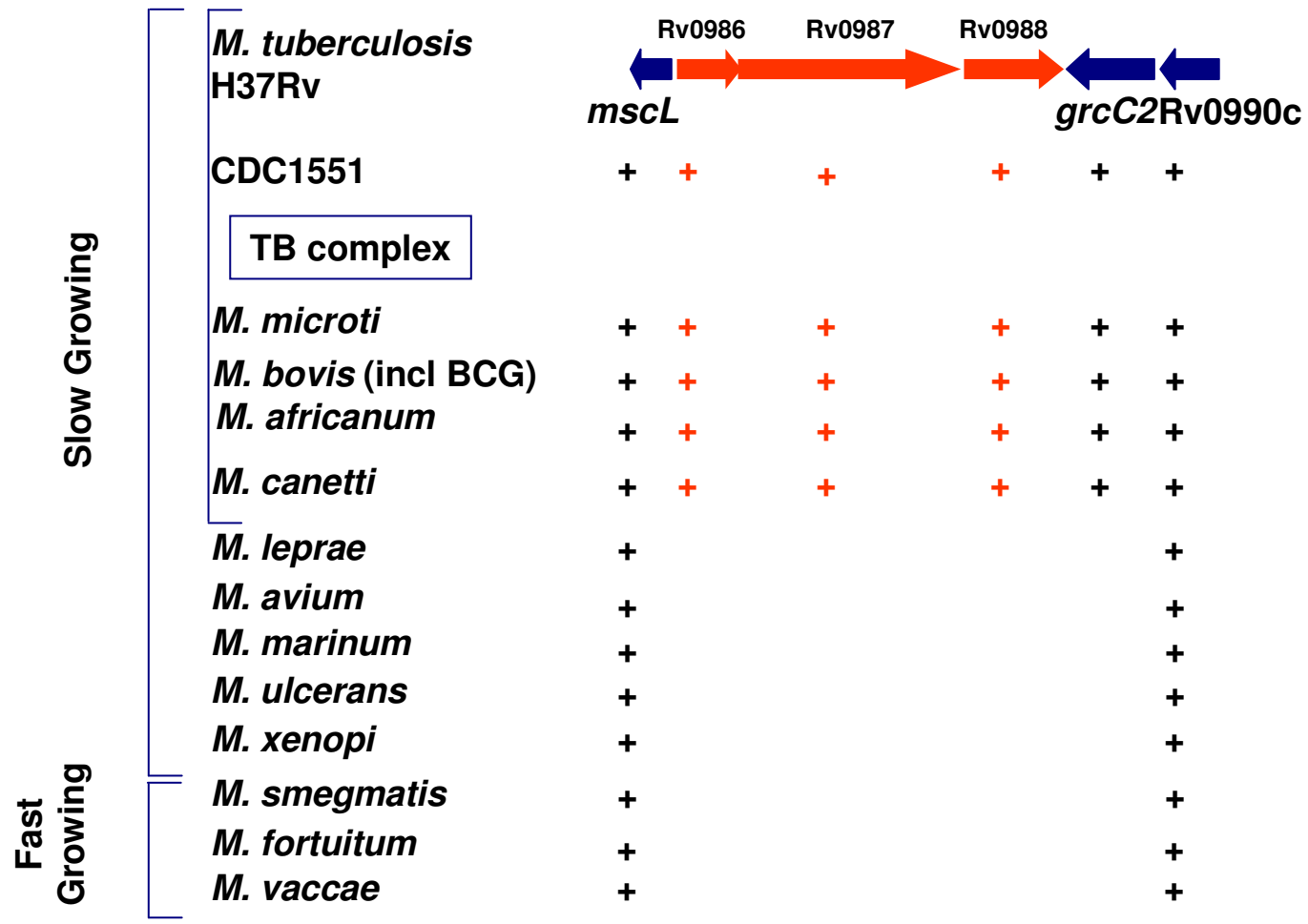
Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, Neyrolles O.
Mol Biol Evol. 2006. 23:1129-35.

Rosas-Magallanes V, Stadthagen-Gomez G, Rauzier J, Barreiro LB, Tailleux L, Boudou F, Griffin R, Nigou J, Jackson M, Gicquel B, Neyrolles O.
Infect Immun. 2007. 75:504-7.



In *Agrobacterium tumefaciens* *attE-H* genes form an operon thought to encode an ABC transporter involved in the secretion of a host cell adherence factor

Rv0986 was isolated by Pethe *et al.* (PNAS, 2004) in a screening to find mutants defective in the arrest of phagosome maturation

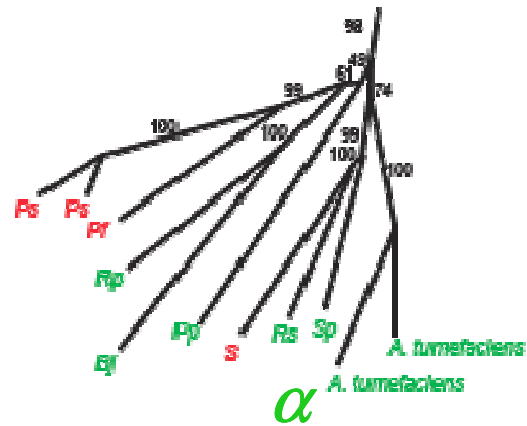
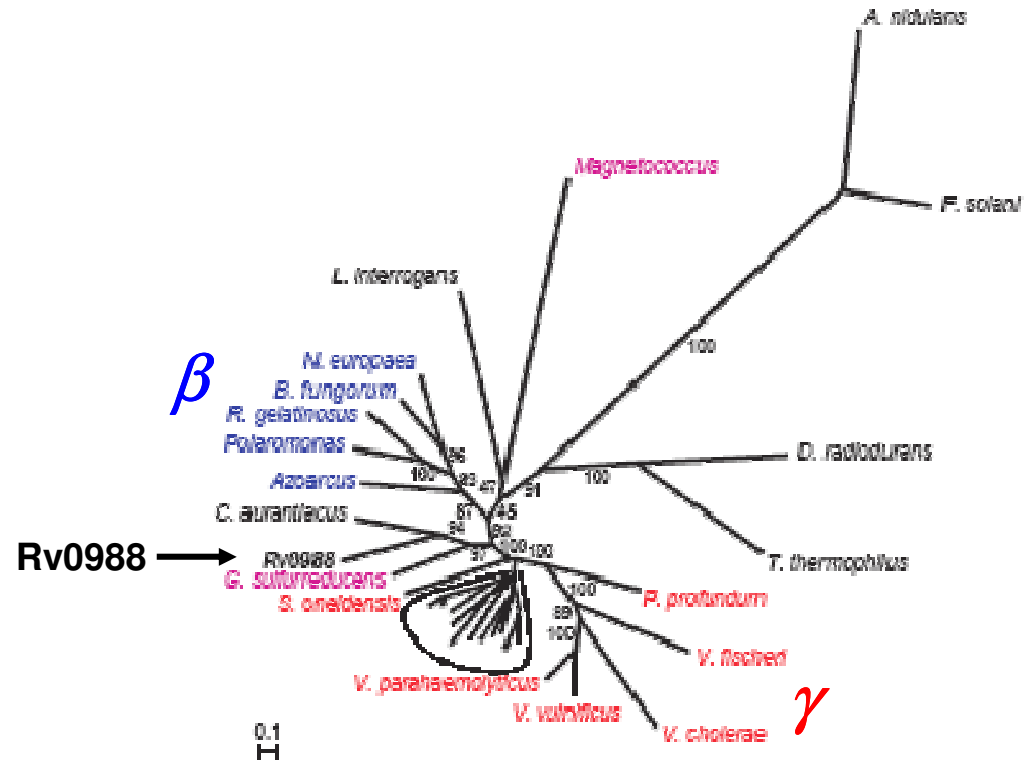


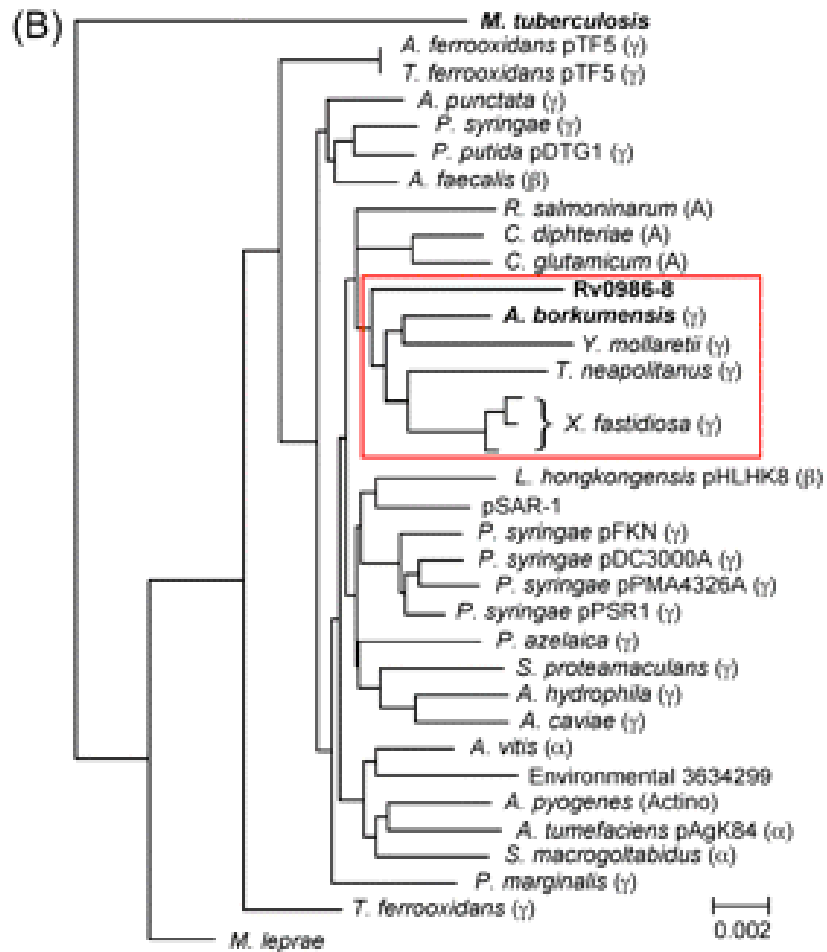
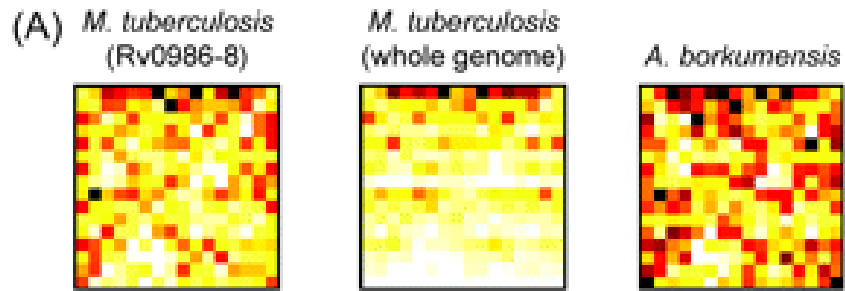
Rv0986-88 is present in all *M. prototuberculosis* strains

**Traces of insertion elements (incl. transposases)
upstream Rv0986 and downstream Rv0988 in *M.*
prototuberculosis D, E, F and G only**

**Low Ka/Ks (0.2 in average) which means a selective
constraint to prevent aminoacid change**

Rv0988





Functions of genes included in genomic islands

Gene Functional Category	No. of genes	% of genomic islands	Average % in complete genome
IS elements, repeated sequences & phages	40	15.6	3.7
Central intermediary metabolims	28	10.9	27.8
Virulence genes	21	8.2	6.5
Regulatory functions	13	5.1	10.3
Cell envelope & possible membrane proteins	12	4.7	17.5
Others	10	3.9	NA
<i>Unknown</i>	132	51.6	48.0
Total	256		

Potential origin of genomic islands

Grou	Genomic 1sland	Closest genera 1	Closest genera 2	Closest genera 3
P 1	Rv2336-Rv2339	<i>Burkholderia</i>	<i>Sphingomonas</i>	<i>Rhizobium</i>
	Rv2491-Rv2494	<i>Agrobacterium</i>	<i>Pseudomonas</i>	<i>Xanthobacter</i>
	Rv2734-Rv2735c	<i>Bradyrhizobium</i>	<i>Alicyclobacillus</i>	<i>Sphingomonas</i>
	Rv3108-Rv3126c	<i>Burkholderia</i>	<i>Corynebacterium</i>	<i>Pseudomonas</i>
	Rv3376-Rv3378c	<i>Agrobacterium</i>	<i>Rhizobium</i>	
	Rv0112-Rv2691	<i>Mycobacterium</i>	<i>Burkholderia</i>	<i>Pseudomonas</i>
2	Rv0986-Rv0989c	<i>Mycobacterium</i>	<i>Escherichia</i>	<i>Pseudomonas</i>
	Rv2954c-Rv2961	<i>Mycobacterium</i>	<i>Mesorhizobium</i>	<i>Pseudomonas</i>
	Rv0298-Rv303	<i>Mycobacterium</i>	Mycobacterial phage	
3	Rv0347-Rv0349	<i>Rhodococcus</i>	<i>Mycobacterium</i>	Bacteriophage
	Rv0836c-Rv0837c	<i>Mycobacterium</i>	<i>Rhodococcus</i>	Mycobacterial phage
	Rv1184c	<i>Mycobacterium</i>	<i>Rhodococcus</i>	Mycobacterial phage
	Rv2267c-Rv2269c	<i>Mycobacterium</i>	Mycobacterial phage	<i>Rhodococcus</i>
	Rv0595c-Rv0614	<i>Nocardia</i>	<i>Mycobacterium</i>	
4	Rv0628c	<i>Actinoplanes</i>	<i>Nocardia</i>	
	Rv1041c-Rv1055	<i>Mycobacterium</i>	<i>Gordonia</i>	
	Rv2159c-Rv2161c	<i>Nocardia</i>	<i>Amycolatopsis</i>	<i>Micromonospora</i>
	Rv2801c-Rv2824c	<i>Mycobacterium</i>	<i>Rhodococcus</i>	

ACKNOWLEDGMENTS

INSTITUT PASTEUR, PARIS

Olivier NEYROLLES
Mina Ebrahimi RAD
Vania ROSAS-MAGALLANES
Tiago DOS VULTOS
Olga MESTRE
Jean RAUZIER
Brigitte GICQUEL

Olivier SCHWARTZ

Roland BROSCH

Luis BARREIRO
Lluis QUINTANA-MURCI

INSERM, PARIS

Jennifer BECQ
Patrick DESCHAVANNE

University of Oslo

Tone TONJUM

University of Saragoza

Carlos MARTIN

Hôpital Necker, Paris

Ivan MATIC

Hospital Ramon y Cajal, Madrid

Jesus BLAZQUEZ