

"A re-evaluation of M. prototuberculosis": continuing the debate.

Sylvain Brisse, Philip Supply, Roland Brosch, Veronique Vincent, M Cristina Gutierrez

▶ **To cite this version:**

Sylvain Brisse, Philip Supply, Roland Brosch, Veronique Vincent, M Cristina Gutierrez. "A re-evaluation of M. prototuberculosis": continuing the debate.. PLoS Pathogens, 2006, 2 (9), pp.e95. 10.1371/journal.ppat.0020095 . pasteur-00146923

HAL Id: pasteur-00146923
https://pasteur.hal.science/pasteur-00146923

Submitted on 15 May 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OPEN ACCESS Freely available online

PLoS PATHOGENS

Opinions

"A Re-Evaluation of *M. prototuberculosis*": Continuing the Debate

Sylvain Brisse, Philip Supply, Roland Brosch, Veronique Vincent, M. Cristina Gutierrez*

In his Opinion [1], Noel Smith questions that the *Mycobacterium tuberculosis* complex (MTBC) is a successful clone that emerged from a much broader and ancient mycobacterial species, whose extant representatives are *Mycobacterium canettii* and other smooth tubercle bacilli [2]. He suggests that our conclusion is weak because he sees it as solely based on comparison of single nucleotide polymorphisms (SNPs) in six housekeeping genes and the 16 sRNA gene, and because some of these genes also showed evidence of horizontal transfer. Smith's criticism mostly comes down to two important assertions: first, MTBC members did not emerge from the smooth strains; second, the smooth strains and MTBC members should not be seen as a single species.

In response, we would stress that using comparison of SNP in housekeeping genes is now a widely accepted practice that has shown the clonal expansion of several other important bacterial pathogens from genetically broader progenitor species. Examples of such analyses are *Yersinia pestis* descended from *Yersinia pseudotuberculosis* [3], serotype Typhi from *Salmonella enterica* [4], *Shigella* clones from *Escherichia coli* [5], or *Bordetella pertussis* from *Bordetella bronchiseptica* [6]. For *M. tuberculosis,* Musser and colleagues [7] used SNP analysis of six housekeeping genes to detect the extremely reduced genetic polymorphisms that exist within the classical members of the MTBC. Most importantly, their results were later confirmed by whole genome comparisons [8–10]. It was therefore particularly exciting to find a much higher rate of sequence polymorphisms in the same six housekeeping genes that were used by Musser and colleagues when we studied smooth tubercle bacilli isolated in East Africa from patients with clinical tuberculosis, and these data represented the backbone of our published analysis [2]. However, we re-emphasize here that the greater genetic polymorphism observed among the smooth tubercle bacilli is not restricted to the six tested housekeeping genes, but also applies to other regions of the genomes, including conventionally used markers for differentiation of tubercle bacilli. These results, shown in supporting Figure S1 of our article [2], demonstrate that the smooth tubercle bacilli exhibited several outstanding features that underline their differences among each other and with the MTBC members. The number of IS*1081* copies constitutes one example. This insertion sequence has been so far described as being specific for tubercle bacilli [11]. Their number is highly conserved among all members of the MTBC, which carry five or six copies [11]. However, the smooth tubercle bacilli showed zero to three copies [2], indicating that this specific IS element was present in most strains of the smooth tubercle bacilli, but with a distinct copy number. In previous studies, the unusually low copy number of IS*1081* was also used to differentiate the smooth tubercle bacillus *M. canettii* from the classical MTBC members [12,13]. Other examples illustrating the wider diversity of smooth tubercle bacilli, mentioned in Protocol S2 and Figure S1, include the variable presence of another insertion sequence specific to smooth bacilli (IS*MycA1*), different MIRU–VNTR variant alleles also uniquely detected in smooth bacilli, and the absence of the direct repeat region in some smooth groups. In a different study, it was also shown that *M. canettii* harbored a completely different set of spacer sequences in the direct repeat region that was not present in the classical MTBC members [14]. Finally, Figure 1 shows our unpublished sequences of the 5′ end of the 16S rRNA gene in different lineages of the smooth tubercle bacilli. This region, which is 100% identical within the classical members of the MTBC, shows SNP variations that were only found within the smooth tubercle bacilli. This comparison also nicely shows that some lineages of smooth tubercle bacilli differ from others in an otherwise highly conserved region. Thus, altogether, our data clearly show that smooth tubercle bacilli harbor more diversity than MTBC members alone, although they share many characteristics in common (see below). More insight in this matter will certainly be provided by genome sequencing projects on *M. canettii* (http://www.sanger.ac.uk/sequencing/Mycobacterium/canetti) and other lineages of smooth tubercle bacilli.

In the case of the smooth tubercle bacilli strains, we demonstrated gene mosaicism, which we interpret as the result of homologous DNA recombination among smooth bacilli strains. Taking a different view, Smith speculates that the sequence diversity of the smooth strains, and their divergence from the MTBC members, mostly results from horizontal genetic transfer (HGT) into the smooth strains of DNA segments from donor strains of another species that remains to be discovered. Under this hypothesis, MTBC members would have been left untouched by HGT, whereas

Editor: Marianne Manchester, Scripps Research Institute, United States of America

Citation: Brisse S, Supply P, Brosch R, Vincent V, Gutierrez MC (2006) "A re-evaluation of *M. prototuberculosis*": Continuing the debate. PLoS Pathog 2(9): e95. DOI: 10.1371/journal.ppat.0020095

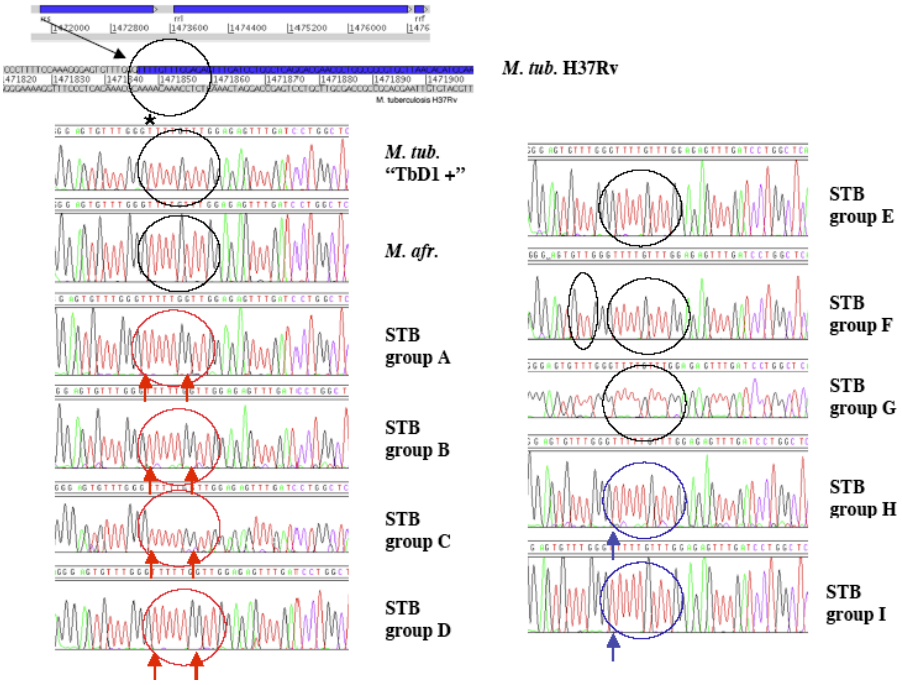
DOI: 10.1371/journal.ppat.0020095

Copyright: © 2006 Brisse et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: HGT, horizontal genetic transfer; MTBC, *Mycobacterium tuberculosis* complex; SNP, single nucleotide polymorphism

Sylvain Brisse and M. Cristina Gutierrez are at Unité de Biodiversité des Bactéries Pathogènes Emergentes, Institut Pasteur, Paris, France. Philip Supply is at INSERM U629, Institut Pasteur de Lille, Lille, France. Roland Brosch is at Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, Paris, France. Veronique Vincent is at the Stop TB Department, World Health Organization, Geneva, Switzerland.

* To whom correspondence should be addressed. E-mail: crisgupe@pasteur.fr



DOI: 10.1371/journal.ppat.0020095.g001

Figure 1. Polymophisms at the 5′ End of the 16S rRNA Gene in *M. tuberculosis* Complex Members and Smooth Tubercle Bacilli

Four alleles are shown by colored circles. The allele of smooth groups H and I (AM283533) consists of a T insertion one nucleotide upstream of the 16S rRNA gene, shown by an arrow; the allele of smooth groups A, B, C, and D (AM283530) consists of the same T insertion and a G to T mutation at nucleotide 6 of the gene (corresponding to the mutation taken into account for the dendrogram shown in Figure 1 in our original article), both shown by an arrow; the allele of smooth group F (AM283532) consists of a T deletion (circled) further upstream of the gene. The allele of smooth groups E and G are identical at the 5′ locus to that of MTBC members (AM283534), but smooth group E (AM283531) has one T to C polymorphism at nucleotide 1017 of the 16S rRNA gene (not shown). The first nucleotide of the 16S rRNA gene is indicated by a star, above the *M. tuberculosis* TbD1+ sequence. M. afr., *Mycobacterium africanum*; M. tub, *M. tuberculosis*; STB, smooth tubercle bacilli.

smooth strains would have been targeted with high frequency by HGT. Obviously, microbiologists are far from having discovered even a tiny fraction of existing biodiversity, and the existence of yet undiscovered mycobacterial species is almost certain. Noel Smith's hypothesis can therefore not be ruled out on the basis that the donor strains have not been discovered. Acquisition of some gene clusters through interspecies HGT has been recently well-described for mycobacteria [15–17], but such interspecies events can be readily recognized by specific genetic signatures. In our case, the levels of nucleotide identity of the recombinant segments in the smooth bacilli with the corresponding MTBC sequences is so high that the hypothetical donor of these recombinant segments would need to be much more closely related to the MTBC members than to any other mycobacterial species known so far (Table 1). Actually, our colleague's hypothesis would require the existence of a donor group that is as closely related (1% to 2% nucleotide divergence in housekeeping genes) to the MTBC members as are strains within single bacterial species [18]. With such a high degree of sequence identity, we think that it is more parsimonious to invoke homologous recombination within a single tubercle bacilli species, and do not see the need to invoke HGT from a hypothetical donor group of phylogenetically very closely related mycobacteria. In addition, our data on insertion sequences, direct repeat region spacers, and genomic deletions would not easily fit in with Smith's hypothesis.

As reflected in his Opinion [1], Noel Smith does not accept the concept of *M. prototuberculosis*. Of course, the bacterial species concept and rules for species definition are subject to debate. To tackle these questions, recommendations have been issued by several international groups [19–21]. Although imperfect, one proposed species definition is pragmatic, operational, and universally applicable: "A species is a category that circumscribes a genomically coherent group of individual isolates/strains sharing a high degree of similarity in independent features (*note*: including phenotypic and clinical features)". The use of nucleotide sequence divergence to define bacterial species is considered as one of the most promising approaches [19,20], and, when taken jointly with phenotypic and clinical parameters, should not necessarily be seen as "arbitrary and without merit." In addition to common features in genotypical markers (see above), the smooth strains and MTBC were grouped into a compact phylogenetic clade remote from the closest mycobacterial species when we analyzed their phylogenetic relatedness with other mycobacterial species defined so far. The overall level of housekeeping gene nucleotide sequence similarity in this clade is as high as 98.46%, a level higher than observed in many current bacterial species. Moreover, smooth tubercle bacilli cause a disease whose clinical and radiological presentation is indistinguishable from tuberculosis caused by classical *M. tuberculosis*, but clearly distinct from diseases caused by other mycobacterial species. Because of this genetic and clinical coherence, we have proposed to give the name *M.*

Table 1. Level of Nucleotide Identities of Putative Recombinant Fragments of Smooth Tubercle Bacilli with the Corresponding Sequences of *M. tuberculosis* H37Rv, and with the Closest Mycobacterial Species (Excluding MTBC Members)

Gene Segment	Smooth Group(s)	Similarity to H37Rv	Closest Other Mycobacterial Species (Similarity)
gyrB 648-1324	Groups F/G/H	99%	*M. kansasii* (87%)
gyrB 870-1449	Group I	99%	*M. malmoense* (88%)
gyrA 462–918	Groups C/D/E/H/I	98%	*M. avium subsp. paratuberculosis* (89%)
katG 1388–1860	Groups E/F/G/H	99%	*M. leprae* (84%)

DOI 10.1371/journal.ppat.0020095.t001

prototuberculosis to the phylogenetic entity constituted by the smooth strains (including *M. canettii*) and the MTBC members. Indeed, most smooth strain groups do not conform to all the formal criteria that define MTBC members and/or *M. canettii,* and a new name was therefore needed. Nevertheless, we want to make clear that we do not propose to abandon the usage of the current species names of the MTBC, such as *M. tuberculosis* or *M. bovis*. Nomenclature and evolutionary biology of bacteria have distinct premises and purposes, and distinct designations can coexist as long as the context in which they are used is clear. We thus propose to consider that the biological species *M. prototuberculosis*—which has no standing in nomenclature—comprises strains that conform to a set of nomenspecies with standing in nomenclature, such as *M. tuberculosis*. In the same way, for example, *Shigella* species or *Y. pestis* names are still retained in the nomenclature although is it known that they belong to a single genomic species together with *E. coli* or *Y. pseudotuberculosis,* respectively.

Noel Smith underlines the uncertainty inherent in the use of molecular clocks as well as methodological simplifications in our analysis. If one would accept his view that recombinant segments come from unrelated donor strains, then age estimates would be nonsense. However, under our hypothesis, the nucleotide diversity was generated within the *M. prototuberculosis* species, within which it has subsequently been shuffled among lineages. We agree that the age estimate may be overestimated because the available polymorphisms are not fixed in the population, and we certainly are aware of molecular clock uncertainties. We ourselves mentioned that the age scenario was speculative and put it forward with the main aim of challenging the widely accepted paradigm that tuberculosis is a recent disease of humans. In this sense, one should remember that the infectious cycle of the tubercle bacillus consists of a balance virtually unique among bacterial pathogens, between often lifelong asymptomatic parasitism and the development of tuberculous cavities in the lungs of the patient, which allows the efficient spread of high numbers of bacteria via aerosol to new hosts. From the perspective of this obligate pathogen, this balance is absolutely essential to maintain its so-successful life cycle, and it is very reasonable to think that this adaptation has evolved during a long-lasting co-evolution between the bacterium and its host. Therefore, our scenario offers a novel working frame to solve the apparent paradox between this exquisite adaptation and the previously hypothesized recent evolutionary origin of the pathogen [22] .

In summary, we thank Noel Smith for his critical comments, which have triggered many stimulating discussions and this response. We hope that we have clarified the different lines of evidence supporting our finding, as we are confident that the greater genetic diversity within this fascinating group of smooth tubercle bacilli is a key to a better understanding of the evolution of the major human pathogen *M. tuberculosis*. ■

Supporting Information

Accession Numbers

The EMBL (http://www.ebi.ac.uk/embl) accession numbers for the representative sequences discussed in Figure 1 are AM283530-AM283534.

Acknowledgments

Author contributions. All authors wrote the paper.

Funding. The authors received no specific funding for this article.

Competing interests. The authors have declared that no competing interests exist.

References

1. Smith NH (2006) A re-evaluation of *M. prototuberculosis*. PLoS Pathog 2 (9): e98 DOI: 10.1371/journal.ppat.0020098
2. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, et al. (2005) Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog 1 (1):e5. DOI: 10.1371/journal.ppat.0010005
3. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, et al. (1999) *Yersinia pestis,* the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 96: 14043–14048.
4. Kidgell C, Reichard U, Wain J, Linz B, Torpdahl M, et al. (2002) *Salmonella typhi,* the causative agent of typhoid fever, is approximately 50,000 years old. Infect Genet Evol 2: 39–45.
5. Pupo GM, Lan R, Reeves PR (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A 97: 10567–10572.
6. Diavatopoulos DA, Cummings CA, Schouls LM, Brinig MM, Relman DA, et al. (2005) *Bordetella pertussis,* the causative agent of whooping cough, evolved from a distinct, human-associated lineage of *B. bronchiseptica*. PLoS Pathog 1 (4):e45. DOI: 10.1371/journal.ppat.0010045
7. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94: 9869–9874.
8. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393: 537–544.
9. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, et al. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol 184: 5479–5490.
10. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, et al. (2003) The complete genome sequence of *Mycobacterium bovis*. Proc Natl Acad Sci U S A 100: 7877–7882.
11. van Soolingen D, Hermans PW, de Haas PE, van Embden JD (1992) Insertion element IS*1081*–associated restriction fragment length polymorphisms in *Mycobacterium tuberculosis* complex species: A reliable tool for recognizing *Mycobacterium bovis* BCG. J Clin Microbiol: 1772–1777.
12. van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, et al. (1997) A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: Characterization of an exceptional isolate from Africa. Int J Syst Bacteriol 47: 1236–1245.
13. Pfyffer GE, Auckenthaler R, van Embden JD, van Soolingen D (1998)

Mycobacterium canettii, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. Emerg Infect Dis 4: 631–634.
14. van Embden JD, van Gorkom T, Kremer K, Jansen R, van Der Zeijst BA, et al. (2000) Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J Bacteriol 182: 2393–2401.
15. Kinsella RJ, Fitzpatrick DA, Creevey CJ, McInerney JO (2003) Fatty acid biosynthesis in *Mycobacterium tuberculosis*: Lateral gene transfer, adaptive evolution, and gene duplication. Proc Natl Acad Sci U S A 100: 10320–10325.
16. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, et al. (2006) Horizontal transfer of a virulence operon to the ancestor of *Mycobacterium tuberculosis*. Mol Biol Evol 23: 1129–1135.
17. Krzywinska E, Krzywinski J, Schorey JS (2004) Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. Microbiology 150: 1707–1712.
18. Spratt BG, Hanage WP, Feil EJ (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr Opin Microbiol 4: 602–606.
19. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, et al. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52: 1043–1047.
20. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, et al. (2005) Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol 3: 733–739.
21. Wayne LG, Brenner DJ, Colwell RR, et al. (1987) International Committee on Systematic Bacteriology. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol 37: 463–464.
22. Parkhill J (2005) Faculty of 1000 biology. Available: http://www.f1000biology.com/article/16201017/evaluation. Accessed 22 August 2006.

