Identification and Characterization of Genomic Variations between *Mycobacterium bovis* and *M. tuberculosis* H37Rv

Ariel Amadio,¹ María Isabel Romano,² Fabiana Bigi,² Ignacio Etchechoury,² Tanja Kubica,³
Stefan Niemann,³ Angel Cataldi,^{2*} and Karina Caimi²

Institute of Microbiology and Agricultural Zoology¹ and Institute of Biotechnology,² CICVyA-INTA, Castelar, Argentina, and Forschungszentrum Borstel, National Reference Center for Mycobacteria, D-23845 Borstel, Germany³

Received 18 June 2004/Returned for modification 9 August 2004/Accepted 17 January 2005

Genetic differences between *Mycobacterium bovis* and *M. tuberculosis* were identified. We found (i) a deletion of Rv3479 specific to *M. bovis*, (ii) that the *rpfA* gene is shortened to various extents in *M. bovis*, and (iii) an insertion in Rv0648 and a duplication of *lppA* common in *M. tuberculosis* complex isolates.

Mycobacterium bovis, the agent of bovine tuberculosis, belongs to the *M. tuberculosis* complex (MTBC), whose members are closely related and show high similarity at the DNA level (2, 5, 8). However, despite this close relationship, they show a large variability in their phenotypic properties, epidemiology, and incidence in human tuberculosis.

The *M. tuberculosis* H37Rv (4) and *M. bovis* AF2122/97 (7) genomic sequences were compared using blastn and MSPcrunch (11) software and visualized with Artemis Comparison Tool version 4.0 for Apple (ACT, www.sanger.ac.uk/Software).

The criteria used to select loci for further studies were that they were not IS6110 insertions, PE-PPE family genes (4), single-nucleotide polymorphisms, or previously published regions of difference (1, 2). Four loci were selected: Rv3479, Rv0648, *lppA* and *rpfA*. In all these cases, the polymorphisms involved only one gene. The bioinformatic analysis yielded the following information: <OL TYPE = "a">.

(i) Rv3479. The Rv3479 gene encodes a hypothetical transmembrane protein of 3,225 bp in both *M. tuberculosis* H37Rv and CDC 1551 (6). However, the orthologue gene in *M. bovis* AF2122/97 carries a 714-bp deletion which starts at base 1254 from the ATG start codon (Fig. 1). This deletion originates two smaller open reading frame with sequences almost identical to Rv3479.

(ii) *rpfA* (Rv0867). *rpfA* is a 1,221-bp gene that encodes for a possible resuscitation-promoting factor. The central part of the protein contains APADLAPP repeats (9). The sequenced *M. bovis* strain lacks part of these repeats (Fig. 1) due to a 240-bp in-frame deletion. These domains are not deleted in *M. tuberculosis* CDC 1551.

(iii) Rv0648. *M. bovis* AF2122/97 has a 533-bp in-frame insertion in the Rv0648 gene relative to *M. tuberculosis* H37Rv (Fig. 1). This characteristic is similar to that of the RvD elements (3) (<http://www.pasteur.fr/recherche/unites/Lgmb->

/Deletion.html). However, the sequence of this insertion indicates that it is not one of the five previously described RvD (3) or TbD1 (2) elements. We propose to refer to this fragment as RvD6. This extra fragment is also present in *M. tuberculosis* CDC 1551.

(iv) *lppA* and *lppB*. The *lppA* and *lppB* genes, encoding lipoproteins, are highly similar. *lppA* is duplicated in *M. bovis* AF2122/97 with respect to *M. tuberculosis* H37Rv (Fig. 1). Interestingly, *lppA* is also duplicated in *M. tuberculosis* CDC 1551.

The four genes were BLAST searched in the *M. smegmatis*, *M. avium*, *M. marinum*, and *M. leprae* genomes (data not shown). Orthologues of *rpfA* genes exist in all of them. However, no Rv3479 and *lppA* orthologues were found in these mycobacteria. The orthologues of Rv0648 present in *M. smegmatis*, *M. avium*, and *M. marinum* have a size similar to that of *M. bovis*, suggesting that *M. tuberculosis* H37Rv suffered a deletion in this gene.

PCRs were applied to *M. tuberculosis* (20 isolates from Argentina and 10 from The Netherlands), *M. bovis* (15 isolates from Argentina and 15 from Mexico), *M. pinnipedii* (2 isolates), and the reference strains *M. tuberculosis* H37Rv, *M. bovis* AN5, and *M. bovis* BCG (five substrains, from the collection of the Institute of Biotechnology, CICVyA-INTA, Castelar, Argentina). *M. microti* strains (provided by Dick van Soolingen, Rijksinstituut voor Volksgezondheid en Milieu, The Netherlands) and *M. bovis* (five isolates) and *M. bovis* subsp. *caprae* (five isolates) from the National Reference Center for Mycobacteria, Borstel, Germany (10) strains were also used. These strains have been extensively characterized by IS6110 restriction fragment length polymorphism and spoligotyping. PCR amplifications were performed using conventional cycling conditions. The primers used are shown in Table 1.

PCR results. **(i) Rv3479.** The shorter band was observed in all the *M. bovis* (wild type and BCG) isolates tested. In all the other MTBC isolates tested, the larger PCR band was detected, including *M. bovis* subsp. *caprae* (Fig. 2). These results indicate that this element, as well as RD4 (2), is specifically deleted in classical *M. bovis* strains, suggesting that the deletion

* Corresponding author. Mailing address: Instituto de Biología, CICVyA/INTA, Los Reseros y Las Cabañas, 1712 Castelar, Argentina. Phone: 54-11-4621-1447. Fax: 54-11-4621-0199. E-mail: acataldi@cnia.inta.gov.ar.

FIG. 1. Schematic representation of gene arrangements at the four loci where differences between *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 were found. Genes are indicated as arrowhead boxes, and sizes of genes are shown. Insertions and deletions of one species with respect to the other are represented as diagonal lines.

occurred after the divergence of *M. bovis* and *M. bovis* subsp. *caprae*.

(ii) **rpfA**. The *rpfA* gene is polymorphic in *M. bovis* because a larger PCR band was observed in *M. bovis* BCG, AN5, and 20/35 wild-type isolates, as well as in the other MTBC species tested. A smaller (6/35 isolates) or an intermediate (9/35) band was amplified in other *M. bovis* isolates (Fig. 2).

(iii) **Rv0648**. The 3,648-bp *Rv0648* gene has high similarity to several α -mannosidases. *M. tuberculosis* is polymorphic at this gene because four isolates produced a small amplification band, while the other six produced a larger band. The other MTBC species tested showed only the larger band.

(iv) **lppA and lppB**. The *lppA* duplication was found in *M. bovis* and in 4/10 *M. tuberculosis* isolates, as suggested by a higher amplification band compared to most *M. tuberculosis* isolates (Fig. 2). *M. bovis* BCG and AN5 have a nonduplicated *lppA* gene. In consequence, *lppA* is not a species-specific marker. Band sizes from *M. pinnipedii* and *M. microti* suggest that there is no *lppA* duplication.

rpfA amplification products were cloned in pGEM-T easy vector (Promega) and sequenced using Dye Terminator reactions. In *M. tuberculosis*, there are five perfectly conserved amino acid repeats (APADLAPP) and eight degenerated and shorter repeats. We observed that there is a progressive loss of internal repeats of *rpfA* in *M. bovis*. While the de-

scribed motif is conserved in *M. bovis* BCG and AN5, a fragment composed of a set of two linkers and four repeats was lost in some local strains (as in the sequenced *M. bovis* strain). Furthermore, two local *M. bovis* strains lack the other linker and six contiguous repeats (data not shown). While the expression of the *rpfA* gene was well demonstrated in *M. tuberculosis* and *M. bovis* BCG, we do not know whether the smaller-size gene variants from *M. bovis* strains are expressed. In any case, *rpfA* is a novel marker to differentiate and type *M. bovis* strains. As with *Rv3479*, the variable shortening of *rpfA* seems to be specific to classical *M. bovis*. This gene appears to perform an important role because it is conserved in several mycobacteria (data not shown).

TABLE 1. Primers used in this study

Primer	Sequence	Annealing temp (°C)
rpfArup2	GCCACCGTAAGCCCACCACA	58
RpfArlow	TCAGCCCGTCAGCCGATGAC	58
Up647-8alt	GTGGAACCCATCTGAGCAGT	53
Low 647-8alt	ATCACTTCGATACGGCCATC	53
Uplppa	CGAAAGCCCACAGCAGTGATCG	58
LowlppB	TGTCGACGCAGAACACCATCAC	57
3479up	GAACCGCCAAGAAGGGAGAGG	58
3479low	CGACGCCGAGCACAAGTAGCA	58

FIG. 2. Example of PCR assessment of the presence of the differences in several MTBC isolates. Primers flanking the differential regions were used for PCR. (*Rv3479* [primers directed to the *Rv3479* gene]) Lanes 1 to 10, *M. tuberculosis* isolates; lane 11, *M. microti*; lane 12, *M. bovis* BCG; lane 13, *M. bovis* AN5; lane 14, *M. pinnipedii*; lanes 15 to 25, *M. bovis* isolates; lanes 26 to 30, *M. bovis* subsp. *caprae* isolates; lane 31, molecular weight marker. (*rpfA*) Lanes 1 to 5, *M. bovis* subsp. *caprae* isolates; lanes 6 to 14, *M. bovis* isolates; lane 15, *M. bovis* AN5; lane 16, *M. microti*; lanes 17 and 18, *M. pinnipedii*; lanes 19 to 23, *M. bovis* BCG substrains; lanes 24 to 33, *M. tuberculosis* isolates; lane 34, molecular weight marker. (*Rv0648*) Lane 1, *M. microti*; lanes 2 and 3, *M. pinnipedii*; lanes 4 to 8, *M. bovis* subsp. *caprae* isolates; lanes 9 to 16, *M. bovis* isolates; lanes 17 to 26, *M. tuberculosis* isolates; lane 27, molecular weight marker. (*lppA*) Lane 1, *M. microti*; lanes 2 and 3, *M. pinnipedii*; lane 4, *M. bovis* BCG; lane 5, *M. bovis* AN5; lanes 6 to 11, *M. bovis* isolates; lanes 12 to 16, *M. bovis* subsp. *caprae* isolates; lanes 17 to 25, *M. tuberculosis* isolates; lane 26, molecular weight marker.

In conclusion, the differences between *M. bovis* and *M. tuberculosis* genes described here could be used as markers for the differentiation and species determination of MTBC isolates.

The present study was supported by grants from the Centro Argentino Brasileño de Biotecnología (CABBIO). M.I.R., F.B., A.A., and A.C. are fellows of the National Research Council of Argentina (CONICET). K.C. has a fellowship from Fundación Antorchas.

We are grateful to Haydee Gil and Valeria Rocha for technical help. We acknowledge The Institute for Genomic Research for the use of unpublished genomic sequences of *M. smegmatis* and *M. avium* and the Sanger Centre for *M. marinum*.

REFERENCES

1. Brosch, R., S. V. Gordon, A. Pym, K. Eiglmeier, T. Garnier, and S. T. Cole. 2000. Comparative genomics of the mycobacteria. *Int. J. Med. Microbiol.* **290**:143–152.
2. Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**:3684–3689.
3. Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon. 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect. Immun.* **67**:5768–5774.
4. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.
5. Cousins, D. V., R. Bastida, A. Cataldi, V. Quse, S. Dow, S. Redrobe, P. Duignan, A. Murray, C. Dupont, N. Ahmed, D. M. Collins, W. R. Butler, D. Dawson, D. Rodriguez, J. Loureiro, M. I. Romano, A. Alito, M. Zumárraga, and A. Bernardelli. 2003. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium tuberculosis* subsp. *pinnipedae* subsp. nov. *Int. J. Syst. Evol. Microbiol.* **53**:1305–1314.
6. Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser. 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* **184**:5479–5490.
7. Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson. 2003. The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. USA* **100**:7877–7882.
8. Kubica, T., S. Rüsche-Gerdes, and S. Niemann. 2003. *Mycobacterium bovis* subsp. *caprae* caused one-third of human *M. bovis*-associated tuberculosis cases reported in Germany between 1999 and 2001. *J. Clin. Microbiol.* **41**:3070–3077.
9. Mukamolova, G. V., O. A. Turapov, D. I. Young, A. S. Kaprelyants, D. B. Kell, and M. Young. 2002. A family of autocrine growth factors in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **46**:623–635.
10. Niemann, S., E. Richter, and S. Rüsche-Gerdes. 2002. Biochemical and genetic evidence for the transfer of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to the species *Mycobacterium bovis* Karlson and Lessel 1970 (approved lists 1980) as *Mycobacterium bovis* subsp. *caprae* comb. nov. *Int. J. Syst. Evol. Microbiol.* **52**:433–436.
11. Sonhammer, E. L. L., and R. Durbin. 1994. A workbench for large scale sequence homology analysis. *Comput. Appl. Biosci.* **10**:301–307.