A 12·7 kb fragment of the *Mycobacterium tuberculosis* genome is not present in *Mycobacterium bovis*

Martin Zumárraga,¹ Fabiana Bigi,¹ Alicia Alito,² María I. Romano¹ and Angel Cataldi¹

Author for correspondence: Angel Cataldi. Tel: +54 1 621 1447. Fax: +54 1 481 2975.
e-mail: acataldi@inta.gov.ar

Instituto de Biotecnología¹
and Instituto de
Patobiología², Centro de
Investigaciones en Ciencias
Veterinarias, Instituto
Nacional de Tecnología
Agropecuaria, 1708 Moron,
Argentina

Southern blotting, sequence analysis and PCR experiments showed that *Mycobacterium bovis* and *Mycobacterium bovis* BCG lack a 12·7 kb fragment present in the genome of *Mycobacterium tuberculosis*. This region is 337 bp downstream of the RD2 region, which was previously described as being absent from some *M. bovis* BCG strains. The 12·7 kb fragment should be useful as a target for a PCR test to differentiate *M. tuberculosis* and *M. bovis*. An analysis of the 12·7 kb region suggests that it represents a deletion in *M. bovis* rather than an insertion in *M. tuberculosis*. The deletion removes most of the *mce-3* operon, one of four highly related operons which may be involved in cell entry, and therefore it may contribute to differences in virulence or host range in the two species.

Keywords: *Mycobacterium tuberculosis*, *Mycobacterium bovis*, sequence analysis, genome, species differentiation

INTRODUCTION

The *Mycobacterium tuberculosis* complex consists of slow-growing, pathogenic mycobacteria that are difficult to differentiate by classical microbiological methods (Runyon *et al.*, 1974). The group includes *M. tuberculosis*, the agent of human tuberculosis; *Mycobacterium bovis*, which causes both bovine and human tuberculosis; *Mycobacterium africanum*, which causes human tuberculosis on the African continent; and *Mycobacterium microti*, which is pathogenic for the vole, a wild rodent. Also included in the *M. tuberculosis* complex are several unclassified isolates that were described as infecting wild seals in Australia and Argentina (Cousins *et al.*, 1990; Romano *et al.*, 1995). These isolates have biochemical, host range and genetic characteristics that fulfil some, but not all, of the properties by which each species of the *M. tuberculosis* complex is defined.

Precise differentiation at species level within the *M. tuberculosis* complex is a difficult task. Present methods are tedious and slow as they are based on cultural and biochemical properties, such as sensitivity to thioxi-phene-2-carboxylic acid (TCH) and the niacin test

(Heifets & Good, 1994). Furthermore, *M. bovis* differs from *M. tuberculosis* in having a low growth rate on egg media supplemented with glycerol, but a faster growth on egg media supplemented with pyruvate (Stonebrink media). As well as being laborious, biochemical tests can sometimes be misleading. For example, *M. bovis* isolates are resistant to pyrazinamide, an antimycobacterial drug, and whilst *M. tuberculosis* strains are generally considered pyrazinamide-sensitive, some resistant isolates have been found (Heifets & Good, 1994).

The current lack of precise methods for species identification within the *M. tuberculosis* complex has led different research groups to look for molecular methods to distinguish *M. bovis* from *M. tuberculosis*. (Cousins *et al.*, 1991; Del Portillo *et al.*, 1991; Rodriguez *et al.*, 1995; Scorpio *et al.*, 1997; Sreevatsan *et al.*, 1996). DNA-based methods for differentiation, such as PCR, have the advantage over classical methods because culture of the organism is not required, greatly speeding up the diagnostic process. The identification of genetic differences among members of the *M. tuberculosis* complex will also lead to a better understanding of virulence and host range differences displayed by the members of the complex.

In a previous study (Fisanotti *et al.*, 1997), we designed a pair of primers that amplified only *M. bovis* DNA and

The GenBank accession number for the nucleotide sequence determined in this work is AL022073.

not *M. tuberculosis* DNA; in consequence the amplified fragment was named MBSS (for *Mycobacterium bovis* specific sequence). However, this fragment hybridized to DNA of both species, showing a strong polymorphism between *M. bovis* and *M. tuberculosis*. This region is downstream of the RD2 region of the *M. bovis* genome identified by Mahairas *et al.* (1996). Following the complete sequencing of the *M. tuberculosis* genome (Cole *et al.*, 1998), we were able to analyse this region and we identified a novel and important genetic difference between *M. tuberculosis* and *M. bovis*.

METHODS

Bacterial strains. *M. tuberculosis* H37Rv and *M. bovis* AN5 were used as reference strains. *M. tuberculosis* isolates were obtained from human patients from Argentina, Brazil and Venezuela. *M. bovis* isolates from cattle were obtained from animal health services from Argentina, Brazil and Paraguay. *M. bovis* isolates from humans were obtained from a tuberculosis reference service in Argentina. *M. microti* and *M. africanum* were obtained from D. Van Soolingen (RIVM, Bilthoven, The Netherlands). *M. bovis* and *M. tuberculosis* strains were grown on Stonebrink and Lowenstein-Jensen slants (Difco), respectively.

Computer analysis. DNA alignments and protein structure predictions were performed with MegAlign and Protean software (DNASTAR), respectively. DNA and protein homology searches were performed by e-mail using the FASTA and Blitz services from EMBL. Prosites signature scanning was performed using the Web service at <http://expasy.hcuge.ch/sprot/prosite.html>.

Southern blots. These were performed as previously described (Fisanotti *et al.*, 1997).

DNA amplification. The amplification conditions were: an initial denaturation step at 94 °C for 4 min; followed by 30 cycles of 94 °C for 30 s for denaturation, different annealing temperatures for 1 min, and 72 °C for 2 min for extension.

Four pairs of primers were used (Fig. 1). One of them was composed of primer 1Umbss (ATCTACTTGCTCACCC-TAACG), which anneals to the region common to both *M. bovis* and *M. tuberculosis* located upstream of the *M. tuberculosis* 12·7 kb fragment, and primer 3L12505 (CTGT-

GCTGCGGGCTGCCG), which anneals to the 12·7 kb fragment near its 5' end, giving an amplification product of 1867 bp in *M. tuberculosis*. Another pair was composed of primer 2U415 (ATGAAGGCAAACACCACG), which anneals to the 12·7 kb fragment near the 3' end, and primer 6Lmbss (GCCGCCAAGGCAGCAGAGCAC), which anneals to the region common to both *M. bovis* and *M. tuberculosis* located downstream of the 12·7 kb fragment, giving an amplification product of 969 bp in *M. tuberculosis*. A third pair was formed by 4U4849 (CCGTGACAACGAACTCA) and 5L5984 (CCAGTCCTCGCTGTAGGT), which anneal to the central part of the 12·7 kb fragment, giving a 1135 bp product in *M. tuberculosis*. Amplifications with 1Umbss and 6Lmbss, which both anneal outside the 12·7 kb fragment, gave a 2198 bp product in *M. bovis*.

RESULTS

Identification of the 12·7 kb deletion/insertion

We had previously designed a pair of primers that amplified only *M. bovis* DNA and not *M. tuberculosis* DNA. In consequence the amplified fragment was called MBSS (for *Mycobacterium bovis* specific sequence). However, this fragment hybridized to DNA of both species. MBSS is downstream of the RD2 region of the *M. bovis* genome identified by Mahairas *et al.* (1996). A search for the *M. bovis* MBSS sequence in the *M. tuberculosis* H37Rv sequence database (Cole *et al.*, 1998) at the Sanger Centre (Cambridge, UK) revealed that the MBSS region was disrupted in this species by a 12·7 kb fragment (Fig. 1). The same result was observed when we searched the genome of the *M. tuberculosis* CSU 93 strain currently being sequenced at the Institute for Genomic Research (TIGR, Manassas, VA, USA).

Distribution of the 12·7 kb deletion/insertion

To assess the distribution of the 12·7 kb deletion/insertion in strains of the *M. tuberculosis* complex, a PCR-based strategy was followed. Three pairs of primers directed to both junctions and to the central part of the 12·7 kb region were used (Fig. 1). Only the *M. tuberculosis* genome was amplified with primers 4U4849 and 5L5984 giving products of the expected sizes; no amplification was observed in *M. bovis*, *M. microti*, *M. africanum* and mycobacterial strains isolated from wild seals (Fig. 2a). *M. tuberculosis* DNA, but not *M. bovis* DNA, was amplified with the other two pairs of primers, directed to both junctions (Fig. 2b, c). An additional band of 800 bp of unknown nature was observed in the amplification reactions with primers 1Umbss and 3L12505. Amplifications with 1Umbss and 6Lmbss gave a 2198 bp product only in *M. bovis* (Fig. 2d). To ensure that the genomic *M. tuberculosis* and *M. bovis* DNA in the samples could be amplified, PCRs were performed in parallel using primers against the IS6110 sequence that is present in both species. These PCRs always amplified a product of the expected size for IS6110 (result not shown).

A Southern blot experiment was performed by digesting chromosomal DNA from members of the *M. tubercu-*

Fig. 1. Schematic representation of the 12·7 kb fragment. The annealing sites of primers used in this work are also shown.

Fig. 2. PCR experiments using primers 4U4849 and 5L5984 (a), 2U415 and 6Lmbss (b), 1Umbss and 3L12505 (c), and 1Umbss and 6Lmbss (d). Genomic DNA from different *M. tuberculosis* complex bacteria were used as templates. Lanes: 1–10, *M. tuberculosis* (1–6, strains from Argentina; 7–8, strains from Brazil; 9, strain from Venezuela; 10, strain H37Rv); 11–20 *M. bovis* (11–15, cattle strains from Argentina; 16–17, human strains from Argentina; 18, strain AN5; 19, cattle strain from Paraguay; 20, cattle strain from Brazil); 21, *M. africanum*; 22, *M. microti*; 23, mycobacteria from wild seals. M, molecular mass markers (100 kb ladder, Gibco).

Fig. 3. RFLP using the amplification product of primers 4U4849 and 5L5984 as a probe on *M. tuberculosis* DNA. Lanes: 1–10, *M. bovis* (1–5, cattle strains from Argentina; 6–7, human strains from Argentina; 8, AN5 strain, 9–10, cattle strains from Brazil); 11–17, *M. tuberculosis* (11–14, strains from Argentina; 15, strain from Brazil; 16, strain from Venezuela; 17, strain H37Rv). Molecular mass markers are indicated on the left in kb.

losis complex with *PvuII* and probing it with a labelled amplification product generated by primers 4U4849 and 5L5984. At high stringency (65 °C) *M. tuberculosis* strains showed two hybridization bands of the expected size, 1.2 and 0.6 kb (the amplified portion has a *PvuII* site) (Fig. 3). No signal was observed in *M. bovis* (wild-

type or BCG). Rehybridization at a lower temperature (55 °C) gave the same result (not shown).

Sequence analysis of the 12.7 kb region

The 12.7 kb insertion/deletion was sequenced by the Sanger Centre as part of the virtual clone MTV051 (accession number AL022073); the precise location of the inserted/deleted region is at positions 1745–14473 bp on this clone. It starts at the genomic base 2208004 and extends up to 2220733. This identifies the region as being separated from the deletion point of the RD2 region by 337 bp. According to our search for ORFs (performed before this region was annotated) and to the sequence analysis at the Sanger Centre, there are 13 ORFs in one strand and 1 ORF in the complementary strand. ORFs (Rv1964 to Rv1977) were named according to Cole *et al.* (1998). The most striking observation from Blitz searches was that ORFs Rv1964–Rv1971 showed high homologies with three other regions of the *M. tuberculosis* chromosome, contained on cosmids MTCY19H5, MTCI28 and MTV023. ORF Rv1966 is homologous to the invasin-like protein Mcep, described by Riley and colleagues (Arruda *et al.*, 1993) and encoded in the MTCI28 cosmid (MTCI28.09). The putative proteins encoded in this region have characteristics that identify them as probable membrane proteins, as for example ORFs Rv1964, Rv1965, Rv1970, Rv1971, Rv1972 and Rv1974; most of the putative proteins have possible N-terminal signal sequences (see the Web page of the Sanger Centre for a detailed description). There is also some degree of homology among the ORFs present within the 12.7 kb fragment

itself, especially among Rv1966, Rv1965 and Rv1972 (data not shown).

The deletion in *M. bovis* stretches from the middle of Rv1964 to the middle of Rv1977 in *M. tuberculosis*. An alternative ORF of 585 bp is created, stretching from base 14495 to base 13911 of the sequence published by Mahairas *et al.* (1996).

DISCUSSION

In a previous study we demonstrated that a probe directed to the 3' flanking part of *mpb-64* gene could differentiate between members of the *M. tuberculosis* complex (Fisanotti *et al.*, 1997). We showed that the polymorphic region is located near the 3' end of the RD2 element described by Mahairas *et al.* (1996), a genomic locus present in *M. bovis* but absent from some strains of *M. bovis* BCG. Initially, we decided to clone the region downstream of the *mpb-64* gene using a PCR-based approach. Using this approach, we were only able to amplify *M. bovis* DNA; no amplification was seen with *M. tuberculosis* DNA. Consequently, we believed that this region represented a *M. bovis*-specific locus. However, the lack of amplification in *M. tuberculosis* was not because this sequence is unique to *M. bovis*, but because a fragment of 12.7 kb is present between the annealing sites of the primers in *M. tuberculosis*. The 12.7 kb insertion/deletion was present in all the *M. tuberculosis* strains tested, which included isolates from Argentina, Brazil and Venezuela. We have typed them by IS6110 RFLP and they are not related (data not shown). In addition, the strain sequenced at the TIGR Center also has the insert. All this data could suggest that the 12.7 kb insertion is a general property of *M. tuberculosis* strains.

The 12.7 kb region does not encode excisionases, integrases or other phage-like enzymes, suggesting that it is not a prophage. No significant repetitive sequences were found in the extremes of the insert. The region also does not appear to contain large repeated sequences and, furthermore, its G + C content is similar to that of the whole genome, indicating that it has not been recently acquired by *M. tuberculosis* from another organism. This locus is however different from another 12.7 kb region described by Brosch *et al.* (1998), as being absent from bovine tubercle strains.

According to sequence analysis, most of the ORFs appear to encode membrane or exported proteins. ORF Rv1966 is homologous to the Mcep invasin-like protein described by Riley and colleagues (Arruda *et al.*, 1993), which has been designated *mce-3* by Cole *et al.* (1998). Therefore, this region could play an essential role in cell entry, virulence and/or host range characteristics of *M. tuberculosis*. On the other hand, the genes in the 12.7 kb fragment seem to have an operon-like structure (S. Gordon and S. Cole, personal communication) similar in sequence and organization to three other regions named *mce* operons and cloned in the cosmids MTCY19H5, MTV023 and MTCI28. The multicopy number of the region may serve to protect the virulence

genes from mutations or they may each perform a role at different stages of infection.

Did *M. bovis* suffer a deletion or *M. tuberculosis* gain an insertion? This subject was addressed by analysing the 12.7 kb fragment junctions. An insertion or deletion of a fragment should create hybrid ORFs: two in *M. tuberculosis* or one in *M. bovis*. Both ORFs at the extremes of the fragment seem to be real ones as they are homologous to others in *M. tuberculosis* (in the *mce* operons) or other genera (see Sanger Centre annotation). This fact suggests that *M. bovis* suffered a deletion. A 585 bp ORF is found in *M. bovis* at the site of insertion of the *M. tuberculosis* DNA. A BLAST search revealed no significant homologies to other proteins. This ORF consists of the first 90 bp of ORF Rv1964; the rest lies in Rv1977, but in a different reading frame. The reading frame difference is explained by a base addition in *M. bovis* (data not shown) upstream of the 12.7 deletion point. From an evolutionary point of view, one may hypothesize that an ancestor strain having the 12.7 kb fragment diverged toward *M. tuberculosis* and *M. bovis*, and then *M. bovis* lost the fragment. The deletion must have occurred soon after the divergence because all *M. tuberculosis* and *M. bovis* strains analysed to date were identical with respect to the fragment content. However, this will be further assessed by analysing more strains from different hosts and geographical origins.

As *M. bovis* is difficult to cultivate, a molecular method to distinguish *M. bovis* from *M. tuberculosis* would be welcomed in countries such as Argentina where *M. bovis* infection in humans in certain areas represents 6% of total tuberculosis cases (Latini *et al.*, 1990), or in some regions of Africa where the incidence of tuberculosis in cattle and HIV in humans is very high, increasing the risk of *M. bovis* infection in humans (Cosivi *et al.*, 1998). Moreover, a tuberculosis outbreak provoked by multi-drug-resistant *M. bovis* was described in Spain (Blazquez *et al.*, 1997). The 12.7 kb region may serve as a very useful species-specific probe for differentiating *M. bovis* from *M. tuberculosis* isolates. To date, two species-specific mycobacterial DNA elements in the *M. tuberculosis* complex have been described: the *M. tuberculosis mpt-40* gene (Del Portillo *et al.*, 1991) and a 500 bp fragment specific for *M. bovis* (Rodríguez *et al.*, 1995). MPT40 was originally described as being produced only by *M. tuberculosis*, though later studies showed that it could also be detected in *M. africanum* and some *M. microti* isolates (Liébaona *et al.*, 1996). Further problems with the use of the *mpt-40* gene for speciation were reported by Weil *et al.* (1996) who showed that some strains of *M. tuberculosis* also lack this locus. The 500 bp *M. bovis* specific sequence described by Rodríguez *et al.* (1995) is present in all *M. bovis* isolates tested but not in *M. tuberculosis*. However, database searches indicated that the 3' portion of this 500 bp sequence is also present in *M. tuberculosis*, hence allowing hybridization in Southern blot conditions (data not shown). It appears to us that a PCR-based approach based on the 12.7 kb region offers an

excellent test to identify an isolate as *M. bovis* or *M. tuberculosis*, and we are currently testing this possibility.

ACKNOWLEDGEMENTS

This work was supported by the Centro Argentino Brasileño de Biotecnología (CABBIO) and Fondo Mixto de Cooperación Hispano-Argentino. We thank Laurita Boschioli for discussions and Hayed Gill for technical help. The highly valuable suggestions of Stephen Gordon and Stewart Cole (Institut Pasteur, Paris, France) are gratefully acknowledged.

M. I. R. and A. C. are fellows of the National Research Council of Argentina (CONICET).

REFERENCES

- Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T. & Riley, L. W. (1993). Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261**, 1454–1457.
- Blazquez, J., Delosmonteros, L. E. E., Samper, S., Martin, C., Guerrero, A., Cobo, J., Vanembden, J., Baquero, F. & Gomez-mampaso, E. (1997). Genetic characterization of multidrug-resistant *Mycobacterium bovis* strains from a hospital outbreak involving human immunodeficiency virus-positive patients. *J Clin Microbiol* **35**, 1390–1393.
- Brosch, R., Gordon, S. V., Billault, A., Garnier, T., Eiglmeyer, K., Soravito, C., Barrell, B. G. & Cole, S. T. (1998). Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Immun* **66**, 2221–2229.
- Cole, S. T., Brosch, R., Parkhill, J. & 39 other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.
- Cosivi, O., Grange, J. M., Darbon, C. J. & 7 other authors (1998). Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. *Emerg Infect Dis* **4**, 59–70.
- Cousins, D. V., Francis, B. R., Gow, B. L., Collins, D. M., McGlashan, C. H., Gregory, A. & Mackenzie, R. M. (1990). Tuberculosis in captive seals: bacteriological studies on an isolate belonging to the *Mycobacterium tuberculosis* complex. *Res Vet Sci* **48**, 196–200.
- Cousins, D., Wilton, S. & Francis, B. (1991). Use of DNA amplification for the rapid identification of *Mycobacterium bovis*. *Vet Microbiol* **27**, 187–195.
- Del Portillo, P., Murillo, L. A. & Patarroyo, M. E. (1991). Amplification of a species-specific DNA fragment of *Mycobacterium tuberculosis* and its possible use in diagnosis. *J Clin Microbiol* **29**, 2163–2168.
- Fisanotti, J. C., Romano, M. I., Alito, A., Bigi, F. & Cataldi, A. (1997). A *mpb-64* flanking sequence specific for *Mycobacterium bovis*. *Res Microbiol* **148**, 427–435.
- Heifets, L. B. & Good, R. (1994). Current laboratory methods for the diagnosis of tuberculosis. In *Tuberculosis: Pathogenesis, Protection, and Control*. Edited by B. Bloom. Washington, DC: American Society for Microbiology.
- Latini, M. S., Latini, O. A., Lopez, M. L. & Ceconi, J. O. (1990). Tuberculosis bovina en seres humanos. *Rev Argent Torax* **51**, 13–16.
- Liébana, E., Aranaz, A., Francis, B. & Cousins, D. (1996). Assessment of genetic markers for species differentiation within the *Mycobacterium tuberculosis* complex. *J Clin Microbiol* **34**, 933–936.
- Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. (1996). Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* **178**, 1274–1282.
- Rodriguez, J. G., Mejia, G. A., Del Portillo, P., Patarroyo, M. E. & Murillo, L. A. (1995). Species-specific identification of *Mycobacterium bovis* by PCR. *Microbiology* **141**, 2131–2138.
- Romano, M. I., Alito, A., Bigi, F., Fisanotti, J. C. & Cataldi, A. (1995). Genetic characterization of mycobacteria from South American wild seals. *Vet Microbiol* **47**, 89–98.
- Runyon, E. H., Wayne, L. G. & Kubica, G. P. (1974). Family II. *Mycobacteriaceae*. In *Bergey's Manual of Determinative Bacteriology*, 8th edn, p. 681. Edited by R. E. Buchanan & N. E. Gibbons. Baltimore: Williams & Wilkins.
- Scorpio, A., Collins, D., Whipple, D., Cave, D., Bates, J. & Zhang, Y. (1997). Rapid differentiation of bovine and human tubercle bacilli based on a characteristic mutation in the bovine pyrazinamidase gene. *J Clin Microbiol* **35**, 106–110.
- Sreevatsan, S., Escalante, P., Pan, X. & 11 other authors (1996). Identification of a polymorphic nucleotide in *oxyR* specific for *Mycobacterium bovis*. *J Clin Microbiol* **34**, 2007–2010.
- Weil, A., Plikaytis, B. B., Butler, W. R., Woodley, C. L. & Shinnick, T. M. (1996). The *mtp40* gene is not present in all strains of *Mycobacterium tuberculosis*. *J Clin Microbiol* **34**, 2309–2311.

Received 27 July 1998; revised 7 January 1999; accepted 12 January 1999.