Genome structure in the vole bacillus, *Mycobacterium microti*, a member of the *Mycobacterium tuberculosis* complex with a low virulence for humans

Cristiane C. Frota,† Debbie M. Hunt,† Roger S. Buxton,† Lisa Rickman,†§ Jason Hinds,‡ Kristin Kremer,§ Dick van Soolingen‡ and M. Joseph Colston†

†Division of Mycobacterial Research, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
‡Department of Medical Microbiology, St George’s Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK
§National Mycobacteria Reference Laboratory, National Institute of Public Health and the Environment (RIVM), 3720 BA Bilthoven, The Netherlands

*Correspondence* Roger S. Buxton rbuxton@nimr.mrc.ac.uk

**INTRODUCTION**

*Mycobacterium microti* belongs to the *Mycobacterium tuberculosis* complex, which is composed of closely related species that are difficult to distinguish using biochemical properties alone (Tsukamura et al., 1985; Rastogi et al., 2001). Wells first discovered *M. microti* in the 1930s, and named it the ‘vole bacillus’ (Wells, 1937). It was later designated *M. tuberculosis* var. *muris*, because it could not be differentiated from *M. tuberculosis*. At that time, the bacillus was described as the causative agent of an epizootic disease in the wild English field vole (*Microtus agrestis*) (Brooke, 1941). *M. microti* causes disease in a variety of mammalian species, including wood mice (*Apodemus sylvaticus*), bank voles (*Clethrionomys glareolus*) and shrews (*Sorex araneus*). Disease is rare in llamas (*Lama vicugna molina*), cats and pigs, and very rare in humans (Huijtena & Jaartsveld, 1967; Wayne & Kubica, 1986; Kremer et al., 1998; Cavanagh et al., 2002). *M. microti* differs from other members of the *M. tuberculosis* complex in its curved cell morphology (Pattyn et al., 1970), extremely slow...
growth in vitro, and distinct, host-specific pathogenicity for laboratory animals (van Soolingen et al., 1998).

*M. microti* has also been employed as a vaccine against tuberculosis and was reported to give a high degree of protection, similar to that seen with BCG (*Mycobacterium bovis* bacille Calmette-Guérin) (Sula & Radkovsky, 1976). An attenuated strain of *M. microti* was used as a vaccine in Czechoslovakia from 1951 to 1969, and about 500 000 people, mostly newborn, were vaccinated intradermally. In a separate study in the UK between 1950 and 1952, a vaccine was prepared from non-attenuated *M. microti* and used in vaccination trials organized by the Medical Research Council (Hart & Sutherland, 1977). The trials in Czechoslovakia and the UK gave similar results, with the *M. microti* vaccine conferring about 75% protection. Furthermore, the *M. microti* vaccine showed low allergenic potency, making it less likely than BCG to compromise the tuberculin test in the vaccinated population. The *M. microti* vaccine induced fewer than 30% positive skin test conversions in response to Purified Protein Derivative (PPD) from *M. tuberculosis* (Brooke, 1941; Sula & Radkovsky, 1976; Bloom & Fine, 1994). A recent comparison of efficacy between the Pasteur substrain of BCG and the *M. microti* vaccine showed that both could provide protection against tuberculosis in rabbits (Dannenberg et al., 1998). Moreover, cases of *M. microti* have been employed as a vaccine against tuberculosis in both immunocompromised and immunocompetent individuals, and we included these strains in our investigation. These strains were identified as *M. microti* using the spacer oligonucleotide typing, ‘poligotyping’, DNA fingerprinting method (van Soolingen et al., 1998).

Identification of *M. microti* by traditional methods is not easy, so the prevalence, geographical distribution and clinical importance of *M. microti* may have been underestimated (van Soolingen, 2001). *M. microti* grows slowly on solid media supplemented with pyruvate, and shows similar biochemical properties to *M. tuberculosis*, giving variable results for pyrazinamidase and urease activity and niacin accumulation (Levy-Frebault & Portaels, 1992; van Soolingen et al., 1998). Two separate studies have shown the usefulness of novel genetic markers (IS6110) to characterize *M. microti* isolates (Krem et al., 1998; van Soolingen et al., 1998). Moreover, cases of *M. microti*-derived tuberculosis in both immunocompromised and immunocompetent human patients have been identified using molecular methodologies (Foudraine et al., 1998; Niemann et al., 2000).

The virulence mechanisms of *M. microti* in its natural hosts, and its low virulence in humans, are not well understood. Like that of *M. tuberculosis*, the *M. microti* cell wall contains 34% by weight of mycolates. In both species, exponential-phase *in vitro* cultures and bacteria harvested from mouse lungs contain a high proportion of ketomycolates, whereas in stationary-phase cultures, ketomycolates decrease rapidly to give proportions similar to those of methoxymycolates (Davidson et al., 1982; Watanabe et al., 2001). Although there is no clear relationship between possession of a particular mycolate and strain pathogenicity, the same mycolate types are present in all *M. tuberculosis* complex strains, differing only in chain length and other structural features (Daffé & Draper, 1998; Kremer et al., 2000).

In this study, comparative genomics, by microarray assay methods, were employed to delineate the genetic differences between *M. tuberculosis* and *M. microti*, and to relate these to virulence. DNA microarrays were used to study a group of 12 strains classified as *M. microti*, including the strain OV254, originally isolated from voles in the 1930s (Wells, 1937). *M. microti* strains isolated from other mammals, such as pig, llama and hyrax, were also included. Lastly, *M. microti* strains have more recently been isolated from human tuberculosis infections (van Soolingen et al., 1998), from both immunocompromised and immunocompetent individuals, and we included these strains in our investigation. These strains were identified as *M. microti* using the spacer oligonucleotide typing, ‘poligotyping’, DNA fingerprinting method (van Soolingen et al., 1998).

**METHODS**

**Bacterial strains.** The 12 strains of *M. microti* used in the work described here are listed in Table 1. *M. tuberculosis* H37Rv was used as a comparison in genomic experiments. All strains were grown in Dubos broth (Beckton Dickinson) supplemented with 0.05% (w/v) Tween 80 (Sigma) and 0.1% (v/v) Dubos medium albumin (Beckton Dickinson). All mycobacterial liquid cultures were grown in 1 litre rolling bottles and incubated in a rolling incubator (2 r.p.m.) at 37°C. Growth was monitored by measurement of OD600 using a Cecil CE 1010 spectrophotometer. All *M. microti* isolates in this study had been subjected to identification by poligotyping DNA fingerprint analysis (van Soolingen et al., 1998).

**DNA extraction.** Chromosomal DNA was isolated by the standard lysozyme method of Larsen (2000).

**Synthesis of labelled DNA.** Genomic DNA was labelled with dCTP coupled to Cy3 or Cy5 dyes (Amersham Pharmacia Biotech). DNA (3 μg) was mixed with 3 μl random hexamer primers (1 μg ml⁻¹; Invitrogen) in a total volume of 41.5 μl, and the primers were annealed to the DNA by incubating the mix at 95°C for 5 min, followed by rapid cooling on ice. The tubes were centrifuged briefly, and the following were added: 5 μl 10× Klenow polymerase buffer (Promega), 1 μl dNTP mix (5 mM dA/G/TTP and 2 mM dCTP, Amersham Pharmacia Biotech), 1-5 μl Cy3- or Cy5-dCTP (25 nM, Amersham Pharmacia Biotech) and 1 μl Klenow DNA polymerase (5 U μl⁻¹; Promega). The reaction mixture was incubated at 37°C for 90 min, and labelled cDNA was eluted in 13 μl distilled water after extraction in a MiniElute column (Qiagen). The probe was mixed with 3-2 μl 20× SSC (0.15 M NaCl, 0.015 M sodium citrate), and 0-3 μl 20% (v/v) SDS, and then denatured by heating at 95°C for 2 min. The mix was allowed to cool briefly before being added to the array.

**Slide processing.** Spotted microarray slides, representing 100% of the genes of the *M. tuberculosis* H37Rv genome, were prepared at the Bacterial Microarray Group, St George's Hospital Medical School, London. The rehydration and blocking steps were carried out as described by Eisen & Brown (1999).

**Hybridization and washing.** The processed array slides were placed in prehybridization solution (3×5× SSC, 0.1%, w/v, SDS and
Table 1. Source of the M. microti strains used in the study

NIMR, derived from the NIMR, UK, strain collection; D. van Soolingen, derived from the National Mycobacteria Reference Laboratory, The Netherlands, strain collection.

<table>
<thead>
<tr>
<th>Source host</th>
<th>Strain</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vole (UK)</td>
<td>OV254</td>
<td>NIMR</td>
</tr>
<tr>
<td></td>
<td>15498</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td></td>
<td>scabl5</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td></td>
<td>00-0421</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td>Pig</td>
<td>15911</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td>Llama</td>
<td>15912</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td>Hyrax</td>
<td>15499</td>
<td>Cavanagh et al., 2002; D. van Soolingen</td>
</tr>
<tr>
<td>Human (immunocompromised; Netherlands)</td>
<td>94-2272</td>
<td>van Soolingen et al., 1998</td>
</tr>
<tr>
<td>Human, immunocompetent</td>
<td>97-770</td>
<td>van Soolingen et al., 1998</td>
</tr>
<tr>
<td></td>
<td>99-1853</td>
<td>D. van Soolingen</td>
</tr>
<tr>
<td></td>
<td>2001-1205</td>
<td>D. van Soolingen</td>
</tr>
<tr>
<td></td>
<td>2001-1206</td>
<td>D. van Soolingen</td>
</tr>
</tbody>
</table>

0.1% fraction V BSA [Sigma] at 60 °C for 20 min, to block non-specific probe-binding sites. The slides were then rinsed sequentially for 1 min in distilled water and 1 min in 2-propanol (Merck), and then centrifuged in 50 ml centrifuge tubes at 433 g for 5 min. Samples of denatured, labelled probe (16 µl) were applied to a dried, prehybridized slide. Post-hybridization and slide-washing steps were performed as described by Eisen & Brown (1999).

Data collection and analysis. Slides were scanned in a dual-laser microarray scanner (Genepix 4000A, Axon Instruments) and the images obtained were analysed using GenePix Pro software (Axon Instruments). The software calculated the mean signal intensity and local background for each spot on the array and subtracted the local background from the signal intensity of each spot. Spots showing a high background or poor hybridization were eliminated from the analysis. Data obtained from six slides, including dye swaps, were analysed with Genespring software, version 4.1.5 (Silicon Genetics). To account for dye swap, the signal channel and control channel measurements for dye swap samples were reversed. Each gene's measured intensity was divided by its control channel value in each sample; if the control channel was below 10, then 10 was used instead. If the control channel and the signal channel were both below 10 then no data were reported. Each measurement was divided by the 50-96th percentile of all measurements in that sample. The percentile was calculated using only genes marked present.

PCR amplification and DNA sequencing of deletions. Primers were designed to flank known deletion sites (Table 2). DNA was amplified using HotStar Taq polymerase (Qiagen), and the products purified using a Qiaquick PCR purification kit (Qiagen). DNA sequencing was carried out on an ABI377 DNA sequencer using a Big Dye terminator kit (ABI).

RESULTS

Microarray hybridization of genomic DNA

Each of the 12 strains of M. microti was compared with M. tuberculosis H37Rv by carrying out competitive hybridization with M. microti and M. tuberculosis DNA. Each comparison was carried out in triplicate. The Cy5/Cy3 fluorescence ratios, calculated for each ORF represented on the array, were normalized using Genespring software, and the genes were considered to be significantly differentially hybridized compared with the M. tuberculosis control if they displayed at least a twofold difference in the ratio. Fig. 1 shows the hybridization ratio between the two species, indicating that several genes are deleted from the M. microti genome. The extent of the deletions in one strain, OV254, was determined by sequencing the PCR products obtained using flanking oligonucleotide primers (Table 2).

Genomic analysis also suggested that the sequence of individual genes was similar in M. tuberculosis and M. microti, giving equivalent hybridization by M. tuberculosis and M. microti DNA to genes present in both species. Mismatched hybridization signals probably reflected non-equivalent representation in the genome, as a result of repeated elements or deletions in the M. microti genes. However, it should be emphasized that deletions represent only a subset of the total genetic variability; single nucleotide polymorphisms would not be detected using the microarray technique used in this study. Nevertheless, the identification of regions that are present in M. tuberculosis and consistently absent from M. microti strains may provide insights into the phenotypic differences between the two species (Brodin et al., 2002; Brosch et al., 2000).

Table 3 shows genes deleted in M. microti and demonstrates that, apart from three strains isolated from humans, the pattern of deleted sequences was different in every strain examined. In comparison with M. tuberculosis H37Rv, each M. microti strain had an average of 9-4 deleted regions. Among the 12 strains, a total of 13 different deleted sequences were detected.

Nine of the deleted sequences were previously identified in BCG (Gordon et al., 1999); therefore the nomenclature for these deletions was based on that of BCG (from RD1 to
RD10). The deleted regions RD3, RD7, RD9 and RD10 were identical to the ones in BCG described by Behr et al. (1999) and Gordon et al. (1999). The RD3 and RD9 deletions have also been identified in clinical isolates of \textit{M. tuberculosis}, where they are referred to as DS5 and DS21, respectively (Kato-Maeda et al., 2001). In addition, four new deletions, MiD1, MiD2, MiD3 (standing for \textit{microti} deletion, named by Brodin et al., 2002) and RD1b (to differentiate it from RD1) were identified. Of the 13 deleted regions, only the RD3 deletion was found in all \textit{M. microti} strains. The RD7, RD8 and MiD1 deletions were found in 11 strains, and RD4, RD6 and MiD2 were found only in a few strains.

**Presence of \textit{esat6} within the RD1 deletion in \textit{M. microti}**

\textit{esat6} (Rv3875) appears in a cluster region along with \textit{lhp} (Rv3874) in \textit{M. tuberculosis} and other GC-rich bacteria, including \textit{M. bovis} and \textit{Mycobacterium leprae} (Gey van Pittius et al., 2001). Recent results analysing the function of the ESAT-6 and CFP-10 proteins, known to form a tight 1:1 complex (Renshaw et al., 2002), have shown them to be required for invasion of lung interstitial tissue (Hsu et al., 2003). The extent of the RD1 deletion, where \textit{esat6} is located, was found to be highly variable in \textit{M. microti}.
Table 3. Distribution of deletions in the *M. microti* strains

The figures in each column indicate the numerical designation of each ORF in the genome of *M. tuberculosis* H37Rv, omitting the 'Rv' prefix. The 'c' that follows some numbers indicates that the gene has a reverse orientation. It should be noted that the extent of the RD1, 4, 5, 6 and 8 deletions is not identical in BCG.

<table>
<thead>
<tr>
<th>M. microti strain</th>
<th>Region deleted</th>
<th>RD1</th>
<th>RD3</th>
<th>RD4</th>
<th>RD5*</th>
<th>RD6</th>
<th>RD7</th>
<th>RD8</th>
<th>RD9</th>
<th>RD10</th>
<th>MiD1</th>
<th>RD1/#</th>
<th>MiD2</th>
<th>MiD3</th>
</tr>
</thead>
</table>

*Strains human 97-0770 and hyrax also have plcD (Rv1755c) deleted.

The figures in each column indicate the numerical designation of each ORF in the genome of *M. tuberculosis* H37Rv, omitting the 'Rv' prefix. The 'c' that follows some numbers indicates that the gene has a reverse orientation. It should be noted that the extent of the RD1, 4, 5, 6 and 8 deletions is not identical in BCG.
strains. The RD8 region was deleted in the majority of the strains, but in each case the deletion was not continuous, with two genes (Rv3619c and Rv3620c, encoding proteins of unknown function) present in all the analysed strains, as evidenced by hybridization to the Rv3619c and Rv3620c microarray probes. However when the microarray probes for these two genes were analysed for potential cross-hybridization with other \textit{M. tuberculosis} genes, each gave very high BLAST homology scores with two other genes (data not shown). Thus the presence of Rv3619c and Rv3620c within the \textit{M. microti} genome could be an artefact due to cross-hybridization. We believe, therefore, that the RD8 deletion is probably continuous in \textit{M. microti}. The RD8 deletion in BCG extends from Rv3617 to Rv3622c.

\textbf{Newly identified deletions: MiD1, MiD2 and MiD3}

Two of the newly identified deletions, MiD1 and MiD3, were deleted in the majority of the \textit{M. microti} strains; MiD1 encodes conserved hypothetical proteins and MiD3 encodes insertion sequences and PE/PPE proteins. The MiD2 region was deleted only in two vole isolates (OV254 and vole 15498), both being strains originally isolated from voles in the 1960s. Therefore, this deletion could have occurred during multiple passages of these strains in culture media. The Rv3345c, Rv3346c and Rv3348 microarray probes from the variable MiD3 region also show cross-hybridization with other genes encoding PE/PPE and PGRS proteins; nevertheless the deleted region from \textit{PE PGRS50} (Rv3345c) to Rv3349c was confirmed by PCR amplification and sequencing (see Table 2).

\textbf{DISCUSSION}

The genomic analysis using DNA microarrays, allied to the confirmation by sequencing across the deleted regions, have been shown to be useful, rapid and simple techniques to identify which genes are missing in \textit{M. microti} relative to \textit{M. tuberculosis}. Since it is based on a \textit{M. tuberculosis} DNA microarray, the technique does not provide information about sequences present in \textit{M. microti} but absent from \textit{M. tuberculosis}. The genomic deletions identified in this
study provide insights into mycobacterial diversity, and suggest that many genes in the genome are non-essential. Although the results are mostly compatible with *M. tuberculosis* and *M. microti* having a common ancestor (Brosch et al., 2002), there are some discrepancies. For example, we found that not all *M. microti* strains had RD9 deleted; we have no ready explanation for this apparent disagreement.

With regard to deleted regions and virulence, this study shows that it is difficult to ascribe virulence to any particular pattern of deletion. Although RD1 of BCG and *M. microti* is thought to be crucial for attenuation (Mahairas et al., 1996; Brodin et al., 2002; Pym et al., 2002), in this study three of the four *M. microti* strains from immunocompetent patients had the RD1 region deleted. It is also noticeable that of the 12 strains studied, only three were identical in deletion pattern; these strains were all isolated from immunocompetent humans, suggesting that they may have arisen from a single source.

The genomic comparison of *M. tuberculosis* with *M. microti* identified 13 deletion regions in *M. microti*, compared to *M. tuberculosis*. Nine deleted regions (RD1 to RD10) had already been reported in BCG (Behr et al., 1999; Gordon et al., 1999), although five of these (RD1, RD4, RD5, RD6 and RD8) were not identical to those described for BCG. Although *M. microti* mainly causes disease in small rodents, it can also, rarely, cause disease in man and other large mammals, indicating that the deleted regions do not totally eliminate virulence (van Soolingen et al., 1998; Horstkotte et al., 2001). RD1 is thought to be the principal deletion resulting in attenuation of BCG and *M. microti* (Mahairas et al., 1996; Brodin et al., 2002; Pym et al., 2002); even so, one of the *M. microti* strains investigated here did not have an RD1 deletion and two more had only a single gene deleted within this region. For this reason, although RD1 deletions undoubtedly contribute to attenuation of *M. microti* (Pym et al., 2002, 2003), they are not the only mechanism of attenuation. In fact, only deletion RD3 was present in all of the strains examined, while RD7, RD8 and MiD1 were found in almost all *M. microti*, and may therefore have some relation to the host range of *M. microti*.

Interestingly, the human isolate 97-0770 has an exceptional and divergent deletion pattern, and since this isolate caused

---

Fig. 3. Overview of the RD5 region, showing the genes deleted in the various *M. microti* strains, as well as *M. bovis* BCG. Only in strain OV254 was sequencing performed across the deletion (see legend to Fig. 2).
pulmonary tuberculosis in an immunocompetent 34 year old male (van Soolingen et al., 1998), it may have unknown extra genes enabling it to cause disease in immunocompetent individuals.

In general, although the results support the evolutionary scenario proposed for the M. tuberculosis complex by Brosch et al. (2002), in that most isolates of M. microti had the RD7, RD8, RD9 and RD10 deletions, some isolates did not appear to fit into this neat picture. In particular, one isolate did not have the RD7, RD8, RD9 or RD10 deletions, another did not have the RD9 deletion and a third lacked both the RD9 and RD10 deletions – an apparent discrepancy for which at present there is no explanation.

ACKNOWLEDGEMENTS

We should like to thank Dr Philip Butcher for providing the facilities of the Bacterial Microarray Group, St George’s Hospital Medical School (BuG@S), and also for the supply of the M. tuberculosis DNA microarrays and for advice. We should also like to thank the Wellcome Trust for funding the Multi-Collaborative Microbial Pathogen Microarray Facility, under its Functional Genomics Resources Initiative. C. C. F. gratefully acknowledges the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil). K. K. and D. v. S, gratefully acknowledge EC grant 2000-0630, on Molecular Epidemiology of Tuberculosis. Work at NIMR was funded by the Medical Research Council.

REFERENCES


Hsu, T., Hingley-Wilson, S. M., Chen, B. & 14 other authors (2003). The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A 100, 12420–12425.


