A polymorphic region in *Mycobacterium abscessus* contains a novel insertion sequence element

Susan T. Howard,1† Thomas F. Byrd1,2 and C. Richard Lyons1,3

Author for correspondence: C. Richard Lyons. Tel: +1 505 272 4450. Fax: +1 505 272 9912.
e-mail: clyons@salud.unm.edu

A polymorphic region was discovered in the genetically uncharacterized opportunistic pathogen *Mycobacterium abscessus*. The region contains a novel 1.7 kb insertion sequence (IS) named ISMab1. ISMab1 contains two complete ORFs and one partial ORF located in segments with over 80% nucleotide identity to *Mycobacterium avium* IS1601 and IS999 and to previously unreported IS-like elements from *Mycobacterium smegmatis*. The marked similarity within this family of elements is supportive of horizontal transfer between environmental mycobacterial species. In clinical isolates, ISMab1 was either present as a single copy or absent. The polymorphic region containing ISMab1 was identified by genomic subtraction between a parental strain and phenotypic variant. The variant has a 14.2 kb genomic deletion and this is flanked in the parental strain by complex arrays of inverted and direct repeats. Clinical isolates of *M. abscessus* were probed for the deletion and flanking sequences and two were found to be missing more than 20 kb. No regional deletions were found in the type strain, ATCC 19977. Although *M. abscessus* is a rapidly growing species, comparative sequence analysis of 23 kb from the polymorphic region showed that most local ORFs have greater amino acid identity to proteins encoded by genes from the slowly growing mycobacteria, *M. avium* and *Mycobacterium tuberculosis*, than to the rapid-grower *M. smegmatis*. Several ORFs also have strong similarity to *Pseudomonas aeruginosa* genes with a potential role in β-oxidation.

**Keywords:** rapidly growing mycobacteria, insertion element, horizontal transfer, genetic variation

**INTRODUCTION**

The genus *Mycobacterium* has been divided into two groups on the basis of growth rate and, in general, this division is supported by phylogenetic analyses of the 16S rRNA gene (Pitulle *et al.*, 1992). The major pathogens *Mycobacterium leprae*, the *Mycobacterium tuberculosis* complex and *Mycobacterium avium* complex, are members of the slowly growing mycobacteria. The second group, the rapidly growing mycobacteria (RGM), includes saprophytic species found in soil and water systems (Falkinham, 1996; Howard & Byrd, 2000). Comparative analysis of 16S rRNA gene sequences indicates that the RGM may be the phylogenetically older group (Pitulle *et al.*, 1992).

Several species of RGM, particularly *Mycobacterium abscessus*, *Mycobacterium chelonae* and *Mycobacterium fortuitum*, are opportunistic pathogens and can cause infections ranging from localized abscesses to pulmonary and disseminated disease (Griffith *et al.*, 1993; Howard & Byrd, 2000; Wright & Wallace, 1995). RGM-associated disease tends to be sporadic and usually associated with injury or surgical procedures (Wright & Wallace, 1995). However, large outbreaks of post-operative wound infections (Chadha *et al.*, 1998) and post-injection abscesses (Galil *et al.*, 1999; Villaneuva *et al.*, 1997) have been attributed to contami-
nation of medical reagents or instruments with M. abscessus. M. abscessus is also the leading cause of pulmonary disease due to RGM (Griffith et al., 1993).

M. abscessus was originally classified as a subspecies of M. chelonae but has been reclassified as a distinct species on the basis of DNA–DNA hybridization and sequence analysis of the 16S rRNA gene (Kusunoki & Ezaki, 1992; Pitulle et al., 1992). Some phylogenetic analyses place M. abscessus close to M. chelonae (Brown et al., 1999; Domenech et al., 1997), whereas others place it on an older branch of the phylogenetic tree (Pitulle et al., 1992; Shojai et al., 1997). Epidemiological investigations utilizing RFLP analysis, PFGE or random amplified PCR (Villaneuva et al., 1997; Wallace et al., 1993; Zhang et al., 1997) have indicated that there is genetic variation within M. abscessus but specific polymorphic regions have not been identified. Molecular characterization has focused primarily on genes that allow rapid speciation of isolates (Kim et al., 1999; Ringuet et al., 1999) and those with a role in drug resistance (Ainsa et al., 1998; Guillemin et al., 1995; Pramananinan et al., 1998). Available genomic data on M. abscessus are primarily limited to partial sequences of housekeeping genes in the GenBank database.

We have been characterizing strain 390R, which is a clinical isolate of M. abscessus, and its mutant derivative, 390S (Byrd & Lyons, 1999). During the course of our investigation, we identified a large polymorphic region in M. abscessus and discovered a novel insertion sequence (IS) with significant nucleotide identity to IS elements from both slowly growing and rapidly growing environmental mycobacteria.

METHODS

Bacterial strains and cultures. Rough (R) and smooth (S) variants of M. abscessus were obtained as described by Byrd & Lyons (1999) and have been designated M. abscessus strain 390R and M. abscessus strain 390S (B. Beamun, personal communication). M. abscessus strains 1056, 1475, 6639, 8243 and 8988 were kindly supplied by P. Conville, Public Health Service, NIH. Strains were verified to be M. abscessus by sequence analysis of the region of the 16S rRNA gene corresponding to Escherichia coli positions 1001–1027 (Brown et al., 1999) (data not shown). The type strain ATCC 19977 was obtained from the American Type Culture Collection. Mycobacteria were cultured as described by Byrd & Lyons (1999).

Extraction of mycobacterial genomic DNA. For routine extraction of DNA, bacterial cells were lysed by bead-beating using 0.1 mm silwica/zirconium beads in a BioSpec Mini beadbeater and DNA was extracted using a Puregene DNA isolation kit (Genta Systems) followed by phenol/chloroform extraction and precipitation with 2-propanol. For the genomic subtraction procedure, DNA was extracted from bead-beaten cells using the Qiagen genomic DNA maxiprep extraction kit with minor modifications.

Genomic subtraction. The protocol of Straus & Ausubel (1990) was followed with some modifications.

(i) Preparation of DNA. DNA from M. abscessus 390S and 390R was used as driver and test DNA, respectively. 390S DNA was sheared into random fragments using a Branson Sonifier 450 cup horn sonicator at 70% output for 2 × 25 s. Fragment sizes ranged from 0.5 to 12 kb, but the majority of fragments were 1–4 kb in size. Sheared DNA was biotinylated using the Clontech photoactivatable biotin-labelling kit, according to the manufacturer’s instructions. After butanol extraction and ethanol/ammonium acetate precipitation, biotinylated DNA was resuspended at a concentration of 1 µg ml⁻¹ in 1 × EX buffer [10 mM N-[2-hydroxyethyl]piperazine-N’-(3-propanesulfonic acid), pH 8.0/1 mM EDTA] (Straus & Ausubel, 1990). 390R DNA was partially digested with the restriction enzyme Sau3AI, treated with phenol/chloroform and ethanol-precipitated. Fragments ranged up to 12 kb in size but the majority of fragments were <600 bp.

(ii) Subtraction. Five micrograms of biotinylated 390S DNA, 0.5 µg Sau3AI-digested 390R DNA and 40 µg yeast tRNA in a final volume of 20 µl 2 × EE were denatured, lyophilized and resuspended in 4 µl 2.5 × EE. After the addition of 1 µl 3 M NaCl, samples were annealed overnight at 65 °C in a Perkin-Elmer Gene Amp 9600 Thermocycler with heated lid. Ninety-five microlitres of EEN (1 × EE/500 mM NaCl) was added and the total mixture was added to MagneSphere streptavidin-coated paramagnetic particles (Promega) prewashed once in 0.5 × SSC and three times in EEN. Samples were incubated with paramagnetic particles for 20 min at room temperature with occasional mixing, the particles were captured, and the supernatant was incubated with a fresh aliquot of prewashed particles. After particle capture, supernatants were pooled, mixed with 20 µg tRNA and precipitated with ethanol. Pellets were washed with 70% ethanol, dried and resuspended in 10 µl EE. After removal of 1 µl for analysis, the sample was mixed with 1 µl EE, 5 µl biotinylated 390S DNA, 1 µl tRNA (20 µg) and 3 µl 10 × EE. The subtraction procedure was then repeated for another four cycles beginning with the denaturation step.

(iii) Addition of adapters and PCR. For amplification of the remaining fragments of 390R DNA, the protocol of Straus & Ausubel (1990) was used. Sau3AI adapters were prepared and ligated to the pool of fragments followed by amplification using a primer that annealed to the adapters. Electrophoresis of the PCR products on an agarose gel showed a range of amplicons <800 bp in size for each sample. Several discrete bands of 300–500 bp were visible beginning with the third round of subtraction.

(iv) Extraction of fragments for hybridization and cloning. From subtraction round five, amplicons between 350 and 500 bp were excised from an agarose gel, purified and re-amplified by PCR. Amplicons between 350 and 500 bp were again purified and then radiolabelled for use as a probe or cloned into the TA cloning vector pCR2.1 (Invitrogen) for sequence analysis.

Southern blotting. For Southern blots, 2 µg genomic DNA was digested with 20 U restriction enzyme and electrophoresed on 0.8% agarose gels. Following alkaline treatment and neutralization (Sambrook et al., 1989), gels were blotted onto GeneScreen Plus hybridization transfer membranes (NEN) by capillary transfer. Membranes were prehybridized in Clontech Express Hybridization solution and then hybridized overnight with [32P]dCTP-labelled DNA probes. Washed membranes were exposed to autoradiographic film.

Cloning of genomic restriction fragments. Restriction fragments of genomic DNA in the desired size range were eluted from agarose gels, ligated into pGEM3zf+ (Promega) and transformed into E. coli. Colony blots were prepared using standard procedures (Sambrook et al., 1989) and clones were
identified by hybridization with overlapping probes. Positive colonies were amplified and the identity of the clones was confirmed by Southern blot analysis of extracted plasmids.

**Cosmid library construction.** 390R DNA was partially digested with Sau3AI, and restriction fragments of 30–40 kb were excised from a 0.8% agarose gel and purified using GeneClean (Bio101) according to the manufacturer’s instructions for large DNA fragments. Purified fragments were ligated into BstII-digested pYUB178 (Pascopella et al., 1994) and ligations were packaged into lambda particles using the Gigapack III XL packaging extract (Stratagene) according to the manufacturer’s instructions. The packaged library was transduced into LB agar containing 25 \( \mu \)g kanamycin ml\(^{-1}\). Colony blots were prepared from the cosmid library and hybridized with a probe to the deletion region. Restriction digests and Southern blot analysis of cos10 showed that it had an insert of approximately 35 kb and that it contained the full deletion and flanking regions (data not shown).

**Sequence analysis and contig assembly.** Sequence analysis was performed on an ABI Prism 377 DNA Sequencer using the ABI BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems). The manufacturer’s modified protocol was used with cosmid DNA. Sequence editing was performed using ABI Editview software. GeneTool 1.0 (BioTools) was used for sequence assembly and for detection of ORFs using a minimum cutoff size of 70 aa. The BLAST programs (Altschul et al., 1997) at the NCBI website were used for sequence comparisons. Additional comparisons were conducted against completed and partial genome sequences at www.tigr.org using BLASTN 2.0 and TBLASTN (W. Gish, http://blat.wustl.edu). The M. smegmatis IS-like sequences ‘IS-A’ and ‘IS-B’ are located in contigs 3311 and 3312, respectively, of the April 8, 2002 update of the TIGR database. Putative functions for some M. tuberculosis genes were verified using the website http://genolist.pasteur.fr/Tuberculist/. The designation ISMa1 was provided by P. Siguier from the IS database registry (www.is.biotoul.fr).

**PC.** Primers S4M13RY (GAT GAC GTT GAG CAA GTG) and Ma-1S4DY (GTC GCT TAG CCA CAA CAT C) which amplify a portion of ISMa1 ORFB were used to detect the IS element. Primers S10T7S (CAT GCA GCA ATT TCA GG) and S4T7A (GAT CGG AGA CAG CTA CG) were used to amplify the 5’-end insertion site.

### RESULTS AND DISCUSSION

**Characterization of deletion R/S-D1**

*M. abscessus* 390S is a smooth colony variant that was spontaneously derived from the rough clinical isolate 390R (Byrd & Lyons, 1999). Previous analysis showed that 390S was also attenuated in vivo (Byrd & Lyons, 1999) although it had normal growth in vitro (T. F. Byrd, unpublished data). As there is little known about this species, we were interested in identifying the genetic differences between 390R and 390S and their relationship to the observed changes. Because the phenotypic changes were stable after in vivo and in vitro passage, it was hypothesized that a major mutation or mutations had occurred in 390S. To examine 390S for deletions, a modified version of the genomic subtraction method (Straus & Ausubel, 1990) was employed using 390S and 390R as driver and tester samples, respectively. After subtraction with 390S DNA, the pool of remaining fragments of 390R DNA was amplified and used to probe a Southern blot of 390R and 390S genomic DNA. The subtracted fragments showed specific hybridization to 390R DNA but not to 390S DNA (data not shown), indicating that they had been lost from the 390S genome. Several restriction fragments that were detected only in 390R were cloned and used in mapping the deletion, which was named R/S-D1. The sequence of the 390R genomic region shown in Fig. 1 was derived from these restriction fragments and from a cosmid (cos10) that contained the entire deletion and flanking regions.

Hybridization analysis of the 390S genome with a 10.5 kb SstI fragment from 390R (R-sst10.5) revealed a smaller fragment in 390S of 4.5 kb (Fig. 1 and data not shown). This fragment, S-sst4.5, contained the joined ends of the deletion and was used to identify the deletion breakpoints. Sequence comparison with 390R indicated that 390S is missing a contiguous region of 14.2 kb which contains 12 major ORFs (Fig. 1).

Examination of the sequence structure flanking the deletion breakpoints (Fig. 2) revealed an imperfect palindromic of 43 bases overlapping the left side of the deletion (‘left’ and ‘right’ refer to the orientation of the region as depicted in Fig. 1). The left breakpoint is near the axis of symmetry of the palindromic (vertical arrow, Fig. 2); the ‘A’ at the axis of symmetry is the first base of the predicted start codon (TTG) of ORF R5c which is transcribed in the opposite direction. Internal palindromes and a direct repeat (DR) are contained within the larger palindromic.

At the right breakpoint, there is a short DR, an inverted repeat (IR) and the region is notably GC-rich with two tandem Smal sites. The IR occurs between the 3’-end of ORF R17 and the 5’-end of R18 (Fig. 2 and data not shown). Interestingly, overlapping palindromic sequences and DRs have been noted as features of some regulatory regions in Gram-positive bacteria (Alland et al., 2000; Dussurget et al., 1999; Lee & Holmes, 2000; Zeng & Saxild, 1999) and therefore the complex sequence features between R5c and R6 and between R17 and R18 may be components of regulatory elements.

There is no obvious similarity between the breakpoints to suggest that the deletion occurred through homologous recombination; however, the features surrounding the breakpoints may have promoted regional instability. DRs and IRs including palindromes can be unstable in eukaryotes and prokaryotes and may lead to deletion of intervening or overlapping sequences (Glickman & Ripley, 1984; Goodchild et al., 1985; Henderson & Petes, 1993; Nasar et al., 2000). Furthermore, coincident palindromes and DRs appear to be favoured deletion sites (Glickman & Ripley, 1984).

**R/S-D1 occurs within a polymorphic region**

Several clinical strains of *M. abscessus* with different colony morphologies were examined but no direct correlation between R/S-D1 and colony morphology...
was found. Hybridization analysis showed that two rough strains, 1056 and 1475, were missing fragment R-sst3 (Fig. 3a), which lies within the deletion (Fig. 1). Further analysis (data not shown) revealed that strains 1056 and 1475 were also missing fragments R-sst10.5 and R-bam9 (Fig. 1), indicating both that this region is not required for a rough phenotype and, furthermore, that it exhibits significant polymorphism. In addition, the deletion fragment R-sst3 was present in the smooth strain 6639 as well as in two strains with an intermediate phenotype (8243 and 8988) (Fig. 3a). These three strains as well as the type strain ATCC 19977, which has a smooth phenotype, produced the same hybridization patterns as 390R using R-sst10.5 and R-bam9 as probes (data not shown). The single exception was strain 8243, which was missing about 2–2.2 kb from a 7–4 kb PvuII fragment that overlaps the R-sst10.5 probe. The reduction in size of this fragment appears to be due to the absence of the IS element ISMab1 from this strain as indicated below.

Comparative analysis of the polymorphic region

The overall G+C content for the 25 kb region shown in Fig. 1 is 60.8 mol%, consistent with the high G+C content characteristic of mycobacteria (Cole et al., 1998, 2001; Shinnick & Good, 1994). Preceding ISMab1, which is described below, 19 major ORFs (R1c–R19) were identified (Fig. 1; Table 1), with ATG, GTG and TTG start sites (Cole et al., 1998) predicted for 14, 3 and 2 ORFs, respectively. For R9, sequence comparisons with other monooxygenases indicate that a small upstream ORF (not shown) contains the original 5’-end; a frameshift appears to have occurred within a string of six adenosine residues (data not shown), possibly due to slippage of the DNA polymerase during replication (Fujii et al., 1999; Timsit, 1999). As the frameshift was present in two fragments that were cloned independently from 390R, this disruption is probably present in the genome.

Comparison of the M. abscessus ORF sequences with
GenBank and completed bacterial genome sequences provided by TIGR databases (www.tigr.com) revealed strongest similarities to the actinomycetes *M. tuberculosis* (Cole et al., 1998) and *Streptomyces coelicolor* (Redenbach et al., 1996) and to the Gram-negative environmental organisms *Pseudomonas aeruginosa* (Stover et al., 2000) and *Caulobacter crescentus* (Nierman et al., 2001) (Table 1). Table 1 also includes the results of tBLASTN searches conducted against the unannotated TIGR databases for *M. avium* and *M. smegmatis*. All of the species listed in Table 1 have GC-rich genomes, and therefore to some degree similarities may reflect preferences for amino acids with GC-rich codons as has been described for *M. tuberculosis* (Cole et al., 1998). However, some ORFs show much stronger similarities to genes from particular species and this may indicate closer functional relationships. R15, for example, has 62% identity and 75% similarity to the *P. aeruginosa* acyl-CoA dehydrogenase PA4435.

The slowly growing species *M. avium* showed strongest similarity to *M. abscessus* ORFs in this region while the rapid-grower *M. smegmatis* had the least similarity to *M. abscessus* among the mycobacteria listed. Most matches to *M. leprae* were to short sequences from pseudogenes (data not shown). *M. abscessus* ORFs within the same protein families were also compared using pairwise BLAST analysis (Tatusova & Madden, 1999). Identities ranged from 26% for the putative transcriptional regulators R3c and R5c to 30–38% among the monooxygenases (R8, R9, R11) and acyl-CoA dehydrogenases (R15, R19). These *M. abscessus* ORFs were therefore generally less similar to each other than to related genes from other species, again suggesting that interspecies matches may reflect functional similarities.

Predicted functions for most of the homologous genes suggest that much of this region is involved in fatty acid metabolism. Acyl-CoA dehydrogenases such as R15 and R19 have a putative role in β-oxidative metabolism and are related to families of *M. tuberculosis* proteins proposed to be involved in the degradation of host cell lipids for energy and metabolic precursors (Cole et al., 1998). It has been noted for *P. aeruginosa* that dehydrogenase genes are often clustered with genes encoding enzymes with related functions such as oxidoreductases and monooxygenases (Stover et al., 2000), and similar clustering is evident within this region of *M. abscessus* (Fig. 1).

Given the change in colony morphology of 390S (Byrd & Lyons, 1999), we had expected to find genes related to the synthesis of glycopeptidolipids (GPLs). Altered colony morphology in *M. avium* has been linked to changes in GPL structure (Belisle et al., 1993a, b) and *M. abscessus* is known to possess GPLs (Lopez-Marín et al., 1994), but none of the missing genes resembles those reported to be involved in GPL synthesis in other mycobacteria (Belisle et al., 1993a, b; Recht et al., 2000). Together with the data showing the absence of this region from some rough strains and its presence in some smooth strains (Fig. 3), this suggests that this region is not directly associated with colony phenotype.

**Structure of ISMab1**

ISMab1 is a novel IS element that was discovered at the 3′-end of R19 (Fig. 1). The proposed structure of ISMab1 is shown in Fig. 4; including the predicted 7 bp terminal IR (Fig. 4, IR 1), the element is 1767 bp in length. Comparative sequence analysis with IS-like elements from *M. avium* and *M. smegmatis* indicates that ISMab1 is composed of two major sections, each of which is highly related to different mycobacterial IS elements (Fig. 4).

The first of these sections begins with the sequence TGAAT (Fig. 4) and is 481 bp long. It contains ORFA and an adjacent out-of-frame sequence, which show 88% aa identity (92% similarity) to contiguous portions of the C-terminus of ORF4 from *M. avium* IS1601. IS1601 ORFs 3 and 4 were noted to be highly hom-
Table 1. Amino acid identities between *M. abscessus* ORFs and ORFs from species with GC-rich genomes

Amino acid identity scores were obtained by **BLASTP** analysis of GenBank database sequences for *M. tuberculosis*, *S. coelicolor* and *P. aeruginosa* and by **TBLASTN** analysis of TIGR database sequences for *M. avium*, *M. smegmatis* and *C. crescentus*. Comparison of **BLASTP** and **TBLASTN** scores from GenBank and TIGR, respectively, for *M. tuberculosis* and *P. aeruginosa* were equivalent within one percentage point. Identities in parentheses are over short sequences and have smallest sum probabilities (*P* values) of less than $1 \times 10^{-3}$. Dashes (–) indicate no significant matches.

<table>
<thead>
<tr>
<th>ORF</th>
<th>Length (aa)</th>
<th>Putative function*</th>
<th><em>M. avium</em></th>
<th><em>M. tuberculosis</em></th>
<th><em>M. smegmatis</em></th>
<th>Streptomyces coelicolor</th>
<th>Pseudomonas aeruginosa</th>
<th>Caulobacter crescentus</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1c</td>
<td>(352)†</td>
<td>Acetyl/propionyl carboxylase, α subunit (<em>accA</em>)</td>
<td>60</td>
<td>60</td>
<td>63</td>
<td>60</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>R2c</td>
<td>533</td>
<td>Acetyl/propionyl carboxylase, β subunit (<em>accD</em>)</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>R3c</td>
<td>224</td>
<td>TetR family transcriptional regulator (tet-r)</td>
<td>30</td>
<td>32</td>
<td>32</td>
<td>30</td>
<td>(29)</td>
<td>(30)</td>
</tr>
<tr>
<td>R4</td>
<td>289</td>
<td>Hypothetical protein</td>
<td>30</td>
<td>28</td>
<td>27</td>
<td>29</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>R5c</td>
<td>229</td>
<td>TetR family transcriptional regulator (tet-r)</td>
<td>33</td>
<td>26</td>
<td>32</td>
<td>28</td>
<td>–</td>
<td>(36)</td>
</tr>
<tr>
<td>R6</td>
<td>284</td>
<td>Oxidoreductase (oxr)</td>
<td>55</td>
<td>38</td>
<td>35</td>
<td>38</td>
<td>39</td>
<td>31</td>
</tr>
<tr>
<td>R7</td>
<td>307</td>
<td>Hypothetical protein</td>
<td>52</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>(32)</td>
<td>–</td>
</tr>
<tr>
<td>R8</td>
<td>507</td>
<td>Monoxygenase (mox)</td>
<td>61</td>
<td>44</td>
<td>38</td>
<td>40</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>R9</td>
<td>340</td>
<td>Monoxygenase (mox)</td>
<td>63</td>
<td>62</td>
<td>34</td>
<td>28</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>R10</td>
<td>321</td>
<td>Esterase/acyl hydrolase, lipase (lip)</td>
<td>39</td>
<td>38</td>
<td>31</td>
<td>45</td>
<td>41</td>
<td>29</td>
</tr>
<tr>
<td>R11</td>
<td>591</td>
<td>Monoxygenase (mox)</td>
<td>40</td>
<td>39</td>
<td>38</td>
<td>40</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td>R12</td>
<td>485</td>
<td>Cytochrome P450 (P450)</td>
<td>44</td>
<td>46</td>
<td>46</td>
<td>22</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>R13</td>
<td>270</td>
<td>Hypothetical protein</td>
<td>27</td>
<td>33</td>
<td>–</td>
<td>29</td>
<td>44</td>
<td>–</td>
</tr>
<tr>
<td>R14</td>
<td>543</td>
<td>Acyl-CoA synthase (<em>fadD</em>)</td>
<td>65</td>
<td>64</td>
<td>66</td>
<td>54</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td>R15</td>
<td>431</td>
<td>Acyl-CoA dehydrogenase (<em>fadE</em>)</td>
<td>56</td>
<td>54</td>
<td>47</td>
<td>52</td>
<td>62</td>
<td>54</td>
</tr>
<tr>
<td>R16</td>
<td>353</td>
<td>Hypothetical protein</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>R17</td>
<td>136</td>
<td>Hypothetical protein</td>
<td>45</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>R18</td>
<td>399</td>
<td>Alcohol dehydrogenase (<em>adhD</em>)</td>
<td>49</td>
<td>50</td>
<td>49</td>
<td>40</td>
<td>36</td>
<td>37</td>
</tr>
<tr>
<td>R19</td>
<td>399</td>
<td>Acyl-CoA dehydrogenase (<em>fadE</em>)</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>39</td>
<td>53</td>
<td>38</td>
</tr>
</tbody>
</table>

*Abbreviations of putative ORF functions used in Fig. 1 are in parentheses. Standard abbreviations (*accA*, *accD*, *fadD*, *fadE*, *adhD*, *fadE*) are derived from *M. tuberculosis* homologues (Cole et al., 1998).

†Length of partial ORF.
Genetic polymorphism in *Mycobacterium abscessus*

**Fig. 4.** Map of ISMab1. ORFs A, B and C are 116, 104 and 273 aa in length, respectively, and transcription direction is indicated by arrows. The speckled rectangle marks an out-of-frame sequence contiguous with ORFA. The inverted repeat IR 1, which flanks ISMab1, is italicized. The left repeat of IR 1 contains the TAG stop codon for R19. The dashed line underlines the start of the sequence matching IRe's IS999 and IS1601; the GAT preceding this is identical to the DR of IS999. DR 1 and IR 2 flanking ORFB are marked by asterisks and inverted arrowheads, respectively. IR 2 and most of DR 1 constitute IR 3 (underlined with thick bars), which flanks ORFs B and C. A thin black bar underlines a 17-base palindrome between ORFs B and C. At the bottom of the figure, bars indicate lengths and matching positions of related sequences from *M. smegmatis* and *M. avium*. Only regions having greater than 80% identity to corresponding sequences of ISMab1 are shown; percentage identities are indicated on the left. The designations 'IS-A' and 'IS-B' refer to uncharacterized IS-like sequences of *M. smegmatis* identified by BLASTN and TBLASTN analysis of the TIGR database at www.tigr.org.

At the nucleotide level, this first section has 82% identity to the region of *M. avium* IS1601 containing ORF4 (Eckstein *et al.*, 2000) and 84% identity to *M. avium* IS999 including the 28 bp IR and 3 bp DR described for that element (J.-P. Laurent & G. Cangelosi, GenBank accession no. AF232829) (Fig. 4 and data not shown). Nucleotide comparisons indicate that IS999 is 80–83% identical to the region of IS1601 containing ORFs 3 and 4 and that the 28 bp IR of IS999 is also present in IS1601 and overlaps the 15 bp IR 2 originally described for that element (Eckstein *et al.*, 2000). Interestingly, a shorter related sequence, designated IS-A in Fig. 4, is also found in *M. smegmatis* and it includes one copy of the IS999 IR.

The IS1601-related sequence in the first section ends abruptly at the left copy of DR 1, suggesting that it was disrupted by the insertion of the section containing ORFs B and C (Fig. 4). This 1.27 kb section may be an independent transposable element: it has 83% nucleotide identity over its entire length to an uncharacterized IS-like sequence from *M. smegmatis* (designated IS-B in Fig. 4), and in both *M. smegmatis* and *M. abscessus* it is flanked by IR 3 (Fig. 4 and data not shown). In addition to flanking the entire section, IR 3 is an inverted copy of the DR 1 and IR 2 sequences which flank ORFB. The IS-B sequence from *M. smegmatis* appears to be part of a larger 1.7 kb IS-like element which is less related to *M. abscessus* at the 5'-end (data not shown). TBLASTN comparisons showed that ORFs B and C have 88% identity (92% similarity) and 86% identity (91% similarity), respectively, to IS-B ORFs.

Both ORFs B and C of ISMab1 show homology to transposases. GenBank searches show that ISMab1 ORFB has 38% aa identity (59% similarity) and 36% aa identity (54% similarity) to, respectively, the transposase of IS6110 (McAdam *et al.*, 1990; Thierry *et al.*, 1990) and a putative transposase from *Saccharopolyspora* (GenBank accession no. AF045021); it is also of similar length to these ORFs. ORFC has 32–35% aa identity to a large group of transposases including the IS3 putative transposase of *E. coli* and *M. avium* IS1601 ORF4.
Distribution of ISMab1

Examination of the clinical isolates for the presence of ISMab1 showed that strains 1056 and 1475, which were missing R/S-D1 and the upstream region, are also missing this element (Fig. 3b). There was also no specific hybridization to strain 8243, which is consistent with the shorter Peull fragment identified by the R-sst10.5 probe (see above). The ISMab1 probe detected single bands in 390R, 390S, and in strains 6639 and 8988. As SstI cuts once within the IS element, the single bands in the SstI digests indicate that there is only one complete copy of the element in these strains. PCR analysis confirmed the distribution of ISMab1 and also indicated that it was present in the type strain ATCC 19977 (data not shown). In addition, sequence analysis of the PCR fragments for 8988 and the type strain showed that ISMab1 was located in the same position in these strains as in 390R.

Conclusions

ISMab1 is the first IS element to be identified in M. abscessus. It is not known, though, whether it functions as a single mobile element. IR 1 is shorter than most IRs described for IS elements (Chandler, 1998) and no DRs were found. Also, only single copies of the element have been detected. ISMab1 is remarkable, though, for its close relationship to IS elements from other environmental mycobacteria. It has been suggested that horizontal transfer of genetic elements may occur between mycobacteria and related species (Gordon et al., 1999). For example, the transposase of IS1552 from M. tuberculosis shares 80% aa identity to a transposase from Rhodococcus (Gordon et al., 1999), and ORFs in M. avium IS1601 show high amino acid similarity to transposases from other slowly growing mycobacteria (Eckstein et al., 2000). Horizontal transfer is further supported by the high degree of nucleotide identity between ISMab1 and elements from M. avium and M. smegmatis.

ISMab1 is located within a region that showed unexpected polymorphism in clinical isolates. Polymorphism associated with large deletions has also been observed in clinical isolates of M. tuberculosis (Ho et al., 2000; Vera-Cabrera et al., 1997) but, as with our M. abscessus findings, the relationship to biological properties is unknown. Although some of the altered phenotypes of M. abscessus 390S (Byrd & Lyons, 1999) may be linked to the loss of this region, this remains to be determined as we have been unable to reintroduce this region into M. abscessus 390S using cosmids. We expect that other genetic changes will be found in 390S, but genomic subtraction was useful for detecting a major genetic difference. This method, or variations of it, has been used to identify bacterial virulence factors and other unique sequences (Emmert et al., 1999; Mahairas et al., 1996; Morrow et al., 1999; Reckseidler et al., 2001; Sawada et al., 1999; Schmidt et al., 1999; Zhang et al., 2000) and in some cases has been successfully used to compare less closely related bacterial strains (Emmert et al., 1999; Mahairas et al., 1996; Morrow et al., 1999; Zhang et al., 2000). It may be a suitable approach for identifying genetic differences in other uncharacterized mycobacterial genomes.

ACKNOWLEDGEMENTS

This investigation was supported by the Department of Veterans Affairs and grants HL64548-02 and HL55776 from the National Institutes of Health. We thank Liana Ponce for excellent technical assistance and enthusiasm. We also wish to thank Blaine Beaman for providing M. abscessus variants 390R and 390S and Patricia Convilie for other M. abscessus strains. Sequencing was performed by the UNM Biochemistry Core Facility. We thank R. Fleischmann, J. Merkel and The Institute for Genomic Research (http://www.tigr.org) for access to unpublished data on M. smegmatis, M. avium and other genomes. Sequencing of M. smegmatis and M. avium at TIGR was accomplished with support from NIAID. TIGR database sequences for S. coelicolor were provided by the Sanger Centre (www.sanger.ac.uk).

REFERENCES


Genetic polymorphism in Mycobacterium abscessus


Received 21 January 2002; revised 19 April 2002; accepted 18 June 2002.