Functional analysis and annotation of the virulence plasmid pMUM001 from *Mycobacterium ulcerans*

Timothy P. Stinear,1,3 Melinda J. Pryor,2† Jessica L. Porter3 and Stewart T. Cole1

1Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Docteur Roux, 75725 Paris Cedex 15, France
2Plate-Forme 4 – Intégration et analyse génomiques, Génopole, Institut Pasteur, 28 Rue du Docteur Roux, 75725 Paris Cedex 15, France
3Australian Bacterial Pathogenesis Program, Department of Microbiology, Monash University, Clayton, 3800, Australia

The presence of a 174 kb plasmid called pMUM001 in *Mycobacterium ulcerans*, the first example of a mycobacterial plasmid encoding a virulence determinant, was recently reported. Over half of pMUM001 is devoted to six genes, three of which encode giant polyketide synthases (PKS) that produce mycolactone, an unusual cytotoxic lipid produced by *M. ulcerans*. In this present study the remaining 75 non-PKS-associated protein-coding sequences (CDS) are analysed and it is shown that pMUM001 is a low-copy-number element with a functional ori that supports replication in *Mycobacterium marinum* but not in the fast-growing mycobacteria *Mycobacterium smegmatis* and *Mycobacterium fortuitum*. Sequence analyses revealed a highly mosaic plasmid gene structure that is reminiscent of other large plasmids. Insertion sequences (IS) and fragments of IS, some previously unreported, are interspersed among functional gene clusters, such as those genes involved in plasmid replication, the synthesis of mycolactone, and a potential phosphorelay signal transduction system. Among the IS present on pMUM001 were multiple copies of the high-copy-number *M. ulcerans* elements IS2404 and IS2606. No plasmid transfer systems were identified, suggesting that trans-acting factors are required for mobilization. The results presented here provide important insights into this unusual virulence plasmid from an emerging but neglected human pathogen.

INTRODUCTION

*Mycobacterium ulcerans* causes Buruli ulcer (BU), a serious disease characterized by chronic, often painless, necrotic ulceration of subcutaneous fat in otherwise healthy individuals. BU-associated mortality is low but morbidity is very high, due to factors such as loss of limbs or limb function. There is no vaccine and no consensus on effective drug treatment. Since 1989, the number of cases has escalated throughout rural Central and West Africa, where the prevalence of BU exceeds leprosy and, in some instances, tuberculosis (Amofah et al., 2002; Debcker et al., 2004). In 1998, the World Health Organization launched the Global Buruli Ulcer Initiative with the principal aim of trying to control the spread of this debilitating disease (Anonymous, 1998).

We recently reported the presence in *M. ulcerans* of a 174 kb circular plasmid, named pMUM001 (Stinear et al., 2004). More than half of the plasmid is composed of three highly unusual polyketide synthase (PKS) genes that are required for the synthesis of the polyketide toxin mycolactone. There is a precedent for plasmid-borne genes involved in secondary metabolite biosynthesis. The pSLA2-L plasmid from *Streptomyces rochei* is rich in genes encoding type I and type II PKS clusters, and non-ribosomal peptide synthetases (Mochizuki et al., 2003). The three mycolactone PKS genes (*mlsA1*, *mlsA2* and *mlsB*) stand out for two

---

*Abbreviations*: BU, Buruli ulcer; CDS, protein-coding sequence(s); EOT, efficiency of transformation; FHA, fork-head-associated; IS, insertion sequence(s); PKS, polyketide synthase(s).

The GenBank/EMBL/DDJB accession number for the sequence of pMUM001 and its annotation reported in this paper is BX649209.

A summary of the 81 predicted CDS in *Mycobacterium ulcerans* pMUM001 is shown in Supplementary Table S1, available with the online version of this paper at http://mic.sgmjournals.org.
reasons. First, they encode some of the largest proteins ever reported (MLSA1, 1·8 MDa; MLSA2, 0·26 MDa; and MLSB, 1·2 MDa); second, there is an extreme level of nucleotide and amino acid sequence conservation (>97% nucleotide identity) among the various functional domains of the 18 modules that comprise the three synthases. This extended level of sequence conservation is unprecedented and points to the very recent evolution of this locus.

Large plasmids harbouring factors that confer key adaptations to new niche environments are a recurring theme amongst bacterial pathogens and allow the host to make large evolutionary leaps over small evolutionary time scales. Well-known examples among human bacterial pathogens include the pXO plasmids of Bacillus anthracis (Okinaka et al., 1999), the pTV plasmids of Yersinia pestis (Parkhill et al., 2001), and the pNV plasmids of Shigella (Buchrieser et al., 2000). Each of these species is highly clonal. They share a nearly identical genome structure and sequence with other species in their genera, and via plasmid acquisition, sometimes in concert with other more subtle genome changes; they have evolved adaptive advantages for new environments, adaptations that may also present a pathogenic phenotype. Recent evolution by horizontal transfer fits well with the hypothesis that M. ulcerans is a clonal derivative of Mycobacterium marinum, a hypothesis formed from multi-locus sequence typing and hybridization analyses that showed a high level of DNA sequence conservation and gene synteny between these phenotypically diverse species (Stinear et al., 2000).

Plasmids have been widely reported among many mycobacterial species (Pashley & Stoker, 2000). However, until the discovery of pMUM001, mycobacterial plasmids have never been directly linked to virulence, and the absence of plasmids among members of the Mycobacterium tuberculosis complex has led researchers to believe that plasmid-mediated lateral gene transfer is not an important factor for mycobacterial pathogenesis. Very few mycobacterial plasmids have been characterized, with complete DNA sequences available for only three mycobacterial episomes: pAL5000, a 4·8 kb circular element from Mycobacterium fortuitum (Rauzier et al., 1988); pCLP, a 23 kb linear element from Mycobacterium celatum (Le Dantec et al., 2001); and pVT2, a 12·9 kb element from Mycobacterium avium (Kirby et al., 2002). There are very few reports of functions being assigned to mycobacterial plasmids, although several studies have suggested that genes involved in different forms of hydrocarbon metabolism are plasmid borne (Coleman & Spain, 2003; Guerin & Jones, 1988; Waterhouse et al., 1991).

There are 81 predicted protein-coding sequences (CDS) on pMUM001 and we have previously described in detail the six CDS that are involved with the synthesis of mycolactone (Stinear et al., 2004). In this present study, the remaining 75 CDS are described with a functional study of the plasmid replication region.

METHODS

Bacterial strains and culture conditions. The bacterial strains used in this study were Escherichia coli strains XL2 Blue (Stratagene) and DH10B (Invitrogen), Mycobacterium ulcerans strain Ag99, Mycobacterium smegmatis mc²155, and Mycobacterium fortuitum (NCTC 10594), M. fortuitum RAM (clinical isolate, kindly provided by Aina Sievers, Victorian Infectious Diseases Reference Laboratory) and Mycobacterium marinum (M strain, kindly provided by Lalita Ramakrishnan, University of Washington). E. coli derivatives were cultured on Luria–Bertani agar plates and broth supplemented with antibiotics as required (100 µg ampicillin ml⁻¹ and 50 µg apramycin ml⁻¹). Mycobacteria were cultured in 7H9 broth and 7H10 agar (Becton Dickinson) at 30 °C with apramycin used at a concentration of 50 µg ml⁻¹ and kanamycin at 25 µg ml⁻¹.

Nucleic acid techniques. General methods for DNA manipulation were as described by Sambrook et al. (1989). For Southern hybridization experiments, DNA was extracted from mycobacteria as described by Boddinghaus et al. (1990). Approximately 1 µg DNA was digested with Nsi I and the resulting fragments were separated by agarose gel electrophoresis. The DNA was then transferred to Hybond-N membranes by alkaline capillary transfer in the presence of 0·4 M NaOH. A DNA probe based on the repA gene was prepared by PCR-mediated incorporation of Digoxigenin dUTP into the 413 bp repA amplification product. This product was obtained using the primer sequences RepA-F, 5′-CTAGAATGGACGCCTCCG-3′ (position 665–684) and RepA-R, 5′-CTGACACCGTCTCGTACGTCG-3′ (position 1077–1058). Genomic DNA from MUAgy99 was used as template. Southern hybridization conditions were as described previously (Stinear et al., 1999).

Construction of the shuttle plasmid pMUDNA2.1. As part of the M. ulcerans genome sequencing project (http://genopole.pasteur.fr/Mulc/BuruList.html), a whole-genome shotgun clone library of M. ulcerans strain Ag99 was prepared in E. coli using the vector pCDNA2.1 (Invitrogen) (Stinear et al., 2004). In this library were several E. coli shotgun clones that contained M. ulcerans sequences overlapping the predicted origin of replication (ori) of pMUM001. One such clone called mu0260E04 with an insert of 6 kb was selected for further study. To permit selection in a mycobacterial background, the apramycin resistance gene aac(3)-IV was cloned into mu0260E04 (Paget & Davies, 1996). This was achieved by PCR amplification and modification of the aac(3)-IV cassette using the oligonucleotides ApraF-Spel (5′-GGACTAGTTCGGGTCTCA-3′) and ApraR-Spel (5′-GGACTAGTCCGGG-CATTGACCGTACGAT-3′) to incorporate flanking SpeI sites (underlined). The resultant PCR product was digested with SpeI and then cloned into the unique XbaI site of mu0260E04, resulting in the hybrid vector pMUDNA2.1 (see Fig. 1). The deletion constructs pMUDNA2.1 and pMUDNA2.1-3 were prepared by double restriction-endonuclease digestion of pMUDNA2.1 with HpaI/SpeI and EcoRV/SpeI, respectively. Two mycobacteria/E. coli shuttle vectors were used as a positive controls in all transformation experiments. These were (i) the autonomously replicating vector pMV261, which is based on the pAL5000 replicon and confers resistance to kanamycin (Snapper et al., 1990) and (ii) an integrating vector pJKD8003, based on mycobacteriophage L5 (Hatfull & Sarkis, 1993) and containing the same aac(3)-IV apramycin resistance gene as used for the construction of pMUDNA2.1. Conditions for the preparation and electroporation of M. smegmatis were as previously described (Snapper et al., 1990). For electroporation of other mycobacteria, cells were harvested at room temperature from late-exponential-phase cultures, washed twice in sterile water, then once in sterile 10% glycerol (v/v) and finally resuspended in 0·01 volume of 10% glycerol (v/v). In all experiments a 200 µl aliquot of freshly prepared cells was used for each electroporation with a BTX
results and discussion

General features of pMUM001

The plasmid pMUM001 is a circular element of 174,155 bp with 81 predicted CDS and a G+C content of 62.7%. The arrangement and key features of these CDS are shown in Fig. 2 and summarized in Supplementary Table S1 (available as supplementary data with the online version of this paper at http://mic.sgmjournals.org). Six genes were predicted to be involved in mycolactone biosynthesis and they account for 60% of the total plasmid sequence. These genes have been described elsewhere, but they encode three type I modular PKS (MUP032, MUP039, MUP040), a type II thioesterase (MUP038), a FabH-like type III ketosynthase (MUP045) and a P450 hydroxylase (MUP053) (Stinear et al., 2004). There were 26 copies of various IS or fragments of IS, including 14 previously unreported elements. The presence of orthologous genes in other bacteria permitted the identification of CDS involved in plasmid functions such as replication and partitioning, as well as a potential regulatory cluster that includes, somewhat unusually for a plasmid, a serine-threonine protein kinase (STPK) gene. There were no CDS encoding plasmid transfer functions. Eleven CDS had features suggesting that they encode membrane-associated proteins, but other than the STPK, none had identifiable functions. There were 26 CDS encoding hypothetical proteins; 11 of these had no homology with other sequences in the public databases and 15 were classified as conserved hypothetical proteins because they had some homology to hypothetical proteins in M. tuberculosis (8), Mycobacterium leprae (1), Rhizobium loti (1), Agrobacterium tumafaciens (1), bacteriophage T7 (1), Streptococcus coelicolor (2) and Streptococcus avermitilis (1). The overall structure of pMUM001 is highly mosaic, with discrete gene cassettes interspersed with IS. Plasmid copy number was estimated to be 1-9 copies per cell, based on the ratio of the mean number of shotgun sequences per 1 kb of pMUM001 relative to the chromosome from the M. ulcerans genome assembly database (http://genopole.pasteur.fr/Mulc/BuruList.html).

Origin of replication

The repA gene, encoding the 368 aa RepA, is responsible for the initiation of replication and was readily identified by sequence comparisons, sharing 83.6% amino acid identity in 366 aa with RepA from the M. fortuitum plasmid pAZ38 (Gavigan et al., 1997) and 55.6% amino acid identity with RepA from the M. avium plasmid pVT2 (Kirby et al., 2002). There was identity to the predicted RepA proteins from many mycobacterial plasmids, with the exception of pAL5000, which appears unrelated. There was also significant identity with the RepA protein from the Rhodococcus plasmid pSOX (Denis-Larose et al., 1998).

Analysis of the sequence −1 to −600 bp of repA revealed several features suggestive of an iteron-containing origin of replication. Iterons are direct repeat sequences that bind...
Fig. 2. Linear map of pMUM001. The positions of the 81 predicted protein-coding DNA sequences (CDS) are indicated as different-coloured blocks labelled sequentially as MUP001 (repA) through to MUP081. Forward- and reverse-strand CDS are shown above and below the black line, respectively, and the colours represent different functional classifications (red, replication; light blue, regulation; light green, hypothetical protein; dark green, cell wall and cell processes; orange, conserved hypothetical protein; cyan, insertion sequence elements; yellow, intermediate metabolism; grey, lipid metabolism). The black arrows indicate the region cloned into pCDNA2.1 to produce the shuttle vector pMUDNA2.1. The regions covered by the light-grey-shaded boxes indicate 8 kb of identical nucleotide sequence encompassing the start of the mycolactone PKS genes mlmA1 and mlmB. The scale is given in bp and each minor division represents 1000 bp.
RepA and exert control over plasmid replication. A single pair of 16 bp iterons was identified in the region 180–550 bp upstream of the repA initiation codon (Fig. 3). The spacing between iterons is usually a multiple of 11, a distance reflecting the helical periodicity of dsDNA, implying that the binding sites for RepA are on the same face of the DNA (del Solar et al., 1998). The spacing for the iteron identified in pMUM001 is 143 bp, a multiple of 11. Low plasmid copy number is a characteristic of iteron plasmids. It has been proposed that as copy number increases, the RepA molecules bound to the iteron of one origin begin to interact with similar complexes generated on other origins, generating a so-called ‘hand-cuffed’ state that suppresses replication (del Solar et al., 1998). Other features commonly associated with iteron-containing replicons are multiple inverted repeats of partial-iteron sequences. These are generally situated immediately upstream of the repA promoter region (del Solar et al., 1998). In pMUM001, the situation appears somewhat different. A single 12 bp partial inverted repeat of the iteron sequence was detected in the region between the iterons. No obvious promoter elements were found in these upstream sequences; however, the region −1 to −261 bp of the repA ATG shares a very high identity with the same region in pJAZ38 (75% nucleotide identity) and a 69 bp subsection of this region is highly conserved among several mycobacterial plasmids (Picardeau et al., 2000). The 16 bp iteron sequences are boxed and the partial inverted repeat of the iteron is shaded grey.

Several strategies have evolved to ensure maintenance of low-copy-number plasmids within a bacterial population. Killing of plasmid-free segregants by a plasmid-encoded toxin/antitoxin locus is one approach and has been reported for the linear mycobacterial plasmid pCLP (Le Dantec et al., 2001). Another widely employed maintenance system uses active partitioning and distribution of plasmid copies to daughter cells. While no candidate ‘killing’ locus was found, approximately 2 kb downstream of repA is parA, a gene encoding a 326 aa putative chromosome partitioning protein. Par loci generally comprise two proteins (ParA and ParB) that form a nucleoprotein partition complex which binds a cis-acting centromere site (ParS) (Gerdes et al., 2000). Par proteins act independently of the replication apparatus and are involved in active segregation of plasmids and chromosomes before cell division. Together with host factors, Par proteins are required to direct and position newly replicated plasmids. ParA contains an ATPase domain and is specifically stimulated by ParB. Par loci share common features among different bacteria, but they are quite heterogeneous and appear to be acquired to stabilize heterologous replicons (Gerdes et al., 2000). The ParA of pMUM001 contains a Walker-type ATPase A-box motif (KGGVGK) and is most similar to ParA from non-mycobacterial species such as Arthrobacter nicotinovorans (35–1% identity in 308 aa), but it also shares some limited homology with ParA from other mycobacteria such as ParA from pCLP (48% in 41 aa). Par loci are generally arranged as an operon. In pMUM001, a candidate parB (MUP004) was identified immediately downstream of parA. MUP004 encodes a predicted 204 aa protein. BLASTP and PSI-BLAST database searches revealed no similarity to known ParB proteins, or any other proteins. A syntenous Par locus is present in pVT2 from M. avium, with a gene encoding a hypothetical protein immediately downstream of a parA orthologue. Heterogeneity among ParB proteins has been reported (Gerdes et al., 2000). A candidate ParB sequence was not identified on pMUM001; however, three direct repeats of the 18 bp sequence GGTGCTGCTGGGGCGG-TG were discovered in the non-coding sequence upstream of parA between positions 5314 and 5410. Iteron-like sequences such as these have been reported in the promoter region for Par operons and can act as binding sites for ParB (Moller-Jensen et al., 2000).
To test the hypothesis that this region contains a functional replication origin, a small-insert (3–6 kb) E. coli shotgun library of pMUM001 was screened and a clone with a 6 kb fragment was selected. This fragment spanned the region from position 172 467 to 4190 that encompassed the 5 end of MUP081 and the putative ori, repA and parA genes. The clone, named pmu0260E04, was modified by the insertion of aac(3)-IV, a gene conferring resistance to apramycin and thus permitting selection in a mycobacterial background (Paget & Davies, 1996). This construct, named pMUDNA2.1 (Fig. 1), was used to try and transform M. smegmatis, M. fortuitum and M. marinum. Transformants were only obtained for M. marinum (Table 1). The efficiency of transformation (EOT) of M. marinum transformed with pMUDNA2.1 was close to the EOT obtained using the pAL5000-based shuttle plasmid pMV261 (Table 1). Deletion studies were then conducted to try and define the minimum region of pMUM001 required for replication. Two deletion constructs of pMUDNA2.1 were made. The first construct (pMUDNA2.1-1) was made by removing the 1300 bp region between the unique SpeI and Hpal sites. This region spans the entire parA gene and 372 bp of upstream sequence (Fig. 1). The second construct (pMUDNA2.1-3) was made by deleting the 2610 bp region between the unique SpeI and EcoRV sites. This 2610 bp segment spanned all of the pMUDNA2.1-1 deletion plus the predicted CDS MUP003 and MUP004. Both these constructs were capable of transformation of M. marinum with an EOT equivalent to that obtained from pMUDNA2.1 (data not shown). This result suggests that the 3327 bp of pMUM001 sequence spanning MUP002, repA, oriM and the partial sequence of MUP081 is sufficient to support replication.

Southern hybridization and back-transformation from M. marinum to E. coli were used to prove that all three constructs were operating as autonomous, episomal elements in M. marinum. The results are summarized in Fig. 4 and confirm that each pMUM001-derived construct supported replication in M. marinum from its own ori.

To test the stability of pMUDNA2.1, a late-exponential-phase culture of M. marinum harbouring pMUDNA2.1 grown in the presence of apramycin was shifted to media without apramycin and then monitored at successive time points by determining plate counts on media with and without the antibiotic. The results of this experiment are summarized in Fig. 5 and show that pMUDNA2.1 was not stably maintained and was rapidly lost from a population of cells in the absence of antibiotic selection. The same results were obtained with the two deletion constructs (data not shown), suggesting either that the putative par locus from pMUM001 is not functional in M. marinum or that additional pMUM001 sequences are required for plasmid maintenance that are outside the 6 kb fragment.

### Table 1. Transformation of mycobacteria with pMUDNA2.1

<table>
<thead>
<tr>
<th>Strain details</th>
<th>EOT [no. transformants (µg plasmid DNA)$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pMUDNA2.1</td>
</tr>
<tr>
<td>M. marinum (M strain)</td>
<td>$1\cdot3 \times 10^9 \pm 0\cdot2$</td>
</tr>
<tr>
<td>M. smegmatis (mc²155)</td>
<td>0</td>
</tr>
<tr>
<td>M. fortuitum (NCTC 10394)</td>
<td>0</td>
</tr>
<tr>
<td>M. fortuitum (RAM)</td>
<td>0</td>
</tr>
</tbody>
</table>

ND, Not done.
used to construct pMUDNA2.1. One such region may be the 18 bp iteron sequences, proposed above as a candidate parS site. These repeats are 1·4 kb upstream of parA and 1·2 kb outside the region of pMUM001 cloned in pMUDNA2.1.


**Regulatory elements**

Between MUP006 and MUP021, in a region without IS disruption, are CDS encoding potential regulatory and membrane-associated proteins (Fig. 2). MUP011 is clearly an STPK with a conserved catalytic kinase domain. It is most closely related to PknJ from *M. tuberculosis* (43% amino acid identity in 523 aa). STPK are transmembrane signal transduction proteins, and in prokaryotes they are known to be involved in the regulation of many cellular processes, including virulence, stress responses and cell wall biogenesis (Boitel et al., 2003). Approximately 3·5 kb downstream of MUP011 is a CDS (MUP018) that may code for a phosphorylation substrate for MUP011. MUP018 encodes a hypothetical transmembrane protein that contains an N-terminal fork-head-associated (FHA) domain, a C-terminal domain with weak similarity to a 2-keto-3-deoxygluconate permease (an enzyme used by bacterial plant pathogens to transport degraded pectin products into the cell), and between these two regions, a helix–turn–helix motif. FHA domains are phosphopeptide recognition sequences that promote phosphorylation-dependent protein–protein interactions (Durocher & Jackson, 2002). The study of FHA-containing proteins in bacteria is a nascent field, but a recent report has suggested that the dual FHA domains of an ABC transporter (Rv1747) in *M. tuberculosis* represent the cognate partner for the STPK PknF (Molle et al., 2004). One possibility is that, given its overall structure, MUP018 may also be involved in substrate transport into the cell. The final CDS in this cluster is MUP021, an orthologue of the putative transcriptional regulator WhiB6 in *M. tuberculosis*. In *M. tuberculosis*, immediately upstream of WhiB6 is the divergently transcribed conserved hypothetical gene Rv3863. A similar linkage is also seen in pMUM001, as MUP018 is an orthologue of Rv3863. The significance of these associations remains to be tested, but the continuity of this region, free of IS disruption, strengthens the idea that these genes may fulfil an important regulatory role. It is also worth noting that, like pMUM001, several mycobacterial phages display a mosaic organization and that one of them, Bxz1, carries an STPK gene (Pedulla et al., 2003). Altered signal transduction pathways may arise from horizontal acquisition of STPK genes by mycobacteria.

**Membrane-associated proteins**

Significant amounts of mycolactone can be detected in an *M. ulcerans* culture supernatant, suggesting that there may be active transport of the molecule out of the bacterial cell. Lipid export in other mycobacteria is known to involve large transmembrane proteins, such as the MmpLs, which in *M. tuberculosis* are found clustered with genes involved in lipid metabolism, including type I PKS (Tekaiia et al., 1999). Analysis of the pMUM001 sequence revealed no mmpL-like genes. Ten hypothetical proteins that may play a role in mycolactone export were identified, as they contained features such as membrane-spanning domains, signal sequences, lipoprotein attachment sites or hydrophobic N-terminal sequences (see Supplementary Table S1). Whatever their function, the 10 CDS listed in Supplementary Table S1 may encode surface-exposed antigens and, given the absence of orthologues in available databases, they may be interesting candidates for testing as *M. ulcerans*-specific antigens with potential application in serodiagnosis or vaccine development.

**Insertion sequences**

Based on the presence of characteristic transposase sequences, 26 copies of various IS or IS-like sequences were identified on pMUM001. They are distributed throughout pMUM001 and interspersed among defined functional CDS clusters (e.g. replication, maintenance, toxin production). Twelve IS were copies of the known *M. ulcerans* elements, IS2404 and IS2606 (Stinear et al., 1999), and the remaining 14 were previously unreported (Fig. 2, Table 2). Transposase sequence comparisons revealed related proteins in other actinomycetes and in more distant genera. There were three copies of a putative IS belonging to the IS4 family (MUP025, MUP028, MUP037).
However, each copy of this element had been disrupted by insertion of another element (IS2404 for MUP028, and IS2606 for MUP025 and MUP037), thus precluding delineation of this IS. The sequences bounded by the ends of the loading module domains of mlsA1 and mlsB and extending through to MUP035 and MUP043 represent 8 kb of identical nucleotide sequence (Fig. 1). This region also contains three different pairs of putative IS (MUP033 and MUP041, MUP034 and MUP042, MUP035 and MUP043). Since the flanking sequences for these IS are also identical, the IS boundaries could not be determined. There is remarkably little distance (90 bp) between the initiation codons of the PKS genes mlsB and mlsA1 and the transposase genes (MUP033 and MUP041) that precede each of them. This raises the possibility that the promoter region for the two PKS genes lies within these IS elements.

MUP051, MUP052 and IS2606 share a very high amino acid identity with transposases found on the 101 kb plasmid pKB1 from the rubber-degrading actinomycete Gordonia westfalica (Broker et al., 2004). The direct significance of this relationship is not known, but it does serve to reinforce the idea that there is considerable genetic exchange between diverse populations of actinomycetes.

BLASTN analysis of the 26 IS sequences against the draft M. ulcerans genome sequence did not reveal any paralogous elements on the M. ulcerans chromosome, with the exception of IS2404 and IS2606.

Table 2. Summary of the 26 putative IS elements detected on pMUM001

<table>
<thead>
<tr>
<th>IS name or MUP CDS no.</th>
<th>Copy no.</th>
<th>Transposase length (aa)</th>
<th>IS family</th>
<th>High-scoring transposase hit [aa identity in overlap (%), size of overlap (aa)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS2404a</td>
<td>1</td>
<td>348</td>
<td>ISAI</td>
<td>Transposase [46, 338] Rhodococcus erythropolis</td>
</tr>
<tr>
<td>IS2404b*</td>
<td>3</td>
<td>348</td>
<td>ISAI</td>
<td>Transposase [67, 414] Gordonia westfalica</td>
</tr>
<tr>
<td>IS2606a</td>
<td>7</td>
<td>444</td>
<td>IS256</td>
<td>Transposase [44, 561] Magnetococcus sp. MC-1</td>
</tr>
<tr>
<td>IS2606b†</td>
<td>1</td>
<td>173 + 302</td>
<td>IS256</td>
<td>Transposase [42, 269] Thermoaerobacter tengcongensis</td>
</tr>
<tr>
<td>025‡, 028‡, 037‡</td>
<td>3</td>
<td>579</td>
<td>IS4</td>
<td>Transposase [54, 71] Streptomyces avermitilis</td>
</tr>
<tr>
<td>027</td>
<td>1</td>
<td>272</td>
<td>IS110</td>
<td>Transposase [68, 94] Gordonia westfalica</td>
</tr>
<tr>
<td>033, 041</td>
<td>2</td>
<td>124</td>
<td>IS6</td>
<td>Transposase [52, 174] Streptomyces avermitilis</td>
</tr>
<tr>
<td>034, 042</td>
<td>2</td>
<td>179</td>
<td>IS3</td>
<td>Transposase [55, 34] Xanthomonas campestris</td>
</tr>
<tr>
<td>035‡, 043</td>
<td>2</td>
<td>351</td>
<td>IS110</td>
<td>Transposase [44, 92] Streptomyces lividans</td>
</tr>
<tr>
<td>044‡</td>
<td>1</td>
<td>46</td>
<td>IS3</td>
<td>Transposase [87, 93] Gordonia westfalica</td>
</tr>
<tr>
<td>049</td>
<td>1</td>
<td>129</td>
<td>IS3</td>
<td>Transposase [66, 277] Gordonia westfalica</td>
</tr>
<tr>
<td>051‡</td>
<td>1</td>
<td>93</td>
<td>IS3</td>
<td>Transposase [1438, 1379] Xanthomonas campestris</td>
</tr>
<tr>
<td>052</td>
<td>1</td>
<td>277</td>
<td>IS3</td>
<td>Transposase [46, 338] Rhodococcus erythropolis</td>
</tr>
</tbody>
</table>

*Contains an internal stop codon.
†Contains a frame-shift mutation.
‡Truncated.

Both forms of IS2404 are predicted to encode a single transposase of 348 aa. IS2404b is the same as IS2404a in all respects, except that it contains an internal stop codon, resulting in predicted transposase fragments of 234 aa and 113 aa. However, there is probably read-through of this stop codon, as there are three copies of IS2404b, suggesting that the element may still be capable of transposition.

Eight copies of the element IS2606 were also identified. It, too, was found to be larger than the 1406 bp initially reported (Stinear et al., 1999). It has a size of 1438 bp, with 31 bp imperfect inverted repeats, producing target-site duplications of 7 bp and encoding a putative transposase of 444 aa. One copy contained a frame-shift mutation (MUP060 and MUP061) within the transposase region.

Concluding comments

Megaplasmids (50–500 kb) are widespread across many bacterial genera and represent a major resource for lateral gene transfer within microbial communities. Genetic mosaicism has emerged as a common structural theme for these elements (Molbak et al., 2003) and is particularly evident in pMUM001, which is similar in size to certain mycobacteriophages such as Bxz1 that also display a mosaic
arrangement (Pedulla et al., 2003). In part, the mosaic arrangement may stem from the large number of IS elements carried by pMUM001. These are present in both direct and inverted orientations, and recombination between these repeats is expected to contribute to variation in both plasmid size and function. An example of this has already been reported (Stinear et al., 2004).

In this study, we have identified the Rep locus, required for replication, and demonstrated functionality. The resultant shuttle plasmid, pMUDNA2.1, should be useful for genetic analysis of both M. marinum and M. ulcerans. Furthermore, the replicon of pMUM001 may even facilitate the production of mycolactone in a heterologous host although, given the large size of the PKS genes and their repetitive nature, this may be technically challenging. Nonetheless, heterologous expression would represent an important step forward in the functional analysis of mycolactone biosynthesis and even open new prophylactic avenues for preventing BU.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of the Génopole programme, the World Health Organization and the Association Française Raoul Follereau.

REFERENCES


Okinaka, R. T., Cloud, K., Hampton, O. & 12 other authors (1999). Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. J Bacteriol 181, 6509–6515.


