Evolution of Mycobacterium ulcerans and other mycolactone-producing mycobacteria from a common Mycobacterium marinum progenitor

MJ Yip
JL Porter
JAM Fyfe
CJ Lavender
F Portaels

See next page for additional authors

Follow this and additional works at: https://scholarworks.wm.edu/vimsarticles

Part of the Aquaculture and Fisheries Commons

Recommended Citation
Yip, MJ; Porter, JL; Fyfe, JAM; Lavender, CJ; Portaels, F; Rhodes, MW; Kator, HI; and Et al., "Evolution of Mycobacterium ulcerans and other mycolactone-producing mycobacteria from a common Mycobacterium marinum progenitor" (2007). VIMS Articles. 1017.
https://scholarworks.wm.edu/vimsarticles/1017

This Article is brought to you for free and open access by W&M ScholarWorks. It has been accepted for inclusion in VIMS Articles by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
Evolution of *Mycobacterium ulcerans* and Other Mycolactone-Producing Mycobacteria from a Common *Mycobacterium marinum* Progenitor⁷,†

Marcus J. Yip,¹ Jessica L. Porter,¹ Janet A. M. Fyfe,² Caroline J. Lavender,² Françoise Portaels,³ Martha Rhodes,⁴ Howard Kator,⁴ Angelo Colorni,⁵ Grant A. Jenkin,¹ and Tim Stinear¹*

Department of Microbiology, Monash University, Wellington Road, Clayton 3800, Australia;² Victorian Infectious Diseases Reference Laboratory, 10 Wreckyn Street, North Melbourne 3051, Australia;³ Institute of Tropical Medicine, Antwerp 2000, Belgium;⁴ Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia 23062;⁵ and Israel Oceanographic and Limnological Research Ltd., National Centre for Mariculture, Eilat 88112, Israel

Received 12 September 2006/Accepted 8 December 2006

It had been assumed that production of the cytotoxic polyketide mycolactone was strictly associated with *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. However, a recent study has uncovered a broader distribution of mycolactone-producing mycobacteria (MPM) that includes mycobacteria cultured from diseased fish and frogs in the United States and from diseased fish in the Red and Mediterranean Seas. All of these mycobacteria contain versions of the *M. ulcerans* pMUM plasmid, produce mycolactones, and show a high degree of genetic relatedness to both *M. ulcerans* and *Mycobacterium marinum*. Here, we show by multiple genetic methods, including multilocus sequence analysis and DNA-DNA hybridization, that all MPM have evolved from a common *M. marinum* progenitor to form a genetically cohesive group among a more diverse assemblage of *M. marinum* strains. Like *M. ulcerans*, the fish and frog MPM show multiple copies of the insertion sequence IS2404. Comparisons of pMUM and chromosomal gene sequences demonstrate that plasmid acquisition and the subsequent ability to produce mycolactone were probably the key drivers of speciation. Ongoing evolution among MPM has since produced at least two genetically distinct ecotypes that can be broadly divided into those typically causing disease in ectotherms (but also having a high zoonotic potential) and those causing disease in endotherms, such as humans.

*Mycobacterium ulcerans* is a significant human pathogen and the causative agent of Buruli ulcer (BU), a disease with severe morbidity characterized by chronic skin ulceration and extensive necrosis of subcutaneous fat (40). Cases of BU have been reported in many parts of the world, with the greatest burden occurring in West and Central Africa (16). However, no cases of BU have ever been reported in the United States or Europe. *M. ulcerans* strains are characterized by the presence of a large circular virulence plasmid called pMUM (31, 33). This plasmid harbors three large genes (*mlsA1, mlsA2*, and *mlsB*) encoding polyketide synthases that are required for the synthesis of the lipid toxin mycolactone, which is the primary virulence factor for the pathogen (33). Comparisons of multiple plasmid and chromosomal genes among 10 *M. ulcerans* clinical isolates from diverse origins have suggested that plasmid acquisition was probably the key event that marked and permitted the recent emergence of *M. ulcerans* from a common *Mycobacterium marinum* progenitor (31). *M. marinum* is phenotypically distinct from *M. ulcerans*, producing photochromogenic pigments and generally growing more quickly. *M. marinum* causes granulomatous lesions in fish and other ectotherms and can also cause granulomatous skin lesions in humans. Comparisons between the 5.8-Mb genome of the *M. ulcerans* African epidemic strain Agy99 and the 6.6-Mb genome of *M. marinum* strain “M” confirmed this hypothesis and showed that *M. ulcerans* has recently passed through an evolutionary bottleneck, evolving from the generalist *M. marinum* to become a specialist bacterium, adapted to a more restricted (possibly host-specific) environment (32, 34; http://genolist.pasteur.fr/BuruList/). Like other niche-adapted pathogens, such as *Yersinia pestis* and *Bordetella pertussis*, *M. ulcerans* has undergone extensive gene loss due to DNA deletions, DNA rearrangements, and pseudogene formation. Many of these changes have been mediated by some of the 213 copies of IS2404 and 91 copies of IS2606 (34). Neither of these insertion sequence elements (ISE) is present in *M. marinum* (32).

It was assumed until recently that mycolactone production was restricted to *M. ulcerans*. However, in 2004, a previously unreported mycobacterium that contained both IS2404 and IS2606 was recovered from cases of an unusual, lethal edematous disease in laboratory-housed *Xenopus laevis* and *Xenopus tropicalis* frogs (37). Limited sequence comparisons of *hsp65*, the 16S rRNA gene, the 16S-23S rRNA gene internal transcribed sequence spacer, and *rpoB* gene fragments showed that this mycobacterium shared greater than 98% nucleotide identity with *M. ulcerans* and *M. marinum* (37). In a subsequent investigation, the frog mycobacterium was shown to contain a version of the *M. ulcerans* pMUM plasmid and to produce a new mycolactone, mycolactone E (13, 19). It was also given the unofficial epithet *Mycobacterium liflandii* (19). In 2005, another new mycobacterium, isolated during an epizootic of mycobac-
teriosis from diseased striped bass (*Morone saxatilis*) in Chesapeake Bay, was also shown to contain IS2606 and IS2022 and another fish pathogen, *M. ulcerans* (26). One of these strains (DL240490) shared identical gene sequences, and they too produced mycolactone F (26). One of these strains (DL240490) shared identical gene sequences, and they too produced mycolactone F (26).

18. M. ulcerans strains and 22 *M. marinum* strains have been described in detail in a previous publication (32). Ten of these strains, each representing a unique multilocus sequence analysis (MLSA) sequence type (ST), are listed in Table 1, because they were used in additional analyses that included DDH, Southern hybridization, and pseudogene and deletion analyses. Table 1 also lists the details of five novel *M. marinum* and seven other mycobacteria that were examined in this study. The mycobacteria were cultivated in Middlebrook 7H9 medium or on egg yolk agar slopes at 30°C (32).

**MATERIALS AND METHODS**

**Bacterial strains and culture conditions.** Among the isolates used in this study, 18 *M. ulcerans* strains and 22 *M. marinum* strains have been described in detail in a previous publication (32). Ten of these strains, each representing a unique multilocus sequence analysis (MLSA) sequence type (ST), are listed in Table 1, because they were used in additional analyses that included DDH, Southern hybridization, and pseudogene and deletion analyses. Table 1 also lists the details of five novel *M. marinum* and seven other mycobacteria that were examined in this study. The mycobacteria were cultivated in Middlebrook 7H9 medium or on egg yolk agar slopes at 30°C (32).

**General DNA methods and oligonucleotides used in this study.** Genomic DNA for DDH was extracted from mycobacteria harvested from egg yolk agar slopes. Briefly, 100 mg (wet weight) cells was resuspended in 2 ml of lysis buffer (15% [wt/vol] sucrose, 0.05 M Tris, pH 8.5, 0.05 M EDTA, 1 mg/ml lysozyme, and 2 mg/ml RNase) and incubated at 37°C for 1 hour. The volume was then increased to 5 ml by the addition of 2 ml of 10% (wt/vol) sodium dodecyl sulfate and 1 ml of 1× TE (10 mM Tris, pH 8.5, 1 mM EDTA) containing 5 mg of proteinase K. Following overnight incubation at 37°C, the DNA was separated by successive rounds of extraction with phenol-chloroform (1:1) and chloroform and then precipitated with a 2× volume of absolute ethanol in the presence of 4% NaCl at -20°C overnight. The resulting pellet was resuspended in 400 μl of 1× TE, and the extraction process was repeated. The final DNA pellet was washed twice in 70% (vol/vol) ethanol and resuspended in 100 μl of 1× TE.

Other methods for PCR, pulsed-field gel electrophoresis, and Southern hybridization were also performed as described previously (32). For the MLSA, sequencing of a 357-bp fragment of the glcB gene was achieved by PCR amplification using glcB-F (5′-GGACTTCACTACCATACCC-3′) and glcB-R (5′-GACTCAAGAGTACGGTCTCT-3′) (Fig. 1). Other primer pairs used for MLSA have been described before (32). The oligonucleotides used for PCR amplification and sequencing of potential pseudogenes are listed in Table S1 in the supplemental material. Genome comparisons between *M. ulcerans* Ag99 and *M. marinum* M identified 157 specific DNA differences between the species, referred to as *M. ulcerans* regions of difference (MURDs) (34). Most of the MURDs are deletions of DNA from *M. ulcerans*. To investigate the distribution of these deletions (MURD12, -54, and -152) among the strains in this study, we...
performed PCR with three oligonucleotides for each MURD. These primers were designed based on genome comparisons between *M. ulcerans* Agg99 and *M. marinum* M (34), where P1 hybridized to conserved sequences in both strains flanking the region deleted in *M. ulcerans*, P2 hybridized to sequence adjacent to P1 in *M. ulcerans*, and P3 hybridized to sequence adjacent to P1 in *M. marinum*. The oligonucleotide sets used for the detection of MURD12, -54, and -152 are listed in Table S2 in the supplemental material.

DNA-DNA hybridization. For each isolate, at least three biological repeats were performed, in which 100 ng of single-stranded genomic DNA was spotted in duplicate onto positively charged nylon membranes using a 96-well vacuum dot bloter (Schleicher & Schuell). DNA was cross-linked to the membranes with UV irradiation (1,200 mJ), washed in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), and probed immediately. Genomic-DNA probes were prepared by incorporating digoxigenin-dUTP by random prime labeling, following the manufacturer’s instructions (Roche). Blots were hybridized and detected as described previously (15). Blot images were digitized and then analyzed with ImageJ (27). A standard spot area was used to measure the integrated pixel density. The relative binding ratio (RBR), defined as 100% for a blot hybridizing to a probe derived from autologous DNA, was calculated by dividing the mean integrated pixel density by the mean integrated pixel density for each strain by the mean pixel density of the autologous reaction mixture and multiplying by 100. The mean was calculated from at least three independent hybridization reactions. Analysis of variance in conjunction with Dunnett’s test was used to determine if differences between RBRs were significant (*P* < 0.05). Where variance was not homogeneous (as determined by Levene’s statistic), Dunnett’s T3 post hoc test was employed. All calculations were performed using SPSS (version 14.0.1; LEAD Technologies, Inc.).

MLSA. The methods used for MLSA were as previously described (32), with the addition of an eighth locus, a 357-bp fragment of the *gcb* gene that encodes malate synthase, the enzyme involved in the second step of the glyoxylate bypass.

The distribution and nucleotide positions of the eight protein-coding DNA sequences (CDS) on the 5.6-Mbp chromosome of *M. ulcerans* Agg99 and the 6.6-Mbp chromosome of *M. marinum* M are shown in Fig. 1. Double-stranded DNA sequences were determined from the PCR products of each locus using dye terminator sequencing as described previously (32). Sequence assembly was managed with Gap4 (3). An allelic profile was constructed for each strain, and the unique combination of alleles defined a sequence type (Table 1; see Table S3 in the supplemental material). The DNA sequences of the alleles are listed in Table S3 in the supplemental material. The allelic profiles for some isolates differed at more than three of the eight loci, so phylogeny was inferred using a distance method rather than a pairwise comparison of the allelic profiles (29); the latter approach is referred to as multilocus sequence typing (10). For each ST, the nucleotide sequences of the alleles were concatenated in frame to generate an artificial CDS. Using the same approach, all MPM were subjected to further MLSA analysis by determining the partial DNA sequences of four pMUM plasmid CDS (repA, parU, MUP045, and the mls loading domain) as described previously (31).

Phylogenetic methods. Phylogenetic analysis was performed with MEGA software version 2.1.2 (18) and Splits Tree version 4 (14). “P” distances were used throughout, as the overall level of sequence divergence was small. Values for synonymous (dS) and nonsynonymous mutation frequencies were calculated by Nei and Gojobori’s method (20), and standard errors (SE) for the means of these values were estimated by the method of Nei and Jin (21) using dSdNq (7).

VNTR analysis. The primers and conditions used for PCR amplification of the nine variable-number tandem repeat (VNTR) loci 1, 4, 6, 8, 9, 14, 15, 18, and 19 were as described previously (2). Amplicon sizes were estimated following electrophoresis of the PCR products through a 2% agarose gel and comparison with a 100-bp DNA ladder (Promega).

Nucleotide sequence accession numbers. The DNA sequences of the alleles have been assigned GenBank accession numbers DQ985290 to DQ985355.

RESULTS

DNA-DNA hybridization. Whole-genome probes derived from the five novel MPM strains with MLSA ST9, -10, -11, -12, and -14 (see below); three *M. ulcerans* strains (ST17, -19, and -20); and a representative non-MPM *M. marinum* strain (ST22) were hybridized to each other in a reciprocal fashion. The five novel MPM strains were then tested against genomic DNAs from three additional *M. marinum* strains (ST1, -2, and -5). The results are summarized in Table 2, and an example is

<table>
<thead>
<tr>
<th>DNA source</th>
<th>ST17</th>
<th>ST20</th>
<th>ST19</th>
<th>ST1</th>
<th>ST10</th>
<th>ST9</th>
<th>ST22</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. ulcerans</em> ST17</td>
<td>100</td>
<td>100</td>
<td>ND^a</td>
<td>100</td>
<td>100</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. ulcerans</em> ST20</td>
<td>97 (88–117)</td>
<td>100</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. ulcerans</em> ST19</td>
<td>94 (84–101)</td>
<td>104 (92–115)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><em>M. marinum</em> ST14</td>
<td>100 (99–109)</td>
<td>107 (95–105)</td>
<td>99 (94–106)</td>
<td>100</td>
<td>100</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. pseudoshottsii</em> ST14</td>
<td>100 (99–109)</td>
<td>107 (95–105)</td>
<td>99 (94–106)</td>
<td>100</td>
<td>100</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. marinum</em> ST12</td>
<td>98 (80–110)</td>
<td>102 (96–115)</td>
<td>101 (95–107)</td>
<td>102 (94–112)</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><em>M. marinum</em> ST1</td>
<td>98 (86–103)</td>
<td>100 (72–113)</td>
<td>88 (71–95)</td>
<td>98 (86–107)</td>
<td>102 (83–110)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><em>M. marinum</em> ST10</td>
<td>103 (93–108)</td>
<td>100 (89–107)</td>
<td>95 (82–100)</td>
<td>100 (79–107)</td>
<td>100 (82–105)</td>
<td>96 (88–105)</td>
<td>100</td>
</tr>
<tr>
<td><em>M. marinum</em> ST9</td>
<td>91 (80–105)</td>
<td>101 (90–109)</td>
<td>88 (79–98)</td>
<td>96 (79–104)</td>
<td>99 (88–108)</td>
<td>100 (90–116)</td>
<td>98 (88–112)</td>
</tr>
<tr>
<td><em>M. marinum</em> ST1</td>
<td>ND^a</td>
<td>ND^b</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. marinum</em> ST2</td>
<td>ND^a</td>
<td>ND^b</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td><em>M. marinum</em> ST5</td>
<td>ND^a</td>
<td>ND^b</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

ST, the nucleotide sequences of the alleles were concatenated in frame to generate an artificial CDS. Using the same approach, all MPM were subjected to further MLSA analysis by determining the partial DNA sequences of four pMUM plasmid CDS (repA, parU, MUP045, and the mls loading domain) as described previously (31).

Phylogenetic methods. Phylogenetic analysis was performed with MEGA software version 2.1.2 (18) and Splits Tree version 4 (14). “P” distances were used throughout, as the overall level of sequence divergence was small. Values for synonymous (dS) and nonsynonymous mutation frequencies were calculated by Nei and Gojobori’s method (20), and standard errors (SE) for the means of these values were estimated by the method of Nei and Jin (21) using dSdNq (7).

VNTR analysis. The primers and conditions used for PCR amplification of the nine variable-number tandem repeat (VNTR) loci 1, 4, 6, 8, 9, 14, 15, 18, and 19 were as described previously (2). Amplicon sizes were estimated following electrophoresis of the PCR products through a 2% agarose gel and comparison with a 100-bp DNA ladder (Promega).

Nucleotide sequence accession numbers. The DNA sequences of the alleles have been assigned GenBank accession numbers DQ985290 to DQ985355.

RESULTS

DNA-DNA hybridization. Whole-genome probes derived from the five novel MPM strains with MLSA ST9, -10, -11, -12, and -14 (see below); three *M. ulcerans* strains (ST17, -19, and -20); and a representative non-MPM *M. marinum* strain (ST22) were hybridized to each other in a reciprocal fashion. The five novel MPM strains were then tested against genomic DNAs from three additional *M. marinum* strains (ST1, -2, and -5). The results are summarized in Table 2, and an example is
given in Fig. 2, showing that by this technique the five new MPM are indistinguishable from strains of *M. ulcerans*, with an overall mean RBR of 98% (range, 79 to 117%). In contrast, the mean RBR between the MPM and non-MPM was significantly lower at 40% (range, 13 to 68%; *P* < 0.05).

**MLSA of chromosomal genes.** To improve chromosome coverage, we expanded our original MLSA method and added a 357-bp fragment of the *glcB* gene to the seven loci we had previously sequenced from 18 isolates of *M. ulcerans* and 22 isolates of *M. marinum* (Fig. 1) (32). A large difference in chromosomal location between the species was noted for some genes, and this was due to the extensive genome rearrangements that have occurred in *M. ulcerans* (34) ([http://genolist.pasteur.fr/BuruList/](http://genolist.pasteur.fr/BuruList/)). All eight loci were then sequenced for the 12 new isolates used in this study, which included five MPM recovered from diseased fish and frogs. Multiple alleles were obtained for each locus among the total of 52 mycobacteria tested. For example, 13 different *criB* alleles, 7 different *adk* alleles, and 10 different *fbpA* alleles were obtained. A unique combination of alleles is referred to as an ST, and 20 STs were obtained (see Table S3 in the supplemental material). The eight alleles for each ST were concatenated in frame in the order *criB, adk, fbpA, aroE, groEL, ppk, sodA,* and *glcB* to produce an artificial 3,210-bp CDS (Fig. 3A and B). DNA sequence alignments and comparisons of this CDS across all 20 STs revealed only 95 variable sites, representing a shared nucleotide identity of 97%, highlighting the very close genetic relationship between all the mycobacteria analyzed in this study (Fig. 3A and B). The MPM were represented by 11 STs, and they exhibited a very high level of nucleotide identity, being identical at the nucleotide level.

**Comparison of plasmid sizes and distributions of IS2404 and IS2606.** Pulsed-field gel electrophoresis and Southern hybridization were used to compare the sizes of the pMUM plasmids among the novel MPM and to explore IS copy numbers. *M. ulcerans* strains from Malaysia and Japan had plasmids of approximately 170 kb, as previously reported for strains from Africa and Australia (Fig. 4) (31, 33). The novel MPM had plasmids of larger sizes, *M. liflandii* (ST14) and *M. marinum* strain CC240299 (ST10) harbor plasmids of approximately 180 kb, while *M. pseudoshottsii* (ST11) and *M. marinum* DL240490 (ST9) and *M. marinum* DL045 (ST12) all harbor plasmids of approximately 200 kb (Fig. 4). The plasmid size correlated with restriction fragment polymorphism, as shown after digestion of DNAs from these strains with Asel and hybridization with probes derived from IS2404 and IS2606. Of particular note are the patterns of bands, which are indistinguishable from each other for the strains with 200-kbp plasmids (ST9, -11, and -12). MLSA supported this finding with 99.9% shared nucleotide identity and only three variable nucleotides from the 3,210 bp sequenced among these strains (Fig. 3). Southern hybridization confirmed that all MPM contain many copies of IS2404, suggesting that this insertion sequence and possibly its expansion to a high copy number were characteristics of the common progenitor. The same is not true for IS2606. It was not detected in *M. marinum* CC240299, a finding confirmed by PCR (data not shown), and was present in low copy numbers in the *M. ulcerans* strain from Japan (ST19) and the novel MPM (Fig. 4), suggesting that it was acquired independently by the two different lineages of MPM.

**Distribution of pseudogenes among novel MPM.** Approximately 14% of the original coding potential of *M. ulcerans* Agy99 has been lost due to point mutations or insertion sequences that have disrupted CDS and led to the formation of pseudogenes. To explore the distribution of pseudogenes in the novel MPM, we used PCR to amplify and then sequence
the region spanning different mutations. These mutations were nucleotide changes that changed the reading frame or introduced a premature stop codon or disruptions by insertion of IS2404. Twenty of the 743 pseudogenes (including 21 mutations) identified in *M. ulcerans* Agy99 were selected for comparison with novel MPM of ST9, -10, and -11. These sequences were selected as a random sample of the total pseudogene complement with potential to explain phenotypic differences between strains. The results are summarized in Table S4 in the supplemental material and show that 15 of these pseudogenes...
are still intact in the novel MPM; of the five that were pseudogenes, only one (an insertion of 5 nucleotides in \textit{arsC}) was caused by the same mutation as in \textit{M. ulcerans} Agy99. The frequency of the G-to-T transition in \textit{crtI}, encoding phytoene dehydrogenase, which is required for carotenoid pigment production, was more widely investigated by testing all 52 mycobacteria. This mutation was found to be restricted to \textit{M. ulcerans} strains belonging to ST17, -18, and -20 (Fig. 3C).

**Distribution of DNA deletions among MPM.** Deletion of DNA fragments occurs frequently in mycobacteria, and their distribution among strains can be used to reconstruct evolutionary pathways (5). Comparative genomics revealed 157 between \textit{M. ulcerans} strain Agy99 (ST17) and \textit{M. marinum} M (ST22) (34). Three of these regions (MURD12, MURD54, and MURD152) that were shown to be deletions in \textit{M. ulcerans} Agy99 were selected for further study. Using PCR, MPM and \textit{M. marinum} were screened for the presence or absence of each MURD, using the approach developed for studying members of the \textit{Mycobacterium tuberculosis} complex (5) (Fig. 5A). Analysis of strains representing each of the 20 STs revealed that MURD12 had been lost from all strains that constituted the endotherm lineage, represented by ST15, -17, -18, -19, and -20 (Fig. 3C and 5). \textit{M. ulcerans} strains of the subcluster within the endotherm lineage (ST17, -18, and -20) have lost MURD54 and MURD152 (Fig. 3C and 5). \textit{M. ulcerans} strain 5114 from Mexico (ST15) did not produce a product in the MURD152 assay, suggesting that this strain has undergone a different chromosomal modification in this region.

**VNTR analysis distinguishes different MPM.** Variable-number tandem repeat loci are widespread in mycobacterial genomes, and their varied distribution has been exploited for strain differentiation. A system of PCR and sequence-based VNTR typing has been described for high-resolution differentiation among \textit{M. marinum} and \textit{M. ulcerans} strains (1, 2, 12, 35), and we applied this technique to the MPM strains in this study. Comparisons of fragment sizes at VNTR loci 1, 4, 6, 8, 9, 14, 15, 18, and 19 permitted the discrimination of the novel MPM from other strains of \textit{M. marinum} and \textit{M. ulcerans}. However, \textit{M. pseudoshottsii} L15 (ST12), \textit{M. marinum} DL240490 (ST9), and \textit{M. marinum} DL045 (ST11) shared the same VNTR profile (data not shown).

**DISCUSSION**

Initial MLSA analyses and subsequent whole-genome comparisons have shown that \textit{M. ulcerans} has recently evolved from \textit{M. marinum} by acquisition of the pMUM plasmid and reductive evolution (31–34). The recent discovery of MPM that are phenotypically distinct from \textit{M. ulcerans} in diseased fish and frogs has highlighted the possibility that pMUM is being transferred among different mycobacterial species (19, 26). In this report, we show by a systematic genetic approach that all MPM are very closely related to each other and have evolved, not by multiple exchanges of pMUM, but from a common \textit{M. marinum} progenitor that acquired the plasmid.

MLSA is widely used to understand the taxonomic relationships among bacterial populations (8, 10, 11), and it was the method we employed in an earlier study to suggest that \textit{M. ulcerans} recently evolved from \textit{M. marinum} (32). To improve genome coverage and increase resolution of the MLSA method for the present investigation, we added an eighth locus to create a 3,210-bp semantide (a large information-bearing molecule). We reanalyzed our original data set and then added sequences from 12 additional isolates that included 5 mycobacteria recently shown to contain pMUM and to produce mycolactones. There was significant sequence diversity among non-mycolactone-producing \textit{M. marinum} isolates, and as shown by others (32, 39), the majority of isolates fell into two distinct clusters represented by ST1, -2, and -6 and ST 3, -4, and -5. The sample size in this study was too small to draw
conclusions linking specific *M. marinum* genotypes to virulence in humans, as has been proposed (39); however, such a correlation seems unlikely, given that *M. marinum* isolates of human origin spanned the spectrum of sequence diversity revealed by MLSA. Two *M. marinum* isolates recovered from armadillos showed intermediate sequence types (ST7 and -8), as did *M. marinum* strain "M" (ST22), whose genome has recently been sequenced (http://www.sanger.ac.uk/Projects/M_marinum).

MLSA unambiguously showed that, despite their varying phenotypes, all MPM have evolved from a single *M. marinum* clone that has since expanded into at least two distinct lineages, and this was supported by the congruent tree topology derived from the four pMUM plasmid loci (Fig. 3). Together with equivalent levels of synonymous nucleotide substitution frequency between chromosome and episome sequences, these data suggest that plasmid acquisition was probably the principal event that enabled an *M. marinum* progenitor to survive in a new environment.

The genetic homogeneity of MPM was also reflected in their high DDH values. DDH has been widely used for over 30 years in bacterial taxonomy to infer relatedness between genomes, and a DDH value greater than 70% is one criterion used to help define a bacterial species. The high DDH values among *M. ulcerans* and the other MPM were further evidence of their common origin and contrasted with the low (<55% RBR) values when the same MPM were tested against non-mycolactone-producing *M. marinum* strains. These data are consistent with an earlier investigation of *M. ulcerans* and *M. marinum* that showed intraspecies DDH values of >90% and interspecies DDH values of <50% (36). The presence of pMUM-like plasmids and the multiple copies of IS2404 in all MPM may explain, at least in part, the striking DDH results. These data also indicate that IS2404 acquisition (and possibly its expansion to high copy numbers) occurred before radiation of MPM around the world. In contrast, IS2606 is present in high copy numbers only in the lineage of *M. ulcerans* strains that contain isolates from Africa, Australia, and Malaysia (ST17, -18, and -20) (30, 32). The varied distribution of this ISE among other MPM and its absence from the MPM *M. marinum* strain CC240299 (ST10) suggest it has been transferred independently to at least two different populations of MPM subsequent to IS2404 and pMUM acquisition.

The *M. ulcerans* genome project and a recent microarray-based study have both revealed extensive DNA deletion polymorphism among *M. ulcerans* strains (34; M. Käser, S. Ronchini, T. Stinear, M. Tessier, C. Mangold, G. Dernick, M. Naegeli, F. Portaels, U. Certa, and G. Pluschke, submitted for publication). In the current work, the pattern of DNA deletion...
observed among MPM for three deletions was in good agreement with MLSA (Fig. 5). MURD12, a 10-kbp fragment containing CDS involved in secondary metabolism, was a marker for distinguishing between the ectotherm and endotherm lineages, as it was absent from M. ulcerans isolates from both Mexico and Japan (ST15 and -19) and the African, Malaysian, and Australian cluster (ST17, -18, and -20) (Fig. 3C and 5C). The MURD54 and MURD152 deletions differentiate the ST17, -18, and -20 subcluster from other MPM and are indicative of more advanced genome reduction in these strains. MURD152 is a 2.8-kbp DNA fragment deleted from M. ulcerans strain Agy99 (ST17) that spans esxA and esxB, genes encoding key components of the ESX-1 secretion apparatus and virulence factor (4, 34). The MURD152 deletion PCR assay confirmed earlier findings that showed that esxA and esxB are absent from M. ulcerans strains from Africa, Australia, and Malaysia but present in other MPM (19, 26). Both the MURD54 and MURD152 assays may have diagnostic applications in countries where Buruli ulcer is endemic, such as Africa and Australia, where it will be useful, particularly when screening environmental samples, to distinguish between M. ulcerans strains that cause Buruli ulcer and other MPM.

The split of MPM into two distinct lineages, which include strains with different species names (e.g., M. marinum, M. liiflandii, and M. pseudoshottii) that typically cause disease in ectotherms but also have a high zoonotic potential and strains of M. ulcerans (ST17, -18, and -20) that cause Buruli ulcer in humans and target other endotherms, is an important finding. Some insights into the genetic basis for this separation have been gleaned from this study and, combined with previous research showing that these strains have different optimal growth temperatures and produce mycolactones with varying potencies (19, 26), they suggest MPM have evolved to occupy different ecological niches. M. ulcerans is not known to infect fish, while the diseases caused by M. marinum and M. ulcerans in humans differ greatly in their clinical, histopathologic, and epidemiologic aspects (32).

The large number of gene deletions and pseudogenes in the M. ulcerans Agy99 genome compared with M. marinum M is indicative of a bacterium adapting to a restricted and privileged environment, where mutations are tolerated in genes that are no longer required for survival. However, testing three MPM from the ectotherm lineage for 20 of these pseudogenes found only five inactivated CDS, indicating a much less advanced level of genome decay and metabolic streamlining, consistent with the hypothesis that they occupy different environments. Only one of the five mutations (an insertion in arSc) was conserved between the two MPM lineages (see Table S4 in the supplemental material), suggesting that a certain level of genome decay had begun before divergence. The other four mutations occurred at different positions in the same genes (sigI, echA13, and accD1), and this may indicate that the products of these CDS are not only redundant but perhaps deleterious for survival of MPM and so have been subjected to independent, purifying selection.

The difference in pseudogene profiles may also help explain the phenotypic variation observed among MPM. For example, crf encodes phytoene dehydrogenase, an enzyme essential for the production of carotenoid pigments in M. marinum (25). M. ulcerans Agy99 has the same crf locus as M. marinum, but it is nonpigmented, and this has been explained by a point mutation in crfI that introduces a premature stop codon and truncates the gene (see Table S4 in the supplemental material) (34). There was complete correlation between lack of pigment production and the disrupted crfI gene, as only MPM of the ST17, -18, and -20 cluster are nonpigmented (24, 26), and it was only this cluster that contained the mutation (Fig. 3C).

In this report, we have sought to clarify the genetic relationships among mycolactone- and non-mycolactone-producing mycobacteria, but this has in turn highlighted the recurring problem of assigning species status to highly related bacteria, as the question remains how much diversity is permissible in a genetically discrete cluster for it to be regarded as a distinct taxon. From a population genetics standpoint, the data presented here do not support the separation of MPM into different species. Employing a subspecies nomenclature might allow a more meaningful naming system that accurately reflects the common origin of MPM. A comprehensive polyphasic and multicenter study of MPM, as performed by Wayne et al. (41), would help decide their taxonomic positions.

Defining the host specificity and natural ecology of MPM is also a research priority. It may be that there are many different MPM but the only strains isolated are those producing mycolactones with sufficient potency to cause disease in humans, fish, frogs, possums, and koalas (23). A better understanding of MPM in the environment will be crucial to halting the spread of the diseases they cause, in particular Buruli ulcer.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support of the National Health and Medical Research Council of Australia. The collection of the Israeli mycobacterial isolates is maintained at NCM, Eilat, Israel, thanks to the generosity of Dan and Florence Green (North American Friends of IOLB).

We thank Amy Shields for statistical advice, Y. Kazumi and P. L. C. Small for providing M. ulcerans strains, and J. Parkhill for the use of unpublished M. marinum genome sequence data produced by the Pathogen Sequencing Group at the Sanger Institute (ftp://ftp.sanger.ac.uk/pub/pathogens/mm).

REFERENCES
