Large Sequence Polymorphisms Unveil the Phylogenetic Relationship of Environmental and Pathogenic Mycobacteria Related to *Mycobacterium ulcerans*^{∇†}

Michael Käser,^{1*} Julia Hauser,¹ Pamela Small,² and Gerd Pluschke¹

Swiss Tropical Institute, Molecular Immunology, Socinstr. 57, 4002 Basel, Switzerland,¹ and University of Tennessee, Department of Microbiology, M409 WLS, Knoxville, Tennessee 37996-0845²

Received 23 February 2009/Accepted 19 June 2009

Mycolactone is an immunosuppressive cytotoxin responsible for the clinical manifestation of Buruli ulcer in humans. It was believed to be confined to its etiologic agent, *Mycobacterium ulcerans*. However, the identification of other mycolactone-producing mycobacteria (MPMs) in other species, including *Mycobacterium marinum*, indicated a more complex taxonomic relationship. This highlighted the need for research on the biology, evolution, and distribution of such emerging and potentially infectious strains. The reliable genetic fingerprinting analyses presented here aim at both the unraveling of phylogenetic relatedness and of dispersal between environmental and pathogenic mycolactone producers and the identification of genetic prerequisites that enable lateral gene transfer of such plasmids. This will allow for the identification of environmental reservoirs of virulence plasmids that encode enzymes required for the synthesis of mycolactone. Based on dynamic chromosomal loci identified earlier in *M. ulcerans*, we characterized large sequence polymorphisms for the phylogenetic analysis of MPMs. Here, we identify new insertional-deletional events and single-nucleotide polymorphisms that confirm and redefine earlier strain differentiation markers. These results support other data showing that all MPMs share a common ancestry. In addition, we found unique genetic features specific for *M. marinum* strain M, the genome sequence strain which is used widely in research.

The production of the cytotoxic and immunosuppressive mycolactone was, since its description, believed to be confined to *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. Likewise, the existence of specific insertion sequence elements (ISEs), IS2404 and IS2606, was thought to be *M. ulcerans* specific. However, genes for mycolactone production, along with IS2404 and IS2606, have been increasingly identified in a number of environmental and/or pathogenic mycobacteria, with broad regional distribution. Therefore, questions arose of where mycolactone is coming from and what role it has in the evolution, adaptation, and pathogenicity of environmental and pathogenic mycobacteria. Mycobacteria carrying a version of the virulence plasmid pMUM001, containing genes for the production of mycolactone, are designated as mycolactone-producing mycobacteria (MPMs). These strains are all closely related, and there is some justification for considering them part of a *Mycobacterium marinum* complex (42). The genome sequence of *M. marinum* strain M is available (37), and it is devoid of mycolactone-producing coding sequences (CDSs) as well as both IS2404 and IS2606. In contrast, a unique mycolactone (mycolactone F) has been isolated from two strains of *M. marinum* (CC240299 and DL240490) that have produced diseased fish in the Red Sea (14, 29, 39). Both of these strains contained IS2404, and in addition, only *M. marinum* DL240490 also contained IS2606. Mycolactone F and both of these ISEs

have also been found in the fish pathogen *Mycobacterium pseudoshottsii* L15, isolated from diseased fish in the Chesapeake Bay in the United States (29, 30). “*Mycobacterium liflandii*” XT128, which was isolated from African clawed frogs that were introduced into research laboratories, also has both of these ISEs and possesses a slightly different plasmid that encodes mycolactone E (14, 25). Mycolactones produced by clinical isolates of *M. ulcerans* cluster geographically, as follows: strains originating from Africa and Australia produce mycolactones A/B and C (8, 9, 12, 24), while a Chinese *M. ulcerans* strain produces mycolactone D (13).

M. marinum causes a granulomatous intracellular infection in fish, mainly in bass cultivated in fisheries. Whereas *M. marinum* also occasionally produces self-limiting granulomatous skin lesions in humans, *M. ulcerans* disease manifests as a progressing skin ulcer that affects the underlining tissue and sometimes also the bones (34, 40). If left untreated, spontaneous healing of Buruli ulcer may result in disfigurement and disabilities with the affected extremities. Buruli ulcer is endemic in more than 30 countries worldwide, with a primary focus on remote areas of West Africa. School-age children are the most often affected group.

Both the distribution and epidemiological dispersal of *M. ulcerans* strains and the transmission pathway are unclear. An environmental reservoir is likely for Africa, as was recently shown for Australia (7, 17, 21). The fact that *M. ulcerans* is unlikely to be communicable but is associated with exposure to stagnant and slow-moving bodies of water has led to the hypothesis that infection results from environmental exposure. For instance, the presence of free-living amoeba may be associated with high Buruli ulcer endemicity (5). Until now, only one *M. ulcerans* isolate could be cultured from the environ-

* Corresponding author. Mailing address: Swiss Tropical Institute, Molecular Immunology, Socinstr. 57, 4002 Basel, Switzerland. Phone: 41-61-2848-171. Fax: 41-61-2848-101. E-mail: m.kaeser@unibas.ch.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

[∇] Published ahead of print on 10 July 2009.

TABLE 1. Mycobacteria under investigation, including information on mycolactone version and ISE abundance

Species and strain	Geographical origin	Isolation site	Mycolactone version ^a	Reference(s)/ source	Quantification of ^b :							
					IS2404				IS2606			
					PCR result ^b	C _T value	SD	No.	PCR result ^a	C _T value	SD	No.
<i>M. marinum</i> M	California, U.S.	Human patient	None	37	None	NA	NA	10	None	NA	NA	10
<i>M. marinum</i> ATCC 927	Philadelphia, PA, U.S.	Fish	None	ATCC ^c	None	NA	NA	5	None	NA	NA	5
<i>M. marinum</i> CC240299	Ma'agan Michael, Israel	<i>Cyprinus carpio</i> (koi)	F	29, 39	+	18.95	0.83	5	None	NA	NA	5
<i>M. liflandii</i> XT128	California, U.S.	<i>X. tropicalis</i>	E	25, 29	+	16.38	0.85	5	+	23.97	2.10	5
<i>M. marinum</i> DL240490	Eilat, Israel	<i>Dicentrarchus labrax</i> (sea bass)	F	29, 39	+	20.56	2.33	5	+	27.73	2.81	5
<i>M. pseudoshottsii</i> L15	Maryland, U.S.	<i>Morone saxatilis</i> (striped bass)	F	29, 30	+	19.10	1.42	5	+	24.19	1.39	5
<i>M. ulcerans</i> 842	Surinam, South America	Human patient	D?	31	+	19.99	0.74	2 ^d	+	28.90	0.59	2 ^d
<i>M. ulcerans</i> 9756	Japan	Human patient	D? A/B?	31	+	19.45	0.98	5 ^e	+	26.64	0.16	5 ^e
<i>M. ulcerans</i> Agy99	Ghana, Africa	Human patient	A/B	38	+	20.44	1.59	30 ^f	+	20.47	1.60	30 ^f

^a Information from reference 29. For strain 9756 from Japan, mycolactone was supposed to be version D, as for strain ITM 980912 from China, but for now remains unclear since a recent investigation determined mycolactone A/B for yet another Japanese strain (26).

^b Information from reference 29.

^c American Type Culture Collection, LGC Standards (<http://www.lgcstandards-atcc.org/ATCC/CulturesandProducts/tabid/979/Default.aspx>).

^d Includes South American haplotype *M. ulcerans* strains (19).

^e Includes Asian haplotype *M. ulcerans* strains (19).

^f Includes *M. ulcerans* strains of the classical lineage (19).

^g NA, not applicable.

ment (28), but there is considerable molecular evidence for the presence of the organism in aquatic habitats. IS2404 has been identified by PCR in water and plants (23, 28, 32, 36, 41), insects (17, 27), and snails and fish (6, 22), but it remains questionable whether the detection of this genetic marker reflects the actual presence of *M. ulcerans* or that of other MPMs. Still, recent PCR amplification of a variable number of tandem repeats and mycolactone genes has provided further evidence for the widespread distribution of *M. ulcerans* in aquatic habitats (41). MPMs have been distinguished from each other using DNA-DNA hybridization, a variable number of tandem repeats, and multilocus sequence analysis (41, 42). However, several questions remain unresolved. What are the natural reservoirs of MPMs? Was mycolactone production acquired multiple times in the evolution of mycobacteria, and how did the diversification of MPMs occur? Is there a role for mycolactone in mycobacterial survival or even in the transition of MPMs from strictly environmental mycobacteria to mycobacteria pathogenic for (mammalian) hosts?

Genotyping of *M. ulcerans* isolates, aiming at microepidemiological analyses, is demanding due to the exceptional lack of strain diversity among African isolates (M. Käser, O. Gutmann, J. Hauser, T. Stinear, S. Cole, D. Yeboah-Manu, G. Dernick, U. Certa, and G. Pluschke, submitted for publication). Large sequence polymorphisms (LSPs), identified by microarray analyses and characterized by PCR, sequencing, and bioinformatics-based sequence comparisons (19, 31), led to the establishment of the first evolutionary tree of *M. ulcerans*, with two major genetic lineages worldwide (19). Thorough analysis of regions of difference (RDs) identified hot spot areas of enhanced genome instability, suggesting that the loss of particular genetic loci has played an important role in host adaptation and virulence (18). Some of the coding sequences in these RDs turned out to be independently deleted in strains of diverse geographic origin (15, 18).

In this study, we addressed the evolutionary relatedness of genomic sequences in a first selection of MPMs. We reasoned that the highly variable chromosomal regions previously described in *M. ulcerans* might also be affected in environmental mycobacteria. Thus, RD analysis of the genomic backbone in these hot spot areas might unveil phylogenetic relationships, as was successfully achieved for members of the *Mycobacterium tuberculosis* complex (2). Based on the RDs identified previously in *M. ulcerans* in comparison to *M. marinum* strain M, we have scanned two such hot spot regions of genome instability, RD9 and RD12, for insertional-deletional polymorphisms (InDels). In addition, strains were characterized by multilocus sequence typing (MLST), and these data were then supplemented by newly identified single-nucleotide polymorphisms (SNPs) and compared to ISE presence and copy number determination.

MATERIALS AND METHODS

Mycobacterial strains and genomic DNA extraction. Nine mycobacterial strains were analyzed, as listed in Table 1. DNA was extracted using an optimized method for mycobacterial DNA preparation (20). Bacterial pellets of about 20 mg (wet weight) were heat inactivated for 1 h at 95°C, followed by cell wall disruption and digestion. DNA was extracted from the supernatants by phenol-chloroform extraction and subjected to ethanol precipitation, as previously described (20). DNA was measured by optical density at 260 nm (NanoDrop 1000; Wilmington, DE).

DNA amplification and sequencing. PCR was performed using 10× FIREPol BD buffer, 0.5 μl FIREPol *Taq* polymerase (Solis BioDyne, Tartu, Estonia), 5 ng genomic DNA, 0.6 μM each of forward and reverse primers, 1.7 mM MgCl₂, and 0.3 mM deoxynucleoside triphosphate at a total volume of 30 μl. Long-range PCR polymerase mix (Fermentas, St. Leon-Rot, Germany) was applied, according to the manufacturer's protocol, to retrieve PCR products longer than 3 kb. PCRs were run in a GeneAmp PCR system 9700 PCR machine. The thermal profile for PCR amplification of mycobacterial genomic DNA included an initial denaturation step of 95 to 98°C for 3 min, followed by 32 cycles of 95°C for 20 s, annealing at 58 to 65°C for 20 s, and elongation at 72°C for 30 s to 4 min. The PCRs were finalized by an extension step at 72°C for 10 min. PCR products were analyzed on 1% agarose gels by gel electrophoresis, using ethidium bromide

staining and the AlphaImager illuminator (Alpha Innotech, San Leandro, CA). PCR fragments produced for analysis of unknown genomic sequences were purified using the NucleoSpin purification kit (Macherey-Nagel, Düren, Germany) and subjected to direct sequencing. Sequencing was done by Macrogen, Seoul, Korea. Primers (Sigma-Aldrich, Steinheim, Germany) were selected based on the genome sequences of *M. marinum* M (GenBank accession numbers CP000854 and CP000895) and *M. ulcerans* Agy99 (GenBank accession number CP000325) using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). For amplification of unknown regions, these were combined with outward directed primers corresponding to sequences within the IS2404 element.

MLST. MLST was undertaken for four genes which clearly differentiate among the mycobacteria used in this study using the following primers: Adk gene (Adk_f, ATCATGATCGGTGTCGGTTT; Adk_r, CTGTGGCACCCTCTGCTAC), AroE gene (AroE_f, AACACCTGGCGAATATCGAG; AroE_r, GGTGTAGTCAACACACGCA), FbpA gene (FbpA_f, TTCCTGACCGAGCTGCGG; FbpA_r, CCCAGTACTCCAGCTGTGC), and Sod gene (SodA_f, AACCCACATCTCAGGTCTCAG; SodA_r, CTCGGGTATTCGAGTCAGC).

Bioinformatic sequence analyses. Comparative in silico sequence analysis was performed using the following tools: the sequence manipulation suite (<http://bioinformatics.org/sms/index.html>), the sequence alignment tool BLAST 2 sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>), the multiple sequence alignment website MultAlin (<http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>), the Artemis software release 9 (33), and the Artemis Comparison Tool software release 6 (3). Retrieved MLST sequences were aligned using the EMBL-EBI ClustalW2 Alignment tool (<http://www.ebi.ac.uk/>), and differing nucleotides were used to create a splits network with split decomposition using SplitsTree4 version 4.10 (16; www.splitstree.org).

Real-time PCR. Primers (Sigma-Aldrich, Steinheim, Germany) and TaqMan probes (Biomers, Ulm, Germany) were designed using Primer Express software version 2.0 (Applied Biosystems, Foster City, CA), and probes were each 5'-end labeled with the fluorescent dye 6-carboxyfluorescein (FAM) and 3'-end labeled with the quencher 6-carboxytetramethylrhodamine (TAMRA). Primers and probes targeted *M. ulcerans* Agy99 sequences of IS2404 (IS2404cf, AAAGCAC CACGCAGCATCTT; IS2404cr, AGCGACCCAGTGGATTG; and IS2404cp, FAM-CCGTCCAACGCGATCGGCA-TAMRA), IS2606 (IS2606f, TGCTGAC GGAGTTGAAAAACC; IS2606r, CCTTTGAGGCCGTCACAGA; and IS2606p, FAM-CGGCGTGGCCGACATCTTCTTC-TAMRA), and GroEL (GroELf, CCT GCTGAGCGTCGAAGTC; GroELr, GGGCACCGAGCTGGAGTT; and GroELp, FAM-CCGAGAGGTATCCCTTGTGCGAAACCG-TAMRA). Real-time PCR mixtures contained 50 fg of template DNA, 900 nM of TaqMan probe, 300 nM of each primer, and TaqMan universal PCR master mix (Applied Biosystems, Foster City, CA) at a total volume of 25 μ l. Amplification and signal detection were performed using the 7500 real-time PCR system (Applied Biosystems, Foster City, CA) under the following conditions: 1 cycle of 50°C for 2 min, 1 cycle of 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. Quantitative TaqMan real-time PCR threshold cycle (C_T) values for the ISEs were normalized by detection of the single-copy GroEL target sequence. Samples were repeated at least twice, and negative controls were included in each assay.

Identification of insertional-deletional genomic events. Mycobacterial strains of diverse origin and species affiliation, as listed in Table 1, were investigated for two selected RDs. Scanning by PCR of overlapping 1-kb steps in MPM genomes revealed identical amplification products over large segments in the 40-kb and 60-kb regions comprising RD9 and RD12, respectively. However, PCR products differing in size or absent from the genome sequence of *M. marinum* M provided evidence for insertional-deletional genomic variations. Thorough determination of the respective ends of the genomic deviations down to some 300 bp made it possible to design primer pairs complementary to regions flanking the polymorphic sites. PCR analysis that bridged the sequence deviations yielded products that, upon sequencing, revealed the sizes of the deletions, their exact breakpoints, and the nature of the inserted DNA. Primers used for genomic characterization in RD9 (see Fig. 1) were KM096 (AACGTCGGTATCCGTTGGT), MK146 (TGGTGCTTTCAATGTGGTA), and KM113 (TTGAATAGTCGG GCAATTCC), and those used for genomic characterization in RD12 (see Fig. 3) were MK904 (CTACCACTGTGCAGCAATCC), MK905 (TCGACAGTTTCA ATGCTTTG), MK893 (AACGCTAACCAATGGCAATCC), and MK632 (AAG CGCAGAGTTGGATGACT).

Nucleotide sequence accession numbers. New sequences and those that had reproducibly slight differences to recently published data (42) were submitted to NCBI, as follows (GenBank accession numbers): *M. marinum* ATCC 927 AroE gene allele 13 (FJ685517), *M. marinum* CC240299 SodA gene allele 7 (FJ685518), *M. liflandii* XT128 SodA gene allele 8 (FJ685519), *M. pseudoshottsii*

L15 SodA gene allele 9 (FJ685520), and *M. marinum* ATCC 927 SodA gene allele 10 (FJ685521).

RESULTS

Identification of new InDels in RD9. A PCR using primers KM096 and MK146 revealed the presence of various DNA sequences in a region designated as RD9 (Fig. 1). Based on the *M. marinum* M genome sequence, a PCR product made up of 543 bp was obtained, as predicted. An identical product was also found in *M. marinum* ATCC 927, *M. marinum* CC, *M. liflandii*, and *M. ulcerans* Agy99 (Fig. 1B, top). However, *M. marinum* strain DL and *M. pseudoshottsii* showed PCR bands which were approximately 1.4 kb larger than expected (Fig. 1B, top, lanes 4 and 6). Sequencing of these products revealed the insertion of an IS2404 element into the CDS of PPE5. We designated this genomic event as InsertRD9F. Breakpoints on either side of the insertion were identical for both strains (Fig. 1A), indicating that this insertion was derived from a common ancestor. No amplification product was retrieved in *M. ulcerans* strains Surinam and Japan, which were known to have large deletions in this region (Δ RD9B of 30.5 kb in the *M. ulcerans* South American haplotype and Δ RD9A of 24.1 kb in the Asian haplotype [19]) (Fig. 1).

When using primer pair KM096 and KM113 (Fig. 1), the expected PCR product size deduced from the *M. marinum* M sequence was too large (6.6 kb) for PCR amplification under the conditions used but yielded a band of 3.1 kb for *M. ulcerans* Agy99 (Fig. 1B, bottom, lane 9). This PCR product corresponds to the described InDel Δ RD9D/MURD95, where 4.2 kb of sequence is substituted by an IS2404 insertion (19). DNA regions in *M. marinum* ATCC 927, *M. marinum* DL, and *M. pseudoshottsii* were also too large to amplify PCR products, but PCR analysis showed that all DNA stretches in between the primers were present (not shown). Primers could not bind in strains of the *M. ulcerans* South American and Asian haplotypes due to the absence of these regions in Δ RD9B and Δ RD9A, respectively (Fig. 1A). However, for *M. marinum* strain CC and *M. liflandii*, this primer pair resulted in an unexpected band of $\sim$ 2.4 kb (Fig. 1B, bottom, lanes 3 and 5). Sequencing of these PCR products revealed an InDel, designated Δ RD9E, with IS2404 substituting 5.4 kb of DNA, comprising parts of the CDSs PPE5 and MMAR_3547 (Fig. 1A). Again, the InDel breakpoints in both strains were identical to each other. However, the deletion differs in size and breakpoints from Δ RD9D/MURD95 in *M. ulcerans* Agy99 (Fig. 1A).

Cluster differentiation by IS2404-specific SNPs. InDel events with perfect correspondence which leave identical scars in the genome are extremely unlikely to occur on two independent occasions. Thus, these newly found InDels provide evidence that a common ancestor was shared by *M. marinum* DL and *M. pseudoshottsii* on the one hand and by *M. marinum* CC and *M. liflandii* on the other hand, dividing these strains into two clusters. However, sequencing of the substituting IS2404 elements in Δ RD9E and InsertRD9F showed SNPs within the newly positioned ISEs. Three SNPs distinguished the cluster defined by Δ RD9E (GTC in *M. marinum* CC versus TCT in *M. liflandii* in the IS2404) (Fig. 1A), and four SNPs differentiated within the other cluster characterized by InsertRD9F (CACC in *M. marinum* DL versus GGTT in *M. pseudoshottsii*) (Fig. 1A).

FIG. 1. Comparison of CDS repertoire in RD9 between mycolactone producers and *M. marinum* strains not producing mycolactone. (A) Schematic view of RD9. Sequence deletions (bars) and insertions (arrows) compared to those of the *M. marinum* M sequence are indicated. Boxes with stripes inclined to the right, IS2404 element. SNPs differentiating between the unique ISEs are indicated. Nomenclature of CDSs is indicated along with the *M. marinum* annotation (http://www.sanger.ac.uk/Projects/M_marinum/), i.e., 3538 stands for MMAR_3538. (B) Ethidium bromide-stained agarose gels showing PCR products retrieved using the indicated primer pairs. Primers are also marked in the schematic view. Arrows indicate the larger amplification product containing InsertRD9F (top) and the products of various sizes due to different deletions (bottom). A band of 3.1 kb is shown for *M. ulcerans* Agy99 (bottom, lane 9, top arrow). However, for *M. marinum* strain CC and *M. liflandii*, the primer pair resulted in an unexpected band of ~2.4 kb (bottom, lanes 3 and 5, bottom arrow). The upper band found in *M. marinum* strain ATCC 927 (top) reflects no size dimorphism but a second binding of primer MK146 some 430 bp downstream, probably due to point mutations confined to this strain.

Identification of new InDels in RD12 reconfirms genetic clustering of MPMs. To provide further evidence to support these newly established mycobacterial clusters, we investigated a second genome region known to be unstable in *M. ulcerans*, RD12 (Fig. 2). Over a sequence stretch of ~60 kb, two deletions (Δ RD12B of 26.7 kb in the *M. ulcerans* South American haplotype and Δ RD12C of 41.3 kb in the Asian haplotype) and several insertions (only in the classical lineage, represented by strain Agy99) were already described (Fig. 2) (18). Here, we established further LSPs of extended sizes within RD12 (Fig. 2). *M. marinum* DL and *M. pseudoshottsii* revealed deletions of 38.2 kb substituted by an IS2404 element, constituting an InDel that we designated Δ RD12D. *M. marinum* CC and *M. liflandii* showed deletions of 14.8 kb, designated Δ RD12E. Again, breakpoints, adjacent regions, and substituting sequences of the LSPs were identical in the two affected strains, reconfirming the clusters found in RD9. No nucleotide exchange was detected in the substituting IS2404 element Δ RD12D. Altogether, this InDel-based analysis divides the fish-associated MPMs into

two clusters, irrespective of their origins and species names, as follows: the saltwater isolates and the freshwater isolates (Table 1).

Differentially inactivated gene repertoire. Analysis of CDSs in RD9 and RD12, whose expression is likely to be ablated among the *M. ulcerans* and MPM strains tested (Fig. 1 and 2), show that there is preferential loss of particular categories of genes. In RD9, with the exception of *M. marinum* M and *M. marinum* ATCC 927, PPE5 is disrupted in all strains. For this CDS, functional gene loss is associated with five independent genetic events in the strains analyzed (Fig. 1A; see also the table in the supplemental material). Similarly, MMAR_3547, a regulatory protein CDS, was deleted by four independent genome rearrangements. In RD12, MMAR_3993 (carrying an acyl coenzyme A dehydrogenase) is disrupted by five independent genetic events, several others (two PE/PPE family proteins, MMAR_3984 and PPE2; three dehydrogenases, MMAR_3992, adhB_1, and MMAR_3998; the hypothetical protein MMAR_3995; and a cytochrome P450 pro-

FIG. 2. Comparison of CDS repertoire in RD12 between mycolactone producers and *M. marinum* strains not producing mycolactone. Indicated are deletions (bars) and insertions (arrows), with ISE elements (IS2404, boxes with stripes inclined to the right; IS2606, boxes with stripes inclined to the left). Point mutations leading to pseudogenization of a CDS are indicated (X). The nomenclature of CDSs is as shown in Fig. 1.

tein, cyp187A4) by four independent genetic events, and many more CDSs by three independent genetic events (Fig. 2; see also the table in the supplemental material). Among the functional protein groups most affected are cell wall/cell processes, intermediary metabolism/respiration, lipid metabolism, and PE/PPE proteins (a detailed list of all CDSs of RD9 and RD12 inactivated in at least three independent events among the tested *M. marinum* and MPMs is found in the table in the supplemental material). Many of them belong to a group of proteins whose loss was already suggested to lead to selective advantages for *M. ulcerans* (18).

Identification of genetic characteristics specific for *M. marinum* M. In RD12, unique features were also found for *M. marinum* strains. A small deletion found in *M. marinum* ATCC 927, designated Δ RD12F, differentiates this strain from all other mycobacteria investigated here (Fig. 2). Δ RD12A/MURD105, formerly described as a deletion identical to those of the *M. ulcerans* classical lineage (MURD105) and members of the ancestral lineage (Δ RD12A, South American and Asian haplotypes), was also confirmed for other mycobacteria (Fig. 2). Only *M. marinum* M has a DNA stretch present, comprising MMAR_3972 carrying a hypothetical nonribosomal peptide synthetase containing a nonribosomal peptide-like adenylation domain. All other environmental and pathogenic strains under investigation instead showed sequences of 23 bp (Fig. 3). Thus, it appears that the 3,938-bp DNA stretch was introduced into *M. marinum* strain M rather than deleted from a common ancestor of all other mycobacteria except *M. marinum* M. This is supported by the following observations (Fig. 3). (i) A nucleotide BLAST analysis did not yield any results for other mycobacteria or actinobacteria, but significant similarity (~82%) was found for a small stretch (60 to 80 bp) of an

adenylation site in *Pseudomonas* species. (ii) An extended intergenic region was found between MMAR_3972 and fdx1. (iii) The GC content of the MMAR_3972 gene differs from that of the average genome and the reverse GC frame plot of the *M. marinum* M sequence, showing an unusual codon usage for this region (Fig. 3). Yet another genomic feature unique for *M. marinum* strain M was observed in RD3 (not shown), where the hypothetical membrane protein-carrying MMAR_3059 and MMAR_3060 genes represent a gene duplication that we could not find in any other mycobacteria listed in Table 1.

Quantification of ISEs in MPMs. The relative ISE genome copy numbers of the selected strains were assessed by quantitative real-time PCR (Table 1). The results are in accordance with the whole-genome sequence information of *M. marinum* M (no copies of either of the ISEs) and *M. ulcerans* Agy99 (213 versus 91 copies of IS2404 and IS2606, respectively). Moreover, our quantitative data reconfirmed previous observations on the presence or absence of ISEs retrieved with conventional PCR, where IS2404 was found in all investigated MPMs and IS2606 in all MPMs but *M. marinum* CC (Table 1) (29). The retrieved C_T values support the hypotheses that *M. marinum* ATCC 927 is devoid of both of the ISEs and that a high copy number of IS2606 is confined to the *M. ulcerans* classical lineage, which is prevalent in Africa and Australia (C_T value, 20.47) (Table 1) (18, 42). Low C_T values for both of the ISEs account for higher abundance in *M. liflandii* (16.38 for IS2404; 23.97 for IS2606). Otherwise, the copy numbers of IS2404 were comparable among the mycobacterial isolates under investigation (C_T values of 18.95 to 20.56) (Table 1).

Various genetic polymorphisms result in a congruent phylogeny. In order to test how these new observations fit in with earlier data and to combine all genomic information available,

FIG. 4. Splits graph of environmental and pathogenic mycobacteria with superimposed InDels, ISEs, and information on mycolactone production. The tree was based on MLST of housekeeping genes, distinguishing the investigated selection (see Materials and Methods) and newly identified SNPs in unique ISEs. InDels characteristic for each cluster are indicated between the connected nodes or isolates (in red). MPMs are clustered (circle), mycolactone versions are marked in green (information on mycolactone can be found in Table 1), and ISE versions are marked in blue. All nodes are covered and supported by specific features of LSPs from within the two analyzed RDs, RD9 and RD12 (indicated by arrows).

we compiled various genetically differentiating features of the analyzed environmental mycobacteria into one phylogenetic tree. For visualization, we created a splits graph based on MLST data (42; this study) and SNP diversity found in RD9 and RD12 in this investigation in order to superimpose our findings on InDel diversity and ISE quantification (Fig. 4). All nodes are covered and supported by specific features of LSPs from within the two analyzed RDs, RD9 and RD12. The phylogeny displayed by the newly identified InDels demonstrated a clustering of saltwater and freshwater MPM isolates, regardless of their geographic origin. The identified ISE-specific SNPs could further differentiate *M. marinum* DL240490 from *M. pseudoshottsii* L15 within the saltwater cluster and *M. marinum* CC240299 from *M. liflandii* XT128 within the freshwater cluster. In addition, all MPMs are characterized by the presence of both IS2404 and a mycolactone-encoding virulence plasmid. *M. marinum* CC240299 is the only MPM under investigation that showed no sign of carrying IS2606, whereas a high copy number of IS2606 is strongly linked with the virulent classical lineage of the human pathogen *M. ulcerans*. The MPMs show a smaller genetic distance to each other than to the two mycobacteria not producing mycolactone. The latter, including both the fish isolate *M. marinum* ATCC 927 and the human disease isolate *M. marinum* M, display distinct and characteristic genetic features.

DISCUSSION

The herein presented genetic characterization of MPMs in comparison to *M. marinum* strains devoid of mycolactone was

undertaken to enhance our understanding of the phylogenetic relatedness and dispersal of environmental and pathogenic mycobacteria related to *M. ulcerans*. We have used a diverse set of genetic tools, including InDel characterization, MLST, SNP analysis, and ISE quantification, to investigate their evolutionary relationship. Upon detailed InDel characterization within chromosomal hot spot regions previously defined in *M. ulcerans*, the identification of particular SNPs within these RDs, and ISE quantification, we added precision to genetic differentiation of MPMs, which resulted in a merged phylogenetic representation. Our results are in complete agreement with earlier approaches to genetic clustering (29, 42). Detailed characterization of LSPs turned out to be the most valuable tool for discriminating between the closely related family of MPMs and *M. marinum* strains not producing mycolactone that occupy comparable ecological environments. While LSPs are middle-rate indicators for genetic allocation when probed only for their absence or presence (1), they are excellent markers when analyzed down to the nucleotide levels of their breakpoints and insertional-deletion sequence content (2, 35). Since the scars left by genome rearrangements attest to unique events having occurred in a common ancestor, they are unequivocal markers for nodes between the branches of a phylogenetic tree and make it possible to reconstruct evolutionary pathways (2). The genotypic tools developed in this study can be extended to additional RDs already identified in *M. ulcerans* and applied to additional MPMs and related mycobacteria to be isolated in the future. Our analyses approved the genetic clustering of all MPMs compared to that of mycobacteria not

producing mycolactone. Although these data speak for a common ancestry of the analyzed MPMs, it still cannot be decided whether or not the respective virulence plasmids were taken up several times in parallel evolution (11, 26, 42).

One important finding of this study is the identification of genomic features specific for the patient isolate *M. marinum* M, isolated in 1992 in California. This is the only *M. marinum* strain included in this study for which a whole-genome sequence is available (37). Both our newly identified InsertRD12A and the gene duplication in RD3 represent unique features that distinguish this strain from all other mycobacteria included in this study and thus from a hypothetical *M. marinum* most-recent common ancestor.

While large numbers of IS2404 are present in the genomes of all MPMs analyzed (29, 42), the distribution of IS2606 presents a more diverse picture. It has been suggested earlier that IS2606 was independently acquired by different groups of MPMs (42). This might explain the absence of IS2606 in *M. marinum* CC240299 and its presence in *M. liflandii* XT128, despite the fact that other criteria such as MLST and InDel analysis place these strains in the same cluster (Fig. 4). The particularly striking fact that high copy numbers of IS2606 are confined to the highly virulent *M. ulcerans* classical lineage (10, 18, 19, 42; this study) renders it tempting to speculate that multicopy IS2606 is a possible prerequisite for enhanced virulence or adaptation.

It is of particular interest that all MPMs have been identified in association with environmental degradation (25, 30, 39). Beyond the presence rather than absence of mycolactone, the phylogenetic analyses in this investigation revealed a genetic correlation of organisms deriving from similar ecological ambience rather than from geographic origin. For instance, two isolates originating several 1,000 kilometers apart (*M. marinum* DL2404090 from the Red Sea, Israel, and *M. pseudoshottsii* L15 from the Chesapeake Bay in Maryland) share contig identity (and delimitation to other MPMs) of over 100 MB in the genetically unstable chromosomal regions RD9 and RD12. In this case, travel activities of marine fish leading to pathogen dispersal throughout the oceans are a plausible explanation. Even more surprising are the sequence identities and common ancestry of the two freshwater isolates *M. marinum* CC240299, from a lake in Israel, and *M. liflandii* XT128, from the African *Xenopus tropicalis*. The herein presented taxonomic relationship reveals that assigned species names may in some cases be misleading in terms of phylogenetic correlations.

Genome reduction, a prominent feature of the *M. ulcerans* strains of the classical lineage (18, 38), is much less pronounced among *M. marinum*, the ancestral *M. ulcerans* haplotypes, and other MPMs analyzed here. It has been suggested that genome reduction in classical lineage *M. ulcerans* and the consequent loss of highly immunogenic antigens and other antivirulence genes may have played a role in the adaptation to a changing habitat or host (4, 15, 18). Other CDSs were shown to be deleted in independent events, representing a general tendency of habitat/host adaptation by gene loss (18, 31). This observation is supported by the present findings of the same genes, identified earlier as pathoadaptive candidate genes, being disrupted in RD9 and RD12 by yet other deleterious events. Whether these parallel genetic deletions play a role for such mycobacteria in their lifestyle as either free living or being

associated with plants, amoeba, insects, or even mammals as natural habitats remains open for further investigation.

The genetic tree renders the question of the natural reservoir of mycolactone yet unresolved. Although all *M. marinum* strains investigated can infect fish, the ability to produce mycolactone may open new ecological niches and enable host-specific adaptation. Whether expression of mycolactone has a selective advantage for the survival and microbial evolution in the environment or host, or whether it is just a side effect in host pathogenicity, remains an enigma. Enhanced adaptation is likely to occur at least in the case of the *M. ulcerans* classical lineage, where the macrolide toxin is present and genome shrinkage is extensive, probably due to a high IS2606 copy number. Isolation of more MPMs and characterization with high-resolution genetic tools will expand our understanding of the emergence, evolution, epidemiology, and host specificity of animal and human mycobacterioses.

ACKNOWLEDGMENT

We thank Kerensa McElroy for contributing to experimental studies.

REFERENCES

- Alland, D., D. W. Lacher, M. H. Hazbón, A. S. Motiwala, W. Qi, R. D. Fleischmann, and T. S. Whittam. 2007. Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of *Mycobacterium tuberculosis* and the utility of LSPs in phylogenetic analysis. *J. Clin. Microbiol.* **45**:39–46.
- Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeyer, T. Garnier, C. Gutierrez, K. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**:3684–3689.
- Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21**:3422–3423.
- Demangel, C., T. P. Stinear, and S. T. Cole. 2009. Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat. Rev. Microbiol.* **7**:50–60.
- Eddiyani, M., J. F. De Jonckheere, L. Durnez, P. Suykerbuyk, H. Leirs, and F. Portaels. 2008. Occurrence of free-living amoebae in communities of low and high endemicity for Buruli ulcer in southern Benin. *Appl. Environ. Microbiol.* **74**:6547–6553.
- Eddiyani, M., D. Ofori-Adjei, G. Teugels, D. De Weirtdt, D. Boakye, W. M. Meyers, and F. Portaels. 2004. Potential role for fish in transmission of *Mycobacterium ulcerans* disease (Buruli ulcer): an environmental study. *Appl. Environ. Microbiol.* **70**:5679–5681.
- Fyfe, J. A. M., C. J. Lavender, P. D. R. Johnson, M. Globan, A. Sievers, J. Azuolas, and T. P. Stinear. 2007. Development and application of two multiplex real-time PCR assays for the detection of *Mycobacterium ulcerans* in clinical and environmental samples. *Appl. Environ. Microbiol.* **73**:4733–4740.
- George, K. M., D. Chatterjee, G. Gunawardana, D. Welty, J. Hayman, R. Lee, and P. L. Small. 1999. Mycolactone: a polyketide toxin from *Mycobacterium ulcerans* required for virulence. *Science* **283**:854–857.
- George, K. M., L. Pascopella, D. M. Welty, and P. L. Small. 2000. A *Mycobacterium ulcerans* toxin, mycolactone, causes apoptosis in guinea pig ulcers and tissue culture cells. *Infect. Immun.* **68**:877–883.
- Guerra, H., J. C. Palomino, E. Falconi, F. Bravo, N. Donaires, E. Van Marck, and F. Portaels. 2008. *Mycobacterium ulcerans* disease, Peru. *Emerg. Infect. Dis.* **14**:373–377.
- Hong, H., C. Demangel, S. J. Pidot, P. F. Leadlay, and T. Stinear. 2008. Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria. *Nat. Prod. Rep.* **25**:447–454.
- Hong, H., P. J. Gates, J. Staunton, T. Stinear, S. T. Cole, P. F. Leadlay, and J. B. Spencer. 2003. Identification using LC-MSn of co-metabolites in the biosynthesis of the polyketide toxin mycolactone by a clinical isolate of *Mycobacterium ulcerans*. *Chem. Commun. (Cambridge)* **22**:2822–2823.
- Hong, H., J. B. Spencer, J. L. Porter, P. F. Leadlay, and T. Stinear. 2005. A novel mycolactone from a clinical isolate of *Mycobacterium ulcerans* provides evidence for additional toxin heterogeneity as a result of specific changes in the modular polyketide synthase. *ChemBiochem* **6**:643–648.
- Hong, H., T. Stinear, P. Skelton, J. B. Spencer, and P. F. Leadlay. 2005. Structure elucidation of a novel family of mycolactone toxins from the frog

- pathogen *Mycobacterium* sp. MU128FXT by mass spectrometry. *Chem. Commun. (Cambridge)* **34**:4306–4308.
15. Huber, C. A., M.-T. Ruf, G. Pluschke, and M. Käser. 2008. Independent loss of immunogenic proteins in *Mycobacterium ulcerans* suggests immune evasion. *Clin. Vaccine Immunol.* **15**:598–606.
 16. Huson, D. H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**:254–267.
 17. Johnson, P. D., J. Azuolas, C. J. Lavender, E. Wishart, T. P. Stinear, J. A. Hayman, L. Brown, G. A. Jenkin, and J. A. Fyfe. 2007. *Mycobacterium ulcerans* in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia. *Emerg. Infect. Dis.* **13**:1653–1660.
 18. Kaser, M., and G. Pluschke. 2008. Differential gene repertoire in *Mycobacterium ulcerans* identifies candidate genes for patho-adaptation. *PLoS Negl. Trop. Dis.* **2**:e353.
 19. Kaser, M., S. Rondini, M. Naegeli, T. Stinear, F. Portaels, U. Certa, and G. Pluschke. 2007. Evolution of two distinct phylogenetic lineages of the emerging human pathogen *Mycobacterium ulcerans*. *BMC Evol. Biol.* **7**:177.
 20. Käser, M., M.-T. Ruf, J. Hauser, L. Marsollier, and G. Pluschke. 2009. Optimized method for preparation of DNA from pathogenic and environmental mycobacteria. *Appl. Environ. Microbiol.* **75**:414–418.
 21. Lavender, C. J., T. P. Stinear, P. D. Johnson, J. Azuolas, M. E. Benbow, J. R. Wallace, and J. A. Fyfe. 2008. Evaluation of VNTR typing for the identification of *Mycobacterium ulcerans* in environmental samples from Victoria, Australia. *FEMS Microbiol. Lett.* **287**:250–255.
 22. Marsollier, L., T. Séverin, J. Aubry, R. W. Merritt, J.-P. Saint André, P. Legras, A.-L. Manceau, A. Chauty, B. Carboneille, and S. T. Cole. 2004. Aquatic snails, passive hosts of *Mycobacterium ulcerans*. *Appl. Environ. Microbiol.* **70**:6296–6298.
 23. Marsollier, L., T. Stinear, J. Aubry, J. P. Saint André, R. Robert, P. Legras, A.-L. Manceau, C. Audrain, S. Bourdon, H. Kouakou, and B. Carboneille. 2004. Aquatic plants stimulate the growth of and biofilm formation by *Mycobacterium ulcerans* in axenic culture and harbor these bacteria in the environment. *Appl. Environ. Microbiol.* **70**:1097–1103.
 24. Mve-Obiang, A., R. E. Lee, F. Portaels, and P. L. Small. 2003. Heterogeneity of mycolactones produced by clinical isolates of *Mycobacterium ulcerans*: implications for virulence. *Infect. Immun.* **71**:774–783.
 25. Mve-Obiang, A., R. E. Lee, E. S. Umstot, K. A. Trott, T. C. Grammer, J. M. Parker, B. S. Ranger, R. Grainger, E. A. Mahrous, and P. L. Small. 2005. A newly discovered mycobacterial pathogen isolated from laboratory colonies of *Xenopus* species with lethal infections produces a novel form of mycolactone, the *Mycobacterium ulcerans* macrolide toxin. *Infect. Immun.* **73**:3307–3312.
 26. Pidot, S. J., H. Hong, T. Seemann, J. L. Porter, M. J. Yip, A. Men, M. Johnson, P. Wilson, J. K. Davies, P. F. Leadlay, and T. P. Stinear. 2008. Deciphering the genetic basis for polyketide variation among mycobacteria producing mycolactones. *BMC Genomics* **9**:462.
 27. Portaels, F., P. Elsen, A. Guimaraes-Peres, P. A. Fonteyne, and W. M. Meyers. 1999. Insects in the transmission of *Mycobacterium ulcerans* infection. *Lancet* **353**:986.
 28. Portaels, F., W. M. Meyers, A. Abiordey, A. G. Castro, K. Chemlal, P. de Rijk, P. Elsen, K. Fissette, A. G. Fraga, R. Lee, E. Mahrous, P. L. Small, P. Stragier, E. Torrado, A. Van Aerde, M. T. Silva, and J. Pedrosa. 2008. First cultivation and characterization of *Mycobacterium ulcerans* from the environment. *PLoS Negl. Trop. Dis.* **2**:e178.
 29. Ranger, B. S., E. A. Mahrous, L. Mosi, S. Adusumilli, R. E. Lee, A. Colorni, M. Rhodes, and P. L. Small. 2006. Globally distributed mycobacterial fish pathogens produce a novel plasmid-encoded toxic macrolide, mycolactone f. *Infect. Immun.* **74**:6037–6045.
 30. Rhodes, M. W., H. Kator, A. McNabb, C. Deshayes, J. M. Reyrat, B. A. Brown-Elliott, R. Wallace, Jr., K. A. Trott, J. M. Parker, B. Liffand, G. Osterhout, I. Kaattari, K. Reece, W. Vogelbein, and C. A. Ottinger. 2005. *Mycobacterium pseudoshottsii* sp. nov., a slowly growing chromogenic species isolated from Chesapeake Bay striped bass (*Morone saxatilis*). *Int. J. Syst. Evol. Microbiol.* **55**:1139–1147.
 31. Rondini, S., M. Käser, T. Stinear, M. Tessier, C. Mangold, G. Dernick, M. Naegeli, F. Portaels, U. Certa, and G. Pluschke. 2007. Ongoing genome reduction in *Mycobacterium ulcerans*. *Emerg. Infect. Dis.* **13**:1008–1015.
 32. Ross, B. C., P. D. Johnson, F. Oppedisano, L. Marino, A. Sievers, T. Stinear, J. A. Hayman, M. G. Veitch, and R. M. Robins-Browne. 1997. Detection of *Mycobacterium ulcerans* in environmental samples during an outbreak of ulcerative disease. *Appl. Environ. Microbiol.* **63**:4135–4138.
 33. Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
 34. Sizaïre, V., F. Nackers, E. Comte, and F. Portaels. 2006. *Mycobacterium ulcerans* infection: control, diagnosis, and treatment. *Lancet Infect. Dis.* **6**:288–296.
 35. Smith, N. H., S. V. Gordon, R. Rua-Domenech, R. S. Clifton-Hadley, and R. G. Hewinson. 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* **4**:670–681.
 36. Stinear, T., J. K. Davies, G. A. Jenkin, J. A. Hayman, F. Oppedisano, and P. D. R. Johnson. 2000. Identification of *Mycobacterium ulcerans* in the environment from regions in Southeast Australia in which it is endemic with sequence capture-PCR. *Appl. Environ. Microbiol.* **66**:3206–3213.
 37. Stinear, T. P., T. Seemann, P. F. Harrison, G. A. Jenkin, J. K. Davies, P. D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M. A. Quail, E. Rabinowitsch, D. Walker, B. White, S. Whitehead, P. L. Small, R. Brosch, L. Ramakrishnan, M. A. Fischbach, J. Parkhill, and S. T. Cole. 2008. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* **18**:729–741.
 38. Stinear, T. P., T. Seemann, S. Pidot, W. Frigui, G. Reysset, T. Garnier, G. Meurice, D. Simon, C. Bouchier, L. Ma, M. Tichit, J. L. Porter, J. Ryan, P. D. Johnson, J. K. Davies, G. A. Jenkin, P. L. Small, L. M. Jones, F. Tekaia, F. Laval, M. Daffe, J. Parkhill, and S. T. Cole. 2007. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res.* **17**:192–200.
 39. Ucko, M., A. Colorni, H. Kvitt, A. Diamant, A. Zlotkin, and W. R. Knibb. 2002. Strain variation in *Mycobacterium marinum* fish isolates. *Appl. Environ. Microbiol.* **68**:5281–5287.
 40. Wansbrough-Jones, M., and R. Phillips. 2006. Buruli ulcer: emerging from obscurity. *Lancet* **367**:1849–1858.
 41. Williamson, H. R., M. E. Benbow, K. D. Nguyen, D. C. Beachboard, R. K. Kimbirauskas, M. D. McIntosh, C. Quaye, E. O. Ampadu, D. Boakye, R. W. Merritt, and P. L. Small. 2008. Distribution of *Mycobacterium ulcerans* in Buruli ulcer endemic and non-endemic aquatic sites in Ghana. *PLoS Negl. Trop. Dis.* **2**:e205.
 42. Yip, M. J., J. L. Porter, J. A. Fyfe, C. J. Lavender, F. Portaels, M. Rhodes, H. Kator, A. Colorni, G. A. Jenkin, and T. Stinear. 2007. Evolution of *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria from a common *Mycobacterium marinum* progenitor. *J. Bacteriol.* **189**:2021–2029.