JB Accepts, published online ahead of print on 30 November 2012
J. Bacteriol. doi:10.1128/JB.02132-12

Complete Genome Sequence of the Frog Pathogen *Mycobacterium ulcerans* ecovar Liflandii

Nicholas J. Tobias[1,2], Kenneth D. Doig[1], Marnix H. Medema[3,4], Honglei Chen[5], Volker Haring[5], Robert Moore[5,6], Torsten Seemann[7] and Timothy P. Stinear[1#]

[1]Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria, Australia

[2]Department of Microbiology, Monash University, Clayton, Victoria, Australia

[3]Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands

[4]Department of Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands

[5]Australian Animal Health Laboratories, CSIRO, Geelong, Australia

[6]ARC Centre of Excellence in Structural and Functional Microbial Genomics, Australia

[7]Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria, Australia

Running Title: Complete Genome sequence of *Mycobacterium* Liflandii

[#]Corresponding author: tstinear@unimelb.edu.au

Abstract

In 2004, a previously undiscovered mycobacterium resembling *Mycobacterium ulcerans* (the agent of Buruli ulcer) was reported in an outbreak of a lethal mycobacteriosis in a laboratory colony of the African clawed-frog *Xenopus tropicalis*. This mycobacterium makes mycolactone and is one of several strains of *M. ulcerans*-like mycolactone-producing mycobacteria recovered from ectotherms around the world. Here we describe the complete 6,399,543 bp genome of this frog pathogen (previously unofficially named *Mycobacterium 'liflandii'*) and we show that it has undergone an intermediate degree of reductive evolution, between *M. ulcerans* Agy99 and the fish pathogen *Mycobacterium marinum* M. Like *M. ulcerans* Agy99, it has the pMUM mycolactone plasmid, over 200 chromosomal copies of the insertion sequence IS*2404* and a high proportion of pseudogenes. However, *M. 'liflandii'* has a larger genome that is closer in length, sequence and architecture to *M. marinum* M than to *M. ulcerans* Agy99, suggesting that *M. ulcerans* Agy99 has undergone accelerated evolution. Scrutiny of genes specifically lost suggests *M. 'liflandii'* is a tryptophan, tyrosine and phenylalanine auxotroph. A once extensive *M. marinum*-like secondary metabolome has also been diminished through reductive evolution. Our analysis shows that *M. 'liflandii'*, like *M. ulcerans* Agy99, has the characteristics of a niche-adapted mycobacterium but with several distinctive features in important metabolic pathways that suggests it is responding to different environmental pressures, supporting earlier proposals that it could be considered an *M. ulcerans* ecotype, hence the name *M. ulcerans* ecovar Liflandii.

Introduction

An interesting member of the *Mycobacterium marinum* and *Mycobacterium ulcerans* complex (MuMC) was discovered in the summer of 2001 when an outbreak of generalized cutaneous lesions developed in a colony of *Xenopus tropicalis* at the University of Berkeley, California (1). Infected frogs developed granulomatous skin lesions, with coelomic distention, generalised oedema and septicaemia (1). Cytological examination confirmed the presence of acid-fast bacilli in smears from the liver, spleen, kidney and skin. Based on histopathology and some molecular testing, it was concluded that these frogs were suffering from a mycobacteriosis caused by a *Mycobacterium ulcerans*-like bacterium (1). There are on-going reports and high lethality of this disease in captive frogs across the United States (US) and those imported from the US to Europe (2-4), but this mycobacterium has not been reported to cause illness in humans. Further characterisation of the frog pathogen revealed that it harboured the *M. ulcerans* pMUM megaplasmid and produced mycolactone E, a unique structural variant of the polyketide toxin that is key for pathogenesis in *M. ulcerans* (5, 6). Limited genotype analysis suggests that a single clone of this pathogen is circulating worldwide in institutions housing and breeding anurans. The species name *Mycobacterium liflandii* was proposed, although as the authors admitted at that time, the data were lacking to conclude that this mycobacterium constituted a separate species (6).

A recent genomic study of 35 MuMC isolates confirmed earlier indications that all mycolactone-producing mycobacteria, including those strains classically considered *M. ulcerans*, evolved by a process of lateral gene transfer and reductive evolution from a common *M. marinum*-like progenitor (7-9). These comparisons also show at least three discrete, deep-branching lineages of mycolactone producers, which include the strains predominantly infecting ectotherms and those strains commonly

causing Buruli ulcer in humans (7-9). All lineages showed evidence of strong selective pressures acting on the same cell wall-associated genes, but the lineages differed in the extent of genome reduction, suggesting each lineage might be adapting to slightly different niche environments (8). The frog pathogen belongs to lineage 1, while *M. ulcerans* isolates from Africa and Australia belong to lineage 3. Hereafter we will refer to this frog pathogen as *M. ulcerans* Liflandii (abbreviated as *Mu*L) in line with our proposition that *Mu*L and other lineage-specific isolates should be considered ecotypes of *M. ulcerans* (8, 10).

Like other MuMC members, *Mu*L grows preferentially around 32°C. However, it exhibits several distinguishing microbiological characteristics. *Mu*L forms rough, light orange, non-photochromogenic colonies on Middlebrook 7H11 agar supplemented with oleic acid, albumin, dextrose and catalase (37, 56). *Mu*L grows better on charcoal medium than Lowenstein-Jensen (LJ) medium whereas classical (lineage 3) *M. ulcerans* isolates do not grow on charcoal. Like *M. ulcerans*, antibiograms of *Mu*L indicate resistance to isoniazid, ethambutol and ethionamide but *Mu*L is reportedly additionally resistant to rifampicin and clarythromycin (53).

To date the MuMC are represented by only two fully assembled genome sequences; an *M. ulcerans* isolate from a Buruli ulcer patient isolated in Ghana in 1999 and a *M. marinum* clinical isolate obtained from a patient in the United States in 1994 (11, 12). The high complexity of *M. ulcerans* genomes with the repeat-rich nature of the mycolactone polyketide synthase genes (harboured on pMUM plasmid) and >200 chromosomal copies of IS*2404* has been a barrier to completing assemblies of genome sequences from other members of this complex. Here we describe and compare the sequenced and fully assembled chromosome of *Mu*L, which together with our previous description of the pMUM002

mycolactone plasmid, represents the first complete genome for the strains of *M. ulcerans* that are increasingly associated with epizoonotics in fish, frogs and other ectotherms around the world.

Materials and Methods

Strain and culture conditions

M. ulcerans Liflandii 128FXT was originally isolated from infected *Xenopus tropicalis* at the University of California, Berkeley (1). The isolate was cultured on Brown and Buckle egg yolk agar slopes at 30ºC. Rifampicin MIC testing was performed using Mycobacterial Growth Indicator Tubes (MGIT) as described (13). Briefly, *Mu*L at a McFarland standard of 0.5 was prepared as recommended by the manufacturer (Becton Dickinson) and diluted 1:5 in sterile saline. MGITs were enhanced with 800 µl of MGIT SIRE supplement (Becton Dickinson) and inoculated with 0.5 ml of the diluted bacterial suspension. MGITs contained two-fold dilutions of rifampicin from 0.0625 µg/ml to 16 µg/ml. Experiments were established in triplicate for each dilution. MGITs were incubated at 30ºC and examined daily for four weeks at a wavelength of 365nm, noting the time to fluorescence.

DNA Methods

Genomic DNA was extracted from *Mu*L as previously described (9). Standard methods were used for PCR and Sanger sequencing.

Whole Genome Sequencing and Assembly

Roche 454 GS-FLX sequencing was employed to obtain 260,654 single-end reads (106.3 Mbp) and assembled using gsAssembler v2.5.3 into 91 scaffolds containing a total of 622 contigs. Scaffolds were ordered by reference to an optical map (see below) and BLAT searches in Projector 2 against the *M.*

marinum M and *M. ulcerans* Agy99 reference genomes (14). Genome finishing was managed with Gap4 (15). Specific oligonucleotides were designed to contig-flanking sequences and PCR with Sanger sequencing was used to close gaps. Wherever it was established that a gap contained a single copy of IS*2404,* the gap was filled with a copy of IS*2404* from the previously sequenced *M. ulcerans* Agy99. It is important to note therefore that the exact sequence of each copy of IS*2404* in *Mu*L has not been established. A 3 kbp Roche 454 GS-FLX mate-pair library was also constructed and yielded 218,423 reads (63.6 Mbp). Finally, 13,778,118 Illumina paired-end sequencing reads (487.61Mbp) obtained from a previous project (8) were mapped to the final assembly using an in-house Python utility called Nesoni (http://vicbioinformatics.com ; Harrison *et al*., in preparation) with SHRiMP 2.2 (51). Nesoni was used to do a global variant analysis and generate a list of differences and correct 454 sequencing errors. The chromosome sequence of *Mu*L 128FXT was submitted to Genbank under BioProject number PRNJA128960.

Optical Mapping

An NheI optical map of *Mu*L was prepared by OpGen (Wisconsin, USA) to guide assembly. Contigs were aligned to the optical map using MapSolver software (v3.2.0).

Annotation and Comparative Genome Analysis

Genome annotation was performed by Prokka (http://vicbioinformatics.com/; Seemann, in preparation), utilising the previously annotated *M. marinum* M and *M. ulcerans* Agy99 genomes as references (11, 12). Manual curation of the annotation was then performed using Wasabi, a web-based annotation editor and database as described previously (11). Regions of difference (RDs) were identified by mapping *M. ulcerans* Agy99 and *M. marinum de novo* contigs to the assembled *Mu*L genome. This mapping and other comparisons were visualized with Circos (16), Mauve (17) and ACT (18). Secondary metabolite biosynthesis gene clusters were identified using antiSMASH (19). Additional pseudogene clusters that were too degraded for detection by antiSMASH were identified by

MultiGeneBlast (http://multigeneblast.sourceforge.net/; Medema *et al*., in preparation). The final, manually determined, cut-off used to define gene cluster orthology was that gene clusters should have at least 40% genes with at least 60% amino acid sequence identity. All gene clusters connected by orthology were grouped together into a single gene cluster family. Chromosomal maps of biosynthetic gene clusters were generated with in-house Python scripts using PySVG (http://codeboje.de/pysvg/).

Results and Discussion

General feature of the *M. ulcerans* Liflandii genome

M. ulcerans Liflandii has a 6,399,542 bp genome comprising a single circular 6,208,954 bp chromosome and the previously described 190,588 bp pMUM002 mycolactone megaplasmid, with a G+C content of 65.61% and 62.9% respectively (20). The chromosome is predicted to contain 4,994 protein coding DNA sequences (CDS) and 436 pseudogenes (Fig. 1). One rRNA operon and 50 tRNA genes were identified (Table 1). Sequence assembly was complicated by the presence of 231 chromosomal copies of the earlier described *M. ulcerans* insertion sequence IS*2404* (208 complete copies) (Fig. 1). The integrity of the finished sequence was verified by reference to a high resolution NheI whole-chromosome restriction map (Fig. 2A).

Comparisons of the complete chromosomal sequences of *M. marinum* M and *M. ulcerans* Agy99 against *Mu*L identified a core genome of 4,237 CDS, which is 85% of total *Mu*L coding sequences (Fig. 2B,C). The genetic functional group distribution amongst the three sequenced genomes did not differ significantly across 'intermediary metabolism and respiration', 'cell wall and cell processes', 'conserved hypothetical' or 'lipid metabolism' classes (Fig. 3A). Nucleotide identity amongst the three-way genome comparison of a subset of 3,391 strict core orthologs (excludes paralogs and pseudogenes) was greater than 97% and showed that the highest sequence similarities are between *Mu*L and *M.*

marinum (Fig. 2B). Similarly, the chromosome architecture of *Mu*L is more closely related to *M. marinum* M (Fig. 2B), with few genome rearrangements and few large DNA deletions such as those seen in the reduced *M. ulcerans* Agy99 genome. These data are not in contradiction with previous genomic comparisons that indicate a common ancestor for the three lineages of mycolactone-producing mycobacteria, including *Mu*L (lineage 1) and *M. ulcerans* Agy99 (lineage 3), and suggest that lineage 3 isolates have undergone accelerated evolution through additional niche adaptation, while *Mu*L has retained a more pleomorphic state.

Insertion sequences and phage elements

Isolates from all lineages of *M. ulcerans* have more than 200 chromosomal copies of IS*2404* (9, 10, 21). This element is known to have promoted, at least in part, significant genomic rearrangements in *M. ulcerans* Agy99, including many large DNA deletions and gene disruptions by insertion (22). The *Mu*L chromosome is 430 kb smaller than *M. marinum*, but considering that IS*2404* sequences contribute another 305 kb, it appears that approximately 780 kb has been lost since diverging from a *M. marinum*-like ancestor. Arguably, this difference is largely due to IS*2404*-mediated deletions. Also, 89 of the 224 chromosomal copies of IS*2404* (40%) have inserted within CDS. This is comparable to the situation in *M. ulcerans* Agy99, where 97 CDS have been disrupted by IS*2404* (45% of the 213 copies), although only six copies of IS*2404* appear in the same location in both isolates, indicating independent expansion of IS*2404* in each lineage. Most startling, however, is that despite the obvious activity of IS*2404* in *Mu*L, the 224 copies of this element have had an unexpectedly modest impact on overall chromosome architecture. There are only four major DNA rearrangements in *Mu*L relative to *M. marinum* M, (all flanked by copies of IS*2404*), such as a 440 kb inversion between nucleotide positions 2,891,467 and 3,144,394. These four regions are unchanged in *M. ulcerans* Agy99 relative to *M. marinum* M. The conserved chromosome structure of *Mu*L suggests positive selection for this

chromosome arrangement, despite the expansion of IS*2404*. One conclusion from these observations is that both strains are undergoing independent processes of reductive evolution, where expansion of IS*2404* has been equally tolerated but with strong selection for maintenance of chromosome structure in *Mu*L.

There are 91 copies of IS*2606* in *M. ulcerans* Agy99 (11, 23). *Mu*L has only 4 copies of IS*2606*, three of which are on pMUM002 (20). The presence of a copy of IS*2606* in the same location on both pMUM001 and pMUM002 suggests this IS originated on pMUM and expanded in lineage 3 (20). The single IS*2606* (Mlif_03910) chromosomal copy is in stark contrast to the 83 copies in *M. ulcerans* Agy99 and also highlights the likely role that IS*2606* has played in promoting the extensive chromosome remodelling seen in the lineage 3 isolates, in particular the preference for IS*2606* to insert in close proximity to IS*2404*. Of the 83 chromosome copies of IS*2606*, there are 39 instances of IS*2606* inserting within 100 bp of IS*2404* (59 copies within 500bp) and the combination of IS*2606* and IS*2404* are associated with at least 30 instances of inversions and/or deletions (>5000bp) in *M. ulcerans* Agy99. It is not clear why IS*2606* has not similarly expanded in *Mu*L.

Two mycobacteriophages, phiMU01 (18 kb) and phiMU02 (24 kb), are variously present in all *M. ulcerans* but not in *M. marinum* M (8). Both phages are present in *Mu*L, yet phiMU02 is significantly smaller (13 kb) and phiMU02 CDSs disrupted by IS*2606* in *M. ulcerans* Agy99 are mostly deleted or have acquired frameshift mutations in *Mu*L.

M. ulcerans Liflandii regions of difference

To further explore genetic features that might help explain the specific phenotypes of *Mu*L, we examined regions of DNA present only in *Mu*L compared to the other two genomes. In the original

204 description of the *M. ulcerans* Agy99 genome regions of difference (MURDs) were stretches of DNA
205 present in *M. marinum* M but absent from *M. ulcerans* Agy99 (11). Here we have defined *Mu*L regions
206 of difference (MULiRD) as regions present only in *Mu*L compared with *M. ulcerans* and *M. marinum*.
207 Eleven MULiRDs, spanning 290 CDSs were identified in the genome (Tables S1 & S2). The two
208 largest regions, MULiRD3 (73.9 kb) and MULiRD7 (18.3 kb) (Fig. 1), harbour CDSs with possible
209 roles in secondary metabolism. Both regions are flanked by IS*2404* elements and it appears that in *M.*
210 *ulcerans* Agy99 these regions were lost by ISE-mediated deletion, where a single copy of IS*2404*
211 remains in these locations. MULiRD7 harbours a seven-gene *hyc* operon (Mlif_03568-Mlif_03574)
212 that is duplicated on the chromosome (Mlif_01806 - Mlif_01813), although *hycQ* (Mlif_01811) is a
213 predicted pseudogene. *M. ulcerans* Agy99 only has one copy of this operon with the associated
214 transcriptional regulator (MUL_1889) disrupted by a copy of IS*2404* and *hycE* (MUL_1896)
215 containing a 317 amino acid C-terminal truncation. The orthologous *hyc* locus in *M. tuberculosis* is
216 thought to encode a formate hydrogenylase complex that is part of a dormancy regulatory network
217 involving MprA and DevR, where in *M. tuberculosis* the ortholog of the transcriptional regulator in this
218 locus (Mlif_1806) is upregulated in response to increased concentrations of nitric oxide (24). The
219 preservation and duplication of this system in *Mu*L suggests a similar dormancy response may be
220 important for the lifestyle of the frog pathogen.
221
222 **Pseudogene composition of *M. ulcerans* Liflandii**
223 Approximately 8% of all CDSs annotated in *Mu*L were predicted to be pseudogenes. Pseudogenes in
224 prokaryote genomes generally occupy 1-5% of all CDSs (25). However, obligate pathogens are thought
225 to be an exception with higher levels of gene inactivation (25). CDSs associated with lipid metabolism
226 (11.5%), cell wall and cell processes (9.3%), and intermediary metabolism-associated proteins (8.0%)
227 were overrepresented with pseudogenes (Fig. 3B). In comparison, *M. ulcerans* Agy99 has 19.8%,

15.5% and 13.4% pseudogenes in the same classes, respectively. The percentage of total coding sequences in *Mu*L that are pseudogenes is intermediate between *M. marinum* and *M. ulcerans* in all families of proteins (Fig. 3B) with 169 pseudogenes conserved in both *M. ulcerans* Agy99 and *Mu*L.

Unique to mycobacteria, the cell envelope-associated PE/PPE proteins have been suggested to modulate host immune responses, among other potential functions (26). High DNA identity in the 3' region of these genes is a substrate for recombination and thought to provide a source of antigenic variation among the mycobacteria (27, 28). The highly reduced genome of *M. leprae* contains very few intact CDSs of this family (29). In *M. ulcerans* Agy99, 48.7% of all PE/PPE CDSs are pseudogenes. A significant proportion of these CDSs are also pseudogenes in the *Mu*L genome (28.8%), compared with only 2.2% in *M. marinum* M.

Mutation and inactivation of certain CDSs (pseudogene formation) in *M. ulcerans* Agy99 is predicted to have caused some significant phenotypic changes. A frameshift mutation in *M. ulcerans* Agy99 has resulted in a disruption of *cydA*, a component of the cytochrome *bd* oxidase transporter. However, in *Mu*L *cydA* is intact. This system is involved in response to anaerobic and hypoxic conditions *in vitro* in *M. tuberculosis* (30). Conservation of this locus in *Mu*L may indicate an increased ability of the frog mycobacterium to survive under low oxygen conditions, although, like *M. ulcerans* Agy99, the selenocysteine-containing formate dehydrogenase complex, with a predicted role in anaerobiosis, is likely to be inactive in *Mu*L, suggesting it too has an impaired anerobic respiration capacity.

Phenolic glycolipids (PGLs) are potent antigens and virulence factors produced by mycobacterial pathogens. PGLs are composed of a polyketide backbone, decorated with species-specific combinations of sugar(s) via a phenolic head group. Genome analysis suggests *Mu*L produces the same

PGL as *M. marinum* M, (sometimes called mycoside G). This is distinct to *M. ulcerans* Agy99, where two genes (MUL_1998 and MUL_2001) have been inactivated by mutation resulting in the synthesis of an aglycosylated molecule with a modified polyketide backbone. These two genes are intact in *MuL* (Mlif_1910 and Mlif_1913). The predicted presence of intact PGL in *Mu*L might have implications for interactions between host and bacteria and again points to likely differences in the lifestyle of *M. ulcerans* Agy99 compared to *Mu*L (11).

ESX loci

Mycobacterial intracellular pathogens such as *M. tuberculosis* and *M. marinum* have at least five ESX (or Type VII) ATP-dependent protein secretory systems named ESX-1, 3, 4, 5 & 6. The best defined of these systems, ESX-1, has been implicated in virulence via secretion of certain effectors including the antigens ESAT-6 and CFP-10 (31). *Mu*L has six predicted ESX loci. ESX-1 (6138025 - 6171960) appears intact in *Mu*L, although there are two mutations that might impact function. A copy of IS*2404* has inserted within the intergenic region between the divergently transcribed *eccCb1* and the PE35 ortholog (Mlif_05720) and this may impact expression of either gene. Also of note, Mlif_05724, the gene immediately downstream of *esxA*, is a pseudogene, although the *M. tuberculosis* ortholog (Rv3876) is thought to be non-essential for ESX function (32). ESX-2, immediately downstream of ESX-1, shares the same arrangement as in *M. marinum* M and might be inactive, whereas Mlif_05735 (the ortholog of MMAR_5460) also appears truncated. The 10 kb region immediately downstream of ESX-2 is disrupted by four copies of IS*2404* and thus distinct to *M. marinum*. ESX-3 (218100..230400), ESX-4 (1319227..1332538) and ESX-5 (2592830..2614870) all appear intact, although ESX-5 has a single copy of IS*2404* between Mlif_02416 and Mlif_02418. ESX-6 (182073..187394) is incomplete with a 7 kb deletion compared to the 12 kb version in *M. marinum M*, although *esxB_2* and *esxA_2* in this locus remain intact. In *M. marinum* M but not in other pathogenic

mycobacteria, the region immediately upstream of ESX-1 contains 11 paralogous predicted membrane proteins of unknown function. This 15.7 kb region is deleted in *Mu*L, having been replaced with a single copy of IS*2404*. This region is also deleted in *M. ulcerans* Agy99. Eleven ESX-1 secretion associated proteins (Esp) are present in *M. marinum* with only four of these proteins intact in the *M. ulcerans* Agy99 genome. *Mu*L has mostly conserved these proteins with 10 intact Esp paralogs and one (Mlif_04556) pseudogenized.

Lipoproteins

Lipoproteins in mycobacteria have been implicated in signal transduction (33), evasion of mammalian cells (34) and some have a direct role in virulence as a part of transport systems (35). These proteins can be surface-exposed and anchored by hydrophobic interactions, potentially to mycolic acids within the cell wall (36, 37). Serine proteases like SppA are responsible for hydrolyzing signal peptides prior to export across the cytoplasmic membrane (38). Pre-lipoproteins are then acylated after export from the cytoplasmic membrane by Lgt prior to cleavage by Lsp and Lnt (39). Lipoprotein synthesis appears to be considerably disrupted in *Mu*L as *sppA* is a pseudogene. In *E. coli* the *sppA* homolog specifically cleaves the signal peptide of a major lipoprotein (40). The predicted inability of *Mu*L to cleave N-terminal signal peptides from lipoproteins is likely to significantly hamper pre-processing of lipoproteins. The *lgt* gene is also likely to be inactivated in *Mu*L (8). While not well studied in mycobacteria, prolipoprotein acylation by Lgt is not essential for cleavage by Lsp in different Gram-positive bacteria (41). Despite this, the absence of SppA and Lgt in these *M. ulcerans* strains may provide a mechanism to reduce the lipoprotein-induced TLR2 response as described in other Gram-positive pathogens (41, 42).

Unique Metabolic features of *M. ulcerans* Liflandii

Genome inspection and metabolic pathway analysis suggests *Mu*L may have some distinctive phenotypic characteristics. Genes encoding DAHP synthases, which are important in the first step of the shikimate enzyme pathway, are pseudogenes. DAHP synthases are responsible for converting erythrose-4-phosphate into 3-deoxy-D-arabino-heptulosonate-7-phosphate during chorismate synthesis (43). Paralogs, *aroG* (Mlif_02016) and *aroG_1* (Mlif_03449) are both disrupted, whilst in *M. marinum* M there are intact copies of both *aroG* (MMAR_3222) and *aroG_1* (MMAR_1854). In *M. ulcerans* *aroG* (MUL_2100) is a pseudogene but *aroG_1* (MUL_3533) is intact. Chorismate is an essential precursor for the synthesis of the aromatic amino acids tyrosine, phenylalanine and tryptophan (43) and in *M. tuberculosis*, this pathway is essential for survival (44, 45). The lack of intact *aroG* suggests that *Mu*L may be a tryptophan, tyrosine and phenylalanine auxotroph, and inhabit an environment where these amino acids are available. Experiments to confirm this predicted auxotrophy using Sauton's media supplemented with aromatic amino acids has so far been unsuccessful.

Antibiotic Susceptibility of *M. ulcerans* Liflandii

In *M. tuberculosis* there is a correlation between isoniazid/ethambutol resistance and the presence of an intact *iniA* gene, as *iniA* is part of an operon proposed to encode an efflux pump involved in resistance to a wide range of antibiotics that target cell wall biosynthesis (46, 47). Upon deletion of *iniA*, *M. tuberculosis* shows increased susceptibility to isoniazid (47). In *Mu*L, all three genes *iniA, iniB* and *iniC* are pseudogenes, whereas these genes are intact in both *M. marinum* and *M. ulcerans* Agy99, suggesting other genetic differences in *Mu*L will explain its reported increased resistance to isoniazid and ethambutol compared to *M. ulcerans* Agy99 (2).

M. ulcerans lineage 3 strains are sensitive to rifampicin, which is used in combination with streptomycin to treat Buruli ulcer. In contrast, *Mu*L is reported to be resistant to rifampicin (2).

Examination of the genome reveals a single amino acid substitution in *Mu*L *rpoB* (T713M) compared to *M. marinum* and *M. ulcerans*. To our knowledge, the mutation has not been previously reported but it may represent a novel rifampicin resistance mutation. We therefore conducted MIC testing and found *Mu*L 128FXT was fully susceptible to rifampicin (MIC <0.0625 μg/mL), indicating the T713M mutation does not cause rifampicin resistance. The discrepancy with the previous report remains to be explained.

Secondary metabolism in *M. ulcerans* Liflandii

Mycobacteria have diverse secondary metabolite repertoires that include toxins, siderophores and complex cell wall lipids (48). *M. marinum* M in particular has one of the largest arrays of secondary metabolite gene clusters yet described among bacteria (12) , although the metabolites produced by most of these clusters are unknown. We explored the secondary metabolome of *Mu*L using the biosynthetic gene cluster search tool antiSMASH (19), and detected 28 distinct secondary biosynthesis clusters on the chromosome. This compares with 33 in *M. marinum* M, 11 in *M. ulcerans* Agy99, and 15 in *M. tuberculosis* F11. To try to compare the secondary biosynthetic potential between these mycobacteria, we classified all 88 antiSMASH-identified gene clusters into 37 families based on their sequence homology (Table S3). *Mu*L has 10 more gene clusters than *M. ulcerans* Agy99, while the number of pseudogenized gene clusters is similar in both strains (Fig. 4). This analysis reflected the general trend already observed, that lineage 3 *M. ulcerans* has proceeded further along a reductive evolutionary trajectory than lineages 1 and 2, given that in the genomes of both species numerous gene clusters appear to have been deactivated by pseudogene formation or have been lost by deletion (Fig. 4A).

Neither *Mu*L nor *M. ulcerans* Agy99 has any gene cluster that is not observed in the *M. marinum* M genome, and all gene clusters of *M. ulcerans* Agy99 (whether intact or pseudogenized) are also present

in *Mu*L (Fig. 4C). This is the expected state for strains that have evolved from a common *M. marinum*-like ancestor. However, there is one intact *M. ulcerans* non-ribosomal peptide synthetase (NRPS) gene cluster that is pseudogenized in *Mu*L (Mlif_01390 - Mlif_01430) and eight intact *Mu*L gene clusters that have been pseudogenized in *M. ulcerans*, indicating that a significant part of reductive biosynthetic evolution has taken place independently in both strains and that the products of these loci are not required by the respective specialized bacteria. Mapping the chromosomal positions of the intact and the pseudogene clusters from *Mu*L and *M. ulcerans*, we observed that gene cluster inactivation has primarily occurred on the leading strand of the right hand replichore, closer to the origin of replication than the terminus (Fig. 5). In both *Mu*L and *M. ulcerans*, seven out of the eight gene clusters, closest to the origin of replication on the leading strand, have been inactivated by pseudogene formation (71% of all pseudogene clusters), even though only two gene clusters are shared in this region between the two genomes. This suggests a general phenomenon and might be explained by the fact that in other bacterial species the same chromosomal region contains the most highly expressed genes (49).

M. ulcerans ecovar Liflandii

There is no consensus on what defines an ecotype, but it has been suggested that an ecotype is a bacterial strain that conserves the genetic potential of a species with genetic differences allowing it to exploit a slightly different ecological niche (50). As discussed previously, *Mu*L fulfils the criteria required for classification within the species *M. ulcerans* and this is also true for the other mycolactone-producing mycobacteria (9, 10). The comparative genome data we have presented here show that *Mu*L like all *M. ulcerans* strains is undergoing reductive evolution and is likely also adapting to a niche environment. However, the pattern of mutations, the conserved arrangement of the chromosome and other features such as the distinctive structure and activity of mycolactone E, suggest

371 *Mu*L is responding to a different set of environmental pressures compared to *M. ulcerans* Agy99. These
372 lines of evidence lead us to propose that *Mu*L is an ecotype within the species *M. ulcerans*.
373
374
375 **Conclusions**
376 *M. ulcerans* Liflandii is a member of the MuMC. The data presented here represent the third complete
377 genome sequence for a mycobacterium from this complex. Like *M. ulcerans*, the frog mycobacterium
378 has the signature of a niche-adapted organism and contains the pMUM virulence plasmid, several
379 hundred copies of the *M. ulcerans*-specific insertion sequence IS*2404* and many pseudogenes.
380 However, *Mu*L lacks the dramatic DNA rearrangements and deletions seen in *M. ulcerans* and has a
381 chromosome architecture more closely aligned with *M. marinum*. *Mu*L shares a large gene repertoire
382 with other members of the MuMC. However, specific mutations in key metabolic pathways such as the
383 *aroG/aroG_1* pseudogenes may have a profound impact regarding the environments in which *Mu*L can
384 survive. As with *M. marinum,* ESX cell wall secretion systems appear to be largely intact in *Mu*L.
385 However, other components of the cell wall are characteristically distinctive, with lipoprotein
386 processing likely to be significantly hampered with non-functional *lgt* and *sppA*, dissimilar to both *M.*
387 *ulcerans* Agy99 and *M. marinum* M. We propose that *Mu*L is an ecotype of *M. ulcerans*, and is
388 adapting to a niche that is related but distinct to other *M. ulcerans* lineages. The genome sequence of
389 *Mu*L provides an important snapshot of short-term reductive evolution as highlighted by the impact on
390 secondary metabolite biosynthesis gene cluster repertoires. Comparisons with the genomes of *M.*
391 *ulcerans* and *M. marinum* show that such repertoires can change very rapidly, and that they can leave
392 distinct genomic scars, which remain visible for some time and can be uncovered when a closely
393 related genome is available for comparison. Future research could be aimed at better understanding the
394 microbiology (*e.g.* physiology, biochemistry and antibiotic susceptibility) of *M. ulcerans* Liflandii, to

try and ascribe the genomic features and predictions made here to confirmed phenotypes for this unusual pathogen.

Acknowledgements

This work was supported in part by grants from the National Health and Medical Research Council of Australia (628640) and the Dutch Technology Foundation STW (STW10463). We are grateful to Rainer Breitling and Eriko Takano for constructive comments that helped improve the manuscript.

References

1. **Trott, K. A., B. A. Stacy, B. D. Lifland, H. E. Diggs, R. M. Harland, M. K. Khokha, T. C. Grammer, and J. M. Parker.** 2004. Characterization of a *Mycobacterium ulcerans*-like infection in a colony of African tropical clawed frogs (*Xenopus tropicalis*). Comp Med **54:**309-317.

2. **Nigou, J., M. Gilleron, and G. Puzo.** 2003. Lipoarabinomannans: from structure to biosynthesis. Biochimie **85:**153-166.

3. **Fremont-Rahl, J. J., C. Ek, H. R. Williamson, P. L. Small, J. G. Fox, and S. Muthupalani.** *Mycobacterium liflandii* outbreak in a research colony of *Xenopus* (Silurana) tropicalis frogs. Vet Pathol **48:**856-867.

4. **Chai, N., O. Bronchain, G. Panteix, S. Godreuil, C. de Medeiros, R. Saunders, T. Bouts, and A. de Luze.** Propagation method of saving valuable strains from a *Mycobacterium liflandii* infection in Western clawed frogs (*Silurana tropicalis*). J Zoo Wildl Med **43:**15-19.

5. **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 (Camb)**:**4306-4308.

6. **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.

7. **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.

8. **Doig, K. D., K. E. Holt, J. A. Fyfe, C. J. Lavender, M. Eddyani, F. Portaels, D. Yeboah-Manu, G. Pluschke, T. Seemann, and T. P. Stinear.** 2012. On the origin of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. BMC Genomics **13:**258.

9. **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.

10. **Pidot, S. J., K. Asiedu, M. Kaser, J. A. Fyfe, and T. P. Stinear.** 2010. *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria should be considered a single species. PLoS Negl Trop Dis **4:**e663.

11. **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.

12. **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. Rabbinowitsch, 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.

13. **Heysell, S. K., C. Mtabho, S. Mpagama, S. Mwaigwisya, S. Pholwat, N. Ndusilo, J. Gratz, R. E. Aarnoutse, G. S. Kibiki, and E. R. Houpt.** 2011. Plasma drug activity assay for treatment optimization in tuberculosis patients. Antimicrob Agents Chemother **55:**5819-5825.

14. **van Hijum, S. A., A. L. Zomer, O. P. Kuipers, and J. Kok.** 2005. Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. Nucleic Acids Res **33:**W560-566.

15. **Bonfield, J. K., K. Smith, and R. Staden.** 1995. A new DNA sequence assembly program. Nucleic Acids Res **23:**4992-4999.

16. **Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra.** 2009. Circos: an information aesthetic for comparative genomics. Genome Res **19:**1639-1645.

17. **Darling, A. E., B. Mau, and N. T. Perna.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE **5:**e11147.

18. **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.

19. **Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, and R. Breitling.** 2011. antiSMASH: rapid identification, annotation and

analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res **39:**W339-346.

20. **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.

21. **Kaser, M., and G. Pluschke.** 2008. Differential gene repertoire in *Mycobacterium ulcerans* identifies candidate genes for patho-adaptation. PLoS Negl Trop Dis **2:**e353.

22. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol Mol Biol Rev **62:**725-774.

23. **Stinear, T., B. C. Ross, J. K. Davies, L. Marino, R. M. Robins-Browne, F. Oppedisano, A. Sievers, and P. D. Johnson.** 1999. Identification and characterization of IS*2404* and IS*2606*: two distinct repeated sequences for detection of *Mycobacterium ulcerans* by PCR. J Clin Microbiol **37:**1018-1023.

24. **Voskuil, M. I., D. Schnappinger, K. C. Visconti, M. I. Harrell, G. M. Dolganov, D. R. Sherman, and G. K. Schoolnik.** 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. The Journal of experimental medicine **198:**705-713.

25. **Moran, N. A.** 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell **108:**583-586.

26. **Ramakrishnan, L., N. A. Federspiel, and S. Falkow.** 2000. Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. Science **288:**1436-1439.

27. **Bottai, D., and R. Brosch.** 2009. Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. Mol Microbiol **73:**325-328.

28. **Delogu, G., and M. J. Brennan.** 2001. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. Infect Immun **69:**5606-5611.

29. **Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell.** 2001. Massive gene decay in the leprosy bacillus. Nature **409:**1007-1011.

30. **Shi, L., C. D. Sohaskey, B. D. Kana, S. Dawes, R. J. North, V. Mizrahi, and M. L. Gennaro.** 2005. Changes in energy metabolism of *Mycobacterium tuberculosis* in mouse lung and under in vitro conditions affecting aerobic respiration. Proc Natl Acad Sci U S A **102:**15629-15634.

31. **Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. M. Collins, S. L. Morris, C. H. King, and W. R. Jacobs, Jr.** 2003. The primary mechanism of

attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A **100:**12420-12425.

32. **Brodin, P., L. Majlessi, L. Marsollier, M. I. de Jonge, D. Bottai, C. Demangel, J. Hinds, O. Neyrolles, P. D. Butcher, C. Leclerc, S. T. Cole, and R. Brosch.** 2006. Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. Infect Immun **74:**88-98.

33. **Steyn, A. J., J. Joseph, and B. R. Bloom.** 2003. Interaction of the sensor module of *Mycobacterium tuberculosis* H37Rv KdpD with members of the Lpr family. Mol Microbiol **47:**1075-1089.

34. **Casali, N., and L. W. Riley.** 2007. A phylogenomic analysis of the Actinomycetales *mce* operons. BMC Genomics **8:**60.

35. **Camacho, L. R., D. Ensergueix, E. Perez, B. Gicquel, and C. Guilhot.** 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol Microbiol **34:**257-267.

36. **Kovacs-Simon, A., R. W. Titball, and S. L. Michell.** 2011. Lipoproteins of bacterial pathogens. Infect Immun **79:**548-561.

37. **Wu, C. H., J. J. Tsai-Wu, Y. T. Huang, C. Y. Lin, G. G. Lioua, and F. J. Lee.** 1998. Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS letters **439:**192-196.

38. **Hussain, M., Y. Ozawa, S. Ichihara, and S. Mizushima.** 1982. Signal peptide digestion in *Escherichia coli.* Effect of protease inhibitors on hydrolysis of the cleaved signal peptide of the major outer-membrane lipoprotein. Eur J Biochem **129:**233-239.

39. **Okuda, S., and H. Tokuda.** 2011. Lipoprotein sorting in bacteria. Annu Rev Microbiol **65:**239-259.

40. **Ichihara, S., N. Beppu, and S. Mizushima.** 1984. Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. J Biol Chem **259:**9853-9857.

41. **Henneke, P., S. Dramsi, G. Mancuso, K. Chraibi, E. Pellegrini, C. Theilacker, J. Hubner, S. Santos-Sierra, G. Teti, D. T. Golenbock, C. Poyart, and P. Trieu-Cuot.** 2008. Lipoproteins are critical TLR2 activating toxins in group B streptococcal sepsis. J Immunol **180:**6149-6158.

42. **Machata, S., S. Tchatalbachev, W. Mohamed, L. Jansch, T. Hain, and T. Chakraborty.** 2008. Lipoproteins of *Listeria monocytogenes* are critical for virulence and TLR2-mediated immune activation. J Immunol **181:**2028-2035.

43. **Pittard, A. J.** 1987. Biosynthesis of the aromatic amino acids., p. 368–394. *In* J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), Escherichia coli and Salmonella typhimurium : Cellular and Molecular Biology. American Society for Microbiology, Washington, DC.

44. **Barker, C., and D. Lewis.** 1974. Impaired regulation of aromatic amino acid synthesis in a mutant resistant to p-fluorophenylalanine. J Gen Microbiol **82:**337-343.

45. **Parish, T., and N. G. Stoker.** 2002. The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. Microbiology **148:**3069-3077.

46. **Alland, D., A. J. Steyn, T. Weisbrod, K. Aldrich, and W. R. Jacobs, Jr.** 2000. Characterization of the *Mycobacterium tuberculosis iniBAC* promoter, a promoter that responds to cell wall biosynthesis inhibition. J Bacteriol **182:**1802-1811.

47. **Colangeli, R., D. Helb, S. Sridharan, J. Sun, M. Varma-Basil, M. H. Hazbon, R. Harbacheuski, N. J. Megjugorac, W. R. Jacobs, Jr., A. Holzenburg, J. C. Sacchettini, and D. Alland.** 2005. The *Mycobacterium tuberculosis iniA* gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. Mol Microbiol **55:**1829-1840.

48. **Chopra, T., and R. S. Gokhale.** 2009. Polyketide versatility in the biosynthesis of complex mycobacterial cell wall lipids. Methods Enzymol **459:**259-294.

49. **Rocha, E. P., J. Fralick, G. Vediyappan, A. Danchin, and V. Norris.** 2003. A strand-specific model for chromosome segregation in bacteria. Mol Microbiol **49:**895-903.

50. **Konstantinidis, K. T., and J. M. Tiedje.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A **102:**2567-2572.

563 # Tables

564

565 **Table 1: Comparison of key genomic features between the three fully assembled genomes of**

566 **the *M. marinum/M. ulcerans* complex**

Feature	*M. ulcerans* Liflandii	*M. ulcerans* Agy99	*M. marinum* M
Chromosome size (bp)	6,208,955	5,631,606	6,636,827
pMUM plasmid	190,588	174,155	Not present
G+C content (%)	65.62	65.47	65.73
Protein CDS	4994	4160	5424
Unique CDS	268	30	395
Pseudogenes	436	771	65
IS*2404* copies	239	213	Not present
IS*2606* copies	4	91	Not present

Figure Legends

Fig. 1. Genome map of *M. ulcerans* Liflandii. DNA sequence mapping of the three mycobacterial genomes using *M. ulcerans* Liflandii as a reference displayed in Circos. The tracks from inside to outside represent GC skew, GC content, *M. marinum* M, *M. ulcerans* Agy99, ancestral copies of IS*2404*, all *M. ulcerans* Liflandii IS*2404* copies, pseudogenes present in both mycolactone-producing strains but absent in *M. marinum* M, all *M. ulcerans* Liflandii pseudogenes, reverse strand CDS, forward strand CDS, assembled *M. ulcerans* Liflandii genome, major regions of difference MULRD3 and MULRD7. Mapping to pMUM002 is also presented with sequences from *M. marinum* M excluded from the plot. Sequences for *M. ulcerans* Agy99 and *M. marinum* M were obtained from Stinear *et al.,* 2007 and Stinear *et al.,* 2008 (11, 12).

Fig. 2. Comparison of *M. ulcerans* Liflandii. A) Alignment of assembled *M. ulcerans* Liflandii chromosome against an NheI optical map. **B)** ACT Comparison of *M. ulcerans* Agy99 vs *M. ulcerans* Liflandii vs *M. marinum* M vs *M. ulcerans* Agy99. Regions in red indicate identical sequence and orientation and blue indicates identical sequence in the reverse orientation. Indicated are the percentage nucleotide identities among core orthologs. **C)** Venn Diagram highlighting the number of orthologs between *M. marinum* M, *M. ulcerans* Agy99 and *M. ulcerans* Liflandii.

Fig. 3. Functional group distribution. Percentages proportions of **A)** total coding sequences and **B)** pseudogenes by functional group for the sequenced *M. ulcerans/M. marinum* isolates. CDS were classified into one of: intermediary metabolism and respiration (IM & R), cell wall and cell processes (CW & CP), conserved hypotheticals (CH), lipid metabolism (LM), unknown, insertion sequence and phages (IS & P), regulatory proteins (RP), PE/PPE proteins, information pathways (IP) and virulence, detoxification and adaptation (VD & A).

Fig. 4. Reductive evolution of secondary metabolite gene clusters in *M. ulcerans* Liflandii. A) Numbers of pseudogene clusters and intact gene clusters in four mycobacterial genomes. Note here that *M. marinum* and *M. tuberculosis* may also contain some pseudogene clusters that may have escaped detection due to the lack of a reference strain in which these gene clusters are still intact. **B)** Phylogram with maximum parsimony-inferred evolutionary events ('+' for gene cluster gain, '-' for gene cluster loss and 'P' for pseudogene cluster) in the four mycobacteria, assuming that the 10 gene clusters shared by all four mycobacteria represented the ancestral gene cluster repertoire. Probabilities of the three event types were regarded as equal. The numbers behind the species names represent the number of intact and pseudogene clusters in their genomes. Note that reality does not necessarily adhere to maximum parsimony: for example, deletions having occurred in the line towards *M. ulcerans* may in fact have been preceded by pseudogenes forming before the divergence with *M. ulcerans* Liflandii. **C)** Venn diagrams showing which intact and pseudogene clusters are shared between the four mycobacterial genomes.

Fig. 5. Inactivation of secondary metabolite gene clusters. The circular chromosomes are displayed in a linearized way, with the origin of replication on the left. Depicted is the predominance of pseudogenes on the leading strand of mycobacterial chromosomes near the origin of replication. Coloured rectangles indicate biosynthetic gene clusters. Colours and connecting lines represent homology between gene clusters in the different genomes. A star marks those clusters that have been inactivated through pseudogene formation.