

Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer

Timothy P. Stinear,^{1,2} Torsten Seemann,³ Sacha Pidot,² Wafa Frigui,¹ Gilles Reysset,¹ Thierry Garnier,¹ Guillaume Meurice,⁴ David Simon,⁴ Christiane Bouchier,⁵ Laurence Ma,⁵ Magali Tichit,⁵ Jessica L. Porter,² Janine Ryan,² Paul D.R. Johnson,⁶ John K. Davies,² Grant A. Jenkin,² Pamela L.C. Small,⁷ Louis M. Jones,⁸ Fredj Tekaia,⁹ Françoise Laval,¹⁰ Mamadou Daffé,¹⁰ Julian Parkhill,¹¹ and Stewart T. Cole^{1,12}

¹Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 75725 Paris Cedex 15, France; ²Department of Microbiology, Monash University, Clayton 3800, Australia; ³Victorian Bioinformatics Consortium, Monash University, Clayton 3800, Australia; ⁴Plate-Forme 4—Intégration et Analyse Génomique, Génopole, Institut Pasteur, 75725 Paris Cedex 15, France; ⁵Plate-Forme 1—Génomique, Génopole, Institut Pasteur, 75725 Paris Cedex 15, France; ⁶Department of Infectious Diseases, Austin Hospital, Heidelberg 3084, Australia; ⁷Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996-0845, USA; ⁸Groupe Logiciels et Banques de données, Institut Pasteur, 75725 Paris Cedex 15, France; ⁹Unité de Génétique Moléculaire des Levures, Institut Pasteur, 75725 Paris Cedex 15, France; ¹⁰Department of Molecular Mechanisms of Mycobacterial Infections IPBS-CNRS, 31077 Toulouse Cedex 14, France; ¹¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB 10 1SA, United Kingdom.

Mycobacterium ulcerans is found in aquatic ecosystems and causes Buruli ulcer in humans, a neglected but devastating necrotic disease of subcutaneous tissue that is rampant throughout West and Central Africa. Here, we report the complete 5.8-Mb genome sequence of *M. ulcerans* and show that it comprises two circular replicons, a chromosome of 5632 kb and a virulence plasmid of 174 kb. The plasmid is required for production of the polyketide toxin mycolactone, which provokes necrosis. Comparisons with the recently completed 6.6-Mb genome of *Mycobacterium marinum* revealed >98% nucleotide sequence identity and genome-wide synteny. However, as well as the plasmid, *M. ulcerans* has accumulated 213 copies of the insertion sequence IS2404, 91 copies of IS2606, 771 pseudogenes, two bacteriophages, and multiple DNA deletions and rearrangements. These data indicate that *M. ulcerans* has recently evolved via lateral gene transfer and reductive evolution from the generalist, more rapid-growing environmental species *M. marinum* to become a niche-adapted specialist. Predictions based on genome inspection for the production of modified mycobacterial virulence factors, such as the highly abundant phthiodiolone lipids, were confirmed by structural analyses. Similarly, 11 protein-coding sequences identified as *M. ulcerans*-specific by comparative genomics were verified as such by PCR screening a diverse collection of 33 strains of *M. ulcerans* and *M. marinum*. This work offers significant insight into the biology and evolution of mycobacterial pathogens and is an important component of international efforts to counter Buruli ulcer.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to Genbank under accession number CP000325.]

In 1935, two rural medical practitioners reported a series of unusual indolent skin ulcers from patients within a remote farming community in the Bairnsdale district of Southeastern Australia (Alsop 1972). Thirteen years later, a team of Australian researchers discovered the etiologic agent of "Bairnsdale ulcer," a hitherto unknown *Mycobacterium* that they named *Mycobacterium ulcerans* (MacCallum et al. 1948). During the 1960s, many cases were reported from the Buruli County in Uganda; hence, the disease became more generally known as Buruli ulcer. The disease occurs in other parts of the world, but impoverished rural communities of West and Central Africa are worst affected. Since 1989, the

disease burden has steadily increased, and the prevalence of Buruli ulcer now exceeds that of leprosy and, in some instances, tuberculosis (Johnson et al. 2005).

The epidemiology of Buruli ulcer is poorly understood, and outbreaks are sporadic and unpredictable; however, proximity to stagnant or slow-flowing watercourses is a recognized risk factor. *M. ulcerans* is associated with aquatic vegetation such as algae (Marsollier et al. 2004b); snails and other organisms that feed on algae may serve as passive hosts (Marsollier et al. 2004a). One report has shown that *M. ulcerans* can multiply in the salivary glands of carnivorous water bugs, such as *Naucoris cimicoides* (Marsollier et al. 2002). It is conceivable that humans become infected through contact with contaminated *Naucoridae* or bites, as this route of infection has been demonstrated in a murine model of the disease (Marsollier et al. 2002).

The extensive subcutaneous necrosis that results in the typi-

¹²Corresponding author.

E-mail stcole@pasteur.fr; fax +33-1-4061-3583.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5942807>.

cal pathology of Buruli ulcer stems from the multiple actions of a lipophilic macrolide toxin named mycolactone. Recently, a virulence plasmid, pMUM001, was discovered in *M. ulcerans* (Stinear et al. 2004, 2005b), and this encodes three giant polyketide synthases (PKS) required for mycolactone synthesis. Injection of purified mycolactone replicates the human disease in rodent and cellular models. Mycolactone is cytotoxic, has anti-phagocytic activity, induces apoptosis of antigen-presenting cells, and inhibits the pro-inflammatory cytokine response (Coutanceau et al. 2005). As a consequence, from a relatively early stage in disease *M. ulcerans* is found primarily in an extracellular location in mammalian tissues, in contrast to other pathogenic mycobacteria (Coutanceau et al. 2005). Intriguingly, mycolactone is required for lysis of the plasmatocytes that transport the bacterium away from the coelomic cavity in *N. cimicoides* (Marsollier et al. 2005), suggesting that mycolactone may have an important role in the adaptation of *M. ulcerans* to an arthropod niche.

Phylogenetic relationships inferred from limited multi-locus sequence comparisons suggest that *M. ulcerans* has recently diverged from a *Mycobacterium marinum* progenitor (Stinear et al. 2000). *M. marinum* does not contain pMUM001 and makes no mycolactone, but causes a tuberculoid-like disease in fish and frogs and, occasionally, a limited cutaneous infection in humans characterized by intracellular multiplication and a granulomatous host response. Unlike *M. ulcerans*, *M. marinum* replicates in 4 not 50 h, produces light-inducible carotenoid pigments, and can utilize glucose, acetate, succinate, and pyruvate as sole carbon sources.

Currently, there is no vaccine to prevent Buruli ulcer and, owing to the painless nature of the lesions, patients often refer to clinicians so late that surgical excision followed by skin grafts is the only effective recourse. Treatment with rifampin and streptomycin shows some efficacy but has not yet been validated by a controlled clinical trial (Etuafu et al. 2005). A recent resolution adopted by the World Health Assembly called for intensified research to develop tools to diagnose, treat, and prevent Buruli ulcer (World Health Organization, <http://www.who.int/buruli/en/>). The genome sequence of *M. ulcerans* described herein provides an important resource for addressing these aims.

Results

General features

The genome comprises 5,805,761 bp and is composed of two circular replicons, a 5,631,606-bp chromosome (Fig. 1) and a 174,155-bp plasmid, pMUM001. The plasmid, which has an average G+C content of 62.5%, contains 81 protein-coding sequences (CDS) and encodes the polyketide synthases and polyketide-modifying enzymes that are required to produce mycolactone (Stinear et al. 2004, 2005a,b).

The chromosome contains 4160 CDS and 771 pseudogenes, and functional information could be attributed to nearly 70% of these. There are 45 genes encoding tRNA and a single rRNA operon located 1.2 Mb from *oriC*. The chromosome harbors two prophage, phiMU01 and phiMU02, and 302 insertion sequence elements (ISE), including 209 complete or partial copies of IS2404 and 83 copies of IS2606 (Table 1; Fig. 1; Stinear et al. 1999). Cumulative GC skew analysis demonstrated a bias in G+C content between the leading and lagging strands, indicating a probable origin of replication around *dnaA*; however, deviations

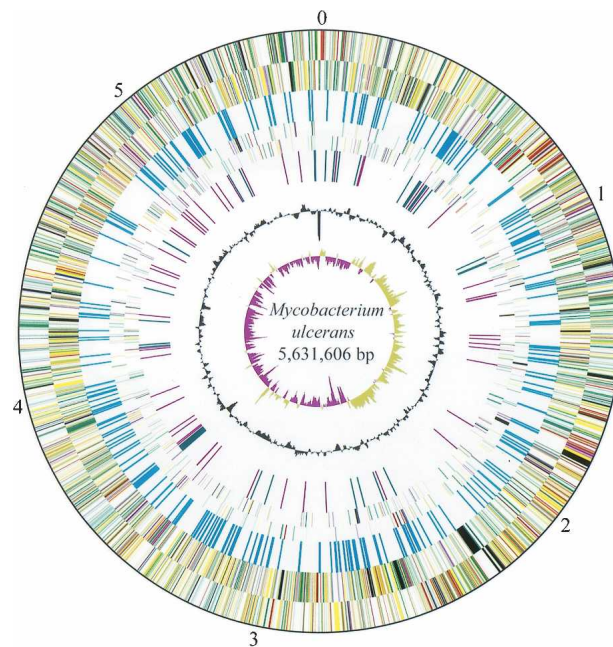


Figure 1. Circular representation of the *Mycobacterium ulcerans* chromosome. The scale is shown in megabases in the outer black circle. Moving inward, the next two circles show forward and reverse strand CDS, respectively, with colors representing the functional classification (red, replication; light blue, regulation; light green, hypothetical protein; dark green, cell wall and cell processes; orange, conserved hypothetical protein; cyan, IS elements; yellow, intermediate metabolism; gray, lipid metabolism; purple, PE/PPE). The location of each copy of IS2404 and IS2606 is then shown (cyan). The following two circles show forward and reverse strand pseudogenes (colors represent the functional classification), followed by the G+C content and finally the GC skew $(G - C)/(G + C)$ using a 20-kb window.

in this bias were also evident (Fig. 1), suggesting the recent occurrence of several large chromosomal rearrangements.

Genome comparisons reveal the origin of *M. ulcerans*

To gain insight into these rearrangements, the recently completed 6,636,827-bp genome sequence of *M. marinum*, with its 5426 CDS, was compared with that of *M. ulcerans* (Fig. 2). This revealed the existence of many deletions, accounting for 1064 kb, and many DNA rearrangements in *M. ulcerans*; some of these are asymmetrical around the origin and terminus of replication (Figs. 1, 2). Symmetrical chromosome rearrangements are commonly seen in bacterial comparisons (Eisen et al. 2000); asymmetric rearrangements that switch genes from the leading to the lagging strands are uncommon, but have been reported in bacteria with large numbers of ISE. Close inspection reveals that in *M. ulcerans* at least 12 of the rearrangements shown in Figure 2 have resulted in transfer to the lagging strand of 273 CDS with leading strand orthologs in *M. marinum* and *M. tuberculosis*, and some of these CDS are predicted to have key roles in cellular metabolism. For example, a 71-kb fragment spanning the region 5,032,986 to 5,104,623 in *M. ulcerans* has been relocated from a region around 945,816 to 1,012,490 in *M. marinum* (Fig. 2). CDS now on the lagging strand in this 71-kb region include (1) *murB* (MUL_4552), encoding a reductase required for peptidoglycan biosynthesis; (2) *icl* (MUL_4536), encoding isocitrate lyase, a key enzyme of the glyoxylate shunt; (3) *proC* (MUL_4570), encoding a carboxylate reductase that is an essential gene in *M. tuberculosis*

Table 1. Summary of DNA sequences present in *M. ulcerans* Agy99 and absent from *M. marinum* M

DNA present in <i>M. ulcerans</i> strain Agy99 but absent from <i>M. marinum</i> strain M	Total length (kb)	Percent of total genome (5806 kb)	No. of copies/fragments	Comments/functional description
pMUM001	174	3.0	1–2	Virulence plasmid, harbors PKS genes for mycolactone production.
IS2404	335	5.8	213	209 chromosomal copies, including 19 pseudogenes. Four copies on pMUM001.
IS2606	129	2.2	91	83 chromosomal copies, including 10 pseudogenes. Eight copies on pMUM001.
PhiMU01 and PhiMU02	29	0.5	2	Prophage, containing three copies of IS2404 and six copies of IS2606 (12.7 kb). PhiMU01 harbors a putative metalloprotease (MUL_3218).
PE/PPE genes	15	0.3	97	Uniqueness arises from extensive C-terminal sequence variation.
Other	22	0.4	32	Includes MUL_2832, hypothetical lipase and MUL_0999, membrane protein.

required for proline biosynthesis; and (4) the *hemACDB* operon (MUL_4598 to MUL_4561), a group of essential genes in *M. tuberculosis* required for porphyrin biosynthesis (Sasseti et al. 2001). Such exchanges of CDS involved in cell wall biosynthesis and carbon, amino acid, and cofactor metabolism to the lagging strand might explain in part the slow growth of *M. ulcerans*, as transcription of these (and other) potentially essential genes could be slowed or disrupted by competition between DNA and RNA polymerases on the lagging strand at the replication forks (Rocha and Danchin 2003). However, it is important to note that chromosome remodeling in *M. ulcerans* has not resulted in a significant net change in strand coding bias as 239 CDS have moved from lagging to leading strand.

Further comparisons between the genomes of *M. ulcerans* and *M. marinum* confirmed a very close relationship between these two species and show that *M. ulcerans* has recently evolved from *M. marinum*. Approximately 90% of the *M. ulcerans* CDS have syntenic orthologs in *M. marinum* with an average DNA

identity level of 98.3% (excluding ISE and PE/PPE genes). This compares with only 60% of *M. ulcerans* CDS that have an *M. tuberculosis* ortholog and an average DNA identity of 78.5%.

Identification of *M. ulcerans*-specific DNA

Many regions of difference are due to insertions and deletions in both species (Supplemental Table S1). In addition to pMUM001, 475 kb of DNA was identified in *M. ulcerans* that is absent from *M. marinum*, and this is distributed throughout the chromosome (Fig. 1; Table 1). Most of this is accounted for by IS2404 and IS2606. The remaining 79 kb includes the prophages and extensive DNA sequence variation among the PE/PPE genes (Table 1). There are 11 predicted CDS that present a diagnostic interest (Table 1). PCR screening for these 11 genes among a diverse panel of *M. ulcerans* and *M. marinum* strains confirmed that they are all *M. ulcerans*-specific and fully conserved (Table 2).

Insertion sequences and phages

Sequencing revealed many more copies of IS2404 and IS2606 (Table 1) than were first estimated (Stinear et al. 1999). These elements have profoundly affected genome plasticity, flanking many of the chromosome inversions, marking the site of DNA deletions, and causing 111 of the 743 predicted pseudogenes by inserting within genes. The *M. marinum* M strain has only 64 pseudogenes and possesses neither IS2404 nor IS2606. There are 10 other putative ISE belonging to various families, including one single-copy element (MUL_2378) belonging to the IS3 family that is present in nine copies in *M. marinum*. The two prophages named phiMU01 (18 kb, 18 CDS) and phiMU02 (24 kb, 17 CDS) resemble other mycobacteriophages in overall structure, integrating near tRNA genes and containing CDS associated with replication functions. However, phiMU02 may be non-functional, as several of its genes have been inactivated by multiple IS2606 insertions.

Pseudogene accumulation

Foremost among pseudogenes in *M. ulcerans* are members of the PE and PPE multigene families encoding the Gly-Ala-rich cell envelope proteins confined to mycobacteria. *M. marinum* has 170 PE and 105 PPE, respectively, compared with *M. ulcerans* that has retained only 70 intact PE and 46 PPE, with the remaining or-

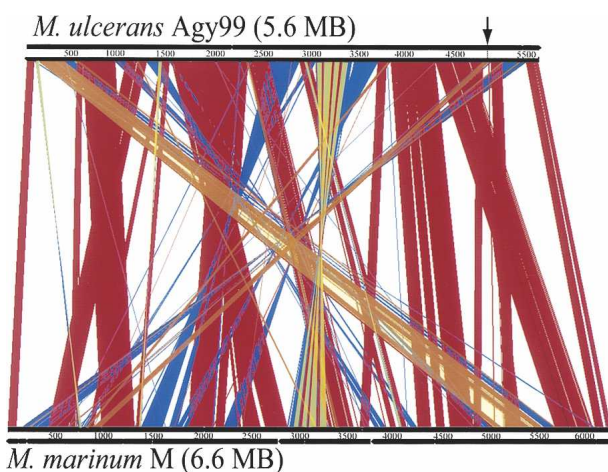


Figure 2. Diagram depicting the rearrangements in, and deletions from, the genome of *M. ulcerans* compared with *M. marinum*. The figure shows a linear genomic comparison generated with ACT (Carver et al. 2005). (Red lines) Regions of DNA:DNA identity, (blue lines) inverted regions, (orange) strand-displacing rearrangements, (arrow) position of the displaced 71-kb fragment.

Table 2. Presence of 11 *M. ulcerans*-specific CDS among different strains of *M. ulcerans* and *M. marinum* as determined by PCR

Locus tag	CDS description	Strain details												
		Ghana Agy99	Ivory Coast Kob	SE Australia Chant	North Australia 13822	Papua New Guinea 941331	Malaysia 1615	Malaysia 941328	People's Republic of China 980912	Japan 753	Mexico 5114	Surinam 842	<i>M. marinum</i> ^a	
MUL_0076	HP ^b	+	+	+	+	+	+	-	+	+	+	+	-	-
MUL_0512	HP phiMU01	+	+	-	+	+	+	+	+	+	+	+	-	-
MUL_0513	HP phiMU01	+	+	-	+	+	+	+	+	+	+	+	-	-
MUL_0515	HP phiMU01	+	+	-	+	+	+	+	+	+	+	+	-	-
MUL_0526	HP phiMU01	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_0527	HP phiMU01	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_0998	HP	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_0999	Membrane protein	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_1001	HP	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_2832	Lipase	+	+	+	+	+	+	+	+	+	+	+	-	-
MUL_3218	CHP ^c phiMU02	+	+	+	+	+	+	+	+	+	+	+	-	-

^aIncludes *M. marinum* M and 22 previously described *M. marinum* strains (Stinear et al. 2005a).^bHypothetical protein.^cConserved hypothetical protein.

(CDS) Protein-coding sequences.

thologs showing widespread sequence variation (Fig. 1; Table 1). These changes suggest a decreased requirement for PE/PPE proteins. The degeneration appears symptomatic of an intermediate stage of reductive evolution with *M. ulcerans* heading toward the highly contracted genome state of *M. leprae*, which is nearly devoid of these proteins (Cole et al. 2001).

The loss of paralogous genes and subsequent decrease in genetic redundancy may also contribute to slow growth via a diminished gene dose effect. For example, the disaccharide trehalose is essential for mycobacterial cell wall biogenesis and can be synthesized by three potential pathways (OtsAB, TreYZ, and TreS). As in *M. tuberculosis*, *M. ulcerans* has two *otsB* genes, *otsB1* and *otsB2*; however, *otsB1* has been disrupted by IS2404. Similarly, one of the three *gltA* genes coding for glutamine synthetase, a central enzyme in the nitrogen regulon, is a pseudogene in *M. ulcerans*, as are three of four phospholipase C paralogs required for phospholipolysis.

Other gene inactivations suggest divergence of the ecological niches of *M. ulcerans* and *M. marinum*. The *crtB* locus in *M. marinum* is responsible for the production of light-inducible carotenoids that protect the bacterium from incident sunlight. *M. ulcerans* does not produce these pigments, suggesting it is not exposed to direct sunlight in its habitat, yet it contains an identical *crtB* region. One explanation for the lack of photochromogenicity may be the introduction of a premature stop codon in *crtI*, encoding phytoene dehydrogenase, the second committed step in pigment synthesis.

Protein families

To better understand the biology of *M. ulcerans*, the predicted proteins were classified into families and compared with those found in other mycobacteria. Only one protein family ($N > 5$), comprising eight metal-dependent hydrolases, was found that had no counterpart in tubercle and leprosy bacilli. Otherwise, the same themes were observed (Tekaija et al. 1999), namely an abundance of proteins involved in lipid metabolism and prominent PE and PPE families. However, the number of potential transport proteins and enzyme systems predicted to be involved in carbon and energy metabolism was higher. *M. marinum* has 192 CDS associated with substrate transport while *M. ulcerans* has only 128.

If *M. ulcerans* occupies several distinct niches, including plant biofilms, insects, and humans, then its metabolism may be versatile. In an aquatic environment, *M. ulcerans* might obtain its energy and carbon from degradation of plant saccharides, as these stimulate growth in vitro (Marsollier et al. 2004b). Little is known, however, about its metabolism in insect tissues, although it may be significant that both *M. marinum* and *M. ulcerans* encode four potential chitinases/transglycosidases (BuruList/mast/P4.27.html), one of which is attached to a PE domain and thus is expected to be in the cell envelope. These enzymes might mediate attachment to, or degradation of, the *N*-acetyl-D-glucosamine polymers that comprise chitin, a major component of the exoskeletons of insects and crustaceans.

Primary metabolism and respiration

Metabolic pathway reconstruction supported the importance of lipid degradation for central carbon metabolism and uncovered intact glycolysis and pentose phosphate pathways, but no Entner-Doudoroff pathway. As in *M. tuberculosis*, *M. ulcerans* has a bifurcated TCA cycle, as it lacks α -ketoglutarate dehydrogenase

(Fig. 3) and thus probably converts α -ketoglutarate to succinate via α -ketoglutarate decarboxylase (Tian et al. 2005). The TCA cycle incorporates the glyoxylate shunt, with two copies of the gene encoding isocitrate lyase. Thus, like *M. tuberculosis*, *M. ulcerans* may direct the two-carbon degradation products (acetyl-CoA) from β -oxidation of host fatty acids into the TCA cycle operating in biosynthetic mode.

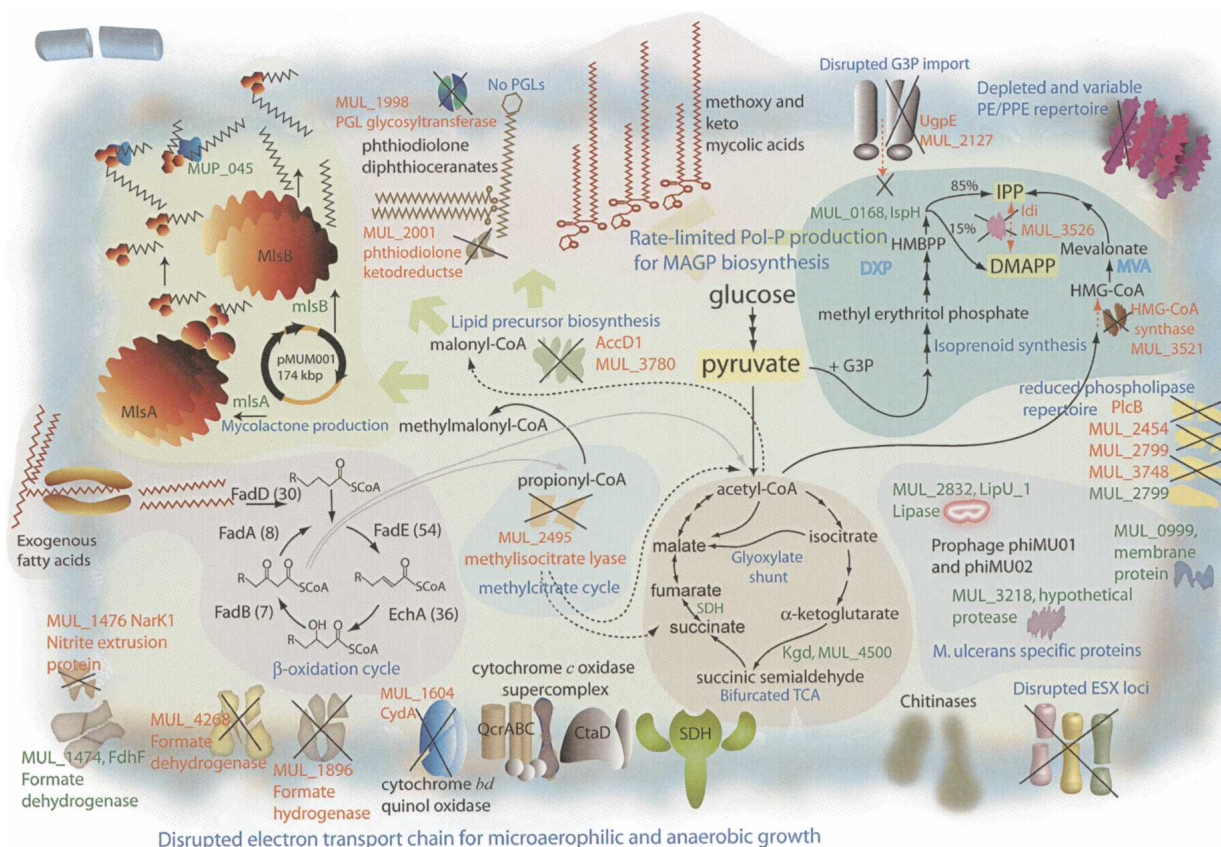
The β -oxidation of fatty acids also generates propionyl-CoA, which can be metabolized in several ways including via the methylcitrate pathway (Fig. 3). *M. ulcerans* appears to have all the methylcitrate enzymes except methylisocitrate lyase, whose putative gene (MUL_2495) has a frameshift mutation. A reduced capacity to metabolize propionic acid might be detrimental for growth; however, in *M. tuberculosis* the combination of the two isocitrate lyases can complement the absence of methylisocitrate lyase (Munoz-Elias and McKinney 2005).

Analysis of potential electron transport chains suggests *M. ulcerans* is capable of growth under aerobic but not anaerobic conditions as it lacks the nitrate and fumarate reductase systems. Based on gene orthology with *M. tuberculosis*, complete NADH oxidase, ubiquinol cytochrome *c* oxidase, and ATP synthase complexes indicate ATP generation by oxidative phosphorylation during aerobic growth. *M. marinum* contains a potential alternative anaerobic respiratory pathway involving nitrite reductase, encoded by *nrfD* and formate dehydrogenase, where formate oxidation is coupled to nitrite reduction. *M. ulcerans* contains an intact nitrite reductase (*nrfD*), but the linked formate dehydrogenase system has been inactivated by multiple mutations (Fig. 3). Loss of this enzyme, a known selenoprotein, is consistent with deletion of the selenocysteine locus from *M. ulcerans*. *M. marinum* may be capable of nitrite reduction via the *nirBD* nitrite reductase and NarK transporter, but while *M. ulcerans* also contains these loci, *nirB* and *narK* are pseudogenes.

Under microaerophilic conditions, mycobacteria up-regulate expression of a high oxygen affinity cytochrome *bd* oxidase (*cydABCD*). This locus is present in both *M. ulcerans* and *M. marinum*, but in *M. ulcerans* the *cydA* ortholog is a pseudogene. A knockout mutation of *cydA* in *Mycobacterium smegmatis* leads to a competitive growth disadvantage under microaerophilic conditions (Kana et al. 2001). The above observations suggest that *M. ulcerans* has depleted respiratory versatility. Nevertheless, it has maintained >400 putative oxidoreductases, dehydrogenases, and mono- and dioxygenases (*M. marinum* has >590), suggesting that a robust and complex respiratory potential remains under aerobic conditions. As reported in other bacteria undergoing reductive evolution, the cytochrome P450 repertoire was particularly prone to pseudogene formation (Cole et al. 2001).

Cell wall lipids and secondary metabolism

The cell envelope of *M. ulcerans* and the secondary metabolites therein have been less well studied than those of other mycobacteria. From the genome sequence several predictions can be made, as reduction of secondary metabolite loci is widespread, notably through depletion of PKS genes. *M. marinum* has 27 PKS genes, but *M. ulcerans* has reduced its complement to 12. Ten of these PKS produce important cell-wall-associated lipids including mycolic acids (*pkS13*), phthiodiolones (*ppsA-E*), phenol phthiodiolones (*pkS15/1*), mannosyl- β 1-phospholipids (*pkS12*), and mycobactins (*mbtC,D*) (Matsunaga et al. 2004; Portevin et al. 2004).



Disrupted electron transport chain for microaerophilic and anaerobic growth

Figure 3. Some important biochemical pathways. The main biochemical activities of *M. ulcerans* discussed in the text are depicted; these are all present in *M. marinum* except for mycolactone biosynthesis encoded by pMUM001. Some key genes and pathways that have been inactivated in *M. ulcerans* are indicated.

Neither *M. ulcerans* nor *M. marinum* contains the *pks2* locus, required for the production of the sulfolipids. A larger PKS operon of unknown function harbors *pks9* and *pks11*, but this operon may be inactive in *M. ulcerans*, as its other PKS genes (*pks7*, *pks8*, *pks10*) have become pseudogenes. One predicted consequence for *M. ulcerans* of downsizing its PKS complement would be liberation of energy (NADP) and substrate (malonyl-CoA and methylmalonyl-CoA) for the production of mycolactone. Contraction of secondary metabolism is also reflected in reduction of the MmpL family of lipid and polyketide transporters from 25 in *M. marinum* to only six in *M. ulcerans*.

Despite the loss of many *pks*, *M. ulcerans* has retained a significant anabolic lipid potential highlighting the central importance of particular lipids to the preservation of the mycobacterial cell wall. Structural analyses of some of the signature lipid species were undertaken and then correlated with the genome findings. *M. ulcerans* and *M. marinum* have been previously shown to produce diunsaturated, methoxy, and ketomycolic acids (Daffe et al. 1991). Central to synthesis of these mycolates and other lipids are the fatty acid synthases I and II. All the principal components of these enzymes are present in *M. ulcerans*, and the key proteins share >90% identity with those of *M. marinum* and *M. tuberculosis*.

M. ulcerans also produces a highly abundant and apolar cell wall lipid consisting of diesters of phthiodiolone (Daffe et al. 1984; Onwueme et al. 2005). The common β -diol backbone of phthiodiolones in *M. ulcerans* is produced from C₁₆–C₁₈ fatty

acids by five type I PKS encoded by the genes *ppsA*–*E*. Processing of *p*-hydroxyphenylalkanoate rather than fatty acids by these PKS results in the production of phenolphthiodiolones, which can be further modified by glycosylation to form the virulence factors and immune modulators, the phenolic glycolipids (PGLs) (Daffe et al. 1992; Reed et al. 2004). In *M. tuberculosis*, Rv2962 encodes the glycosyl transferase that adds the first sugar residue to the highly related lipid, phenolphthiocerol (Perez et al. 2004). The absence of PGLs in *M. ulcerans* Agy99 is readily explained, as *MUL_1998*, the Rv2962 ortholog, is a pseudogene. This observation was supported by MALDI-TOF MS detection of the substrate for the *MUL_1998* gene product, phenolphthiodiolone diphthioceranoate, in *M. ulcerans* but not in its glycosylated form, which was detected in *M. marinum* (Figs. 3, 4).

M. ulcerans and *M. marinum* contain both a mevalonate (MVA) and non-mevalonate (DXP) pathway for polyprenoid synthesis, the first example of such in the mycobacteria and a rare occurrence among bacteria (Rohdich et al. 2004). Polyprenyl phosphate (Pol-P) forms lipid-linked sugar intermediates that are required by bacteria during cell wall biosynthesis. They are formed by sequential condensation of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In other mycobacteria these molecules are synthesized via the DXP pathway; the same appears true for *M. ulcerans*, as the MVA pathway is probably inactive because one of the genes in the MVA operon, *MUL_3521*, encoding a putative hydroxymethylglutaryl-coenzyme A synthase, has been disrupted (Fig. 3).

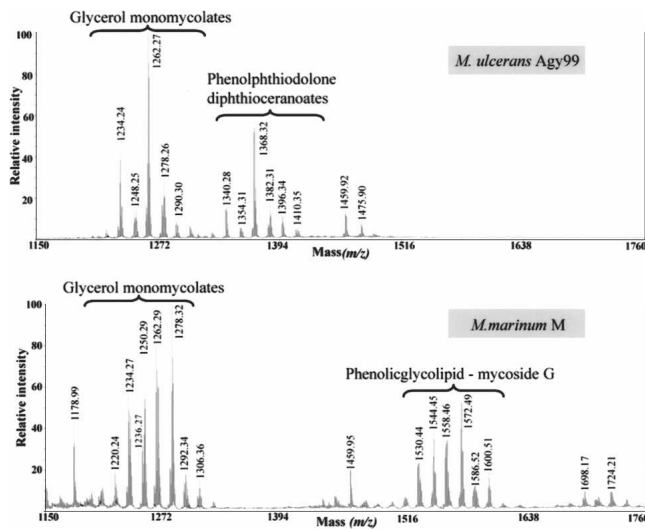


Figure 4. Partial matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectra of lipids from *M. ulcerans* and *M. marinum*. These spectra show the presence of diesters of phthiocerol and phthiodiolone, and of phenolic glycolipids in *M. marinum*, whereas only diesters of phthiodiolone and phenolphthiodiolone occur in *M. ulcerans*.

Implications for pathogenesis from genome reduction

Comparative genomics also shows that the *M. ulcerans* genome is contracting and that extensive pseudogene formation has occurred. There are 157 regions, accounting for 1232 kb, present in *M. marinum* but absent from *M. ulcerans*. These were assigned a *M. ulcerans* Region of Difference (MURD) number (Supplemental Table S1). While some of these represent insertions in *M. marinum*, including seven prophage and 13 ISE, totaling 168 kb, the majority result from the accumulated deletion of 1064 kb of DNA from *M. ulcerans*. In this way many paralogous gene families have been lost in a situation reminiscent of *M. leprae* where considerable genetic downsizing has occurred (Cole et al. 2001). Most notable among these are several secondary metabolism loci and the PE/PPE genes, the latter accounting for 45% of the MURDs. Other deletions affect genes involved in intermediary and energy metabolism such as potassium transport (MURD9), hydrogenase (MURD17), and production of selenocysteine (MURD142), a rare amino acid found in the active sites of formate dehydrogenase and many bacterial antioxidant proteins.

Noteworthy deletions in *M. ulcerans* include the ESX loci. While there are five complete ESX systems in *M. marinum*, only three remain in *M. ulcerans* as the other two have incurred deletions. In *M. tuberculosis*, the *esx1* locus encodes a novel protein secretion apparatus, which contributes to its virulence, intercellular spread, and immunogenicity (Brodin et al. 2004). In *M. ulcerans*, *esx1* has been disrupted by two deletions (MURD151 and MURD152). Effector proteins secreted by ESX systems belong to the ESAT-6 or the EspA families (Fortune et al. 2005), and 13 ESAT-6 proteins exist in *M. ulcerans*. Interestingly, the *espA* gene, encoding a putative substrate of ESX-1, has been deleted (MURD111). Loss of these systems, which trigger granuloma formation by *M. marinum* and *M. tuberculosis*, may contribute to the predominantly extracellular location of *M. ulcerans* in infected tissue. Supplemental Table S1 also lists 102 orthologous *M. tuberculosis* genes that *M. ulcerans* has lost by deletion.

Discussion

Bacteria that have recently passed an evolutionary bottleneck and are now adapting to a new, more stable environment, such as *Mycobacterium leprae*, *Yersinia pestis*, and *Bordetella pertussis* (Parkhill et al. 2003), have some of the following genomic signatures: proliferation of ISE, accumulation of pseudogenes, chromosomal rearrangements, genome downsizing, and acquisition of foreign genes, often via plasmids or bacteriophage, that confer a fitness advantage in the new environment. They also show a high degree of genetic relatedness or clonality. All six genomic signatures are shared by *M. ulcerans*, which appears to be undergoing reductive evolution and, as such, may represent a good model for studying this phenomenon and the response to changing from an environmental to host-adapted, possibly arthropod niche.

Our study provides compelling evidence for *M. ulcerans* being a descendant of an *M. marinum* progenitor strain that acquired the virulence plasmid, pMUM001, from another actinobacterium. This event then led to divergence of the ecological niches of *M. ulcerans* and *M. marinum*, although other factors may have contributed. For instance, the *crtB* locus is responsible for the production of light-inducible carotenoids that protect *M. marinum* from incident sunlight (Ramakrishnan et al. 1997). *M. ulcerans* does not produce these pigments, due to a lesion in *crtI*, so its impaired ability to withstand exposure to direct sunlight would lead to selection for progeny capable of adapting to, and settling in, a more protected niche. The strong selective pressure exerted by mycolactone could have enabled *M. ulcerans* to colonize aquatic insects or other animals sharing its ecosystem.

Once within a more stable and protected habitat, loss of gene function and slower growth would be readily tolerated. Genomic analysis suggests possible mechanisms that have impacted on growth rate including the loss, or reduction in number, of genes for transport of potassium and other solutes; enzymes such as hydrogenase, glutamine synthase, and phospholipase C; and for trehalose and isoprenoid production, among others; in turn, the capacity for free-living would diminish.

Unlike other pathogenic mycobacteria, *M. ulcerans* is found predominantly in extracellular form in infected human tissue, although it passes via a transient intracellular stage (Coutanceau et al. 2005). In part, this is due to the antiphagocytic properties of mycolactone, but loss of the protein secretion system, ESX-1, that mediates the export of the proteins ESAT-6 and EspA may also contribute. Inactivation of ESX-1 in *M. marinum* or *M. tuberculosis* results in reduced uptake by phagocytes and greatly diminished intercellular spread (Brodin et al. 2004). It is also conceivable that some of the functions encoded by the *M. ulcerans*-restricted CDS could contribute to the pathology associated with Buruli ulcer. These genes or their products may also find application in the development of new diagnostic tests, which are required to help control the spread of Buruli ulcer, and their conservation in all isolates of *M. ulcerans* examined is encouraging in this respect.

Methods

M. ulcerans Agy99

M. ulcerans Agy99 was isolated from an ulcerative lesion on the right elbow of a female patient from the Ga district of Ghana in 1999. Multi-locus sequence typing and inter-IS PCR analysis con-

firmed that this strain belonged to the African epidemic genotype. DNA was prepared for sequencing four passages after primary isolation.

Whole-genome sequencing

A whole-genome shotgun library was prepared in the vector pCDNA2.1 using *Escherichia coli* strain XL2-blue (Invitrogen). Genomic DNA was sheared by sonication and size fractionated to produce plasmids with inserts in the size ranges of 2–3 kb, 3–5 kb, and 5–10 kb. These were used as template in DNA cycle sequencing reactions to obtain 75,415 end-sequence reads using an ABI 3700 DNA sequencer (Applied Biosystems), providing theoretical 8× coverage. Sequence assembly was performed by PHRAP and GAP4 as described (Cole et al. 1998). Gap closure was facilitated by scaffolding using a large insert BAC library (Stinear et al. 2005a) augmented with primer walking and PCR.

Genome annotation and comparative genomics

Annotation and database construction was performed with Wasabi, an in-house, Web-interfaced MySQL database for management of genome annotation, Artemis, and GenoList. A complete database is available at <http://genolist.pasteur.fr/BuruList/>. CDS were predicted using GenemarkS and GenoStar, and compared with public sequence databases using the BLAST suite of algorithms. Pseudogenes had one or more mutations that would prevent correct translation. Mutations were checked against original sequence data to exclude sequencing errors. The Artemis Comparison Tool (Carver et al. 2005) was used for comparative genome analysis.

Lipid analysis

Lipids were extracted with CHCl₃/CH₃OH (1:1 v/v) and subjected to MALDI-TOF mass spectrometry using a Voyager DE-STR MALDI-TOF instrument (PerSeptive Biosystems) equipped with a pulse nitrogen laser emitting at 337 nm and analyzed in the reflector mode as previously described (Perez et al. 2004).

Acknowledgments

We thank Paul Harrison, Ivan Moszer, and Roland Brosch for sharing expertise and helpful discussions. We gratefully acknowledge the financial support of the Génopole program, the Association Française Raoul Follereau, the National Health and Medical Research Council of Australia, and the World Health Organization through grants provided by the Nippon Foundation, Tokyo, Japan.

References

Alsop, D. 1972. The Bairnsdale ulcer. *Aust. N. Z. J. Surg.* **41**: 317–319.

Brodin, P., Rosenkrands, I., Andersen, P., Cole, S.T., and Brosch, R. 2004. ESAT-6 proteins: Protective antigens and virulence factors? *Trends Microbiol.* **12**: 500–508.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: The Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., et al. 2001. Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007–1011.

Coutanceau, E., Marsollier, L., Brosch, R., Perret, E., Goossens, P., Tanguy, M., Cole, S.T., Small, P.L., and Demangel, C. 2005. Modulation of the host immune response by a transient intracellular

stage of *Mycobacterium ulcerans*: The contribution of endogenous mycolactone toxin. *Cell. Microbiol.* **7**: 1187–1196.

Daffe, M., Laneelle, M.A., Roussel, J., and Asselineau, C. 1984. Specific lipids from *Mycobacterium ulcerans*. *Ann. Microbiol. (Paris)* **135A**: 191–201.

Daffe, M., Laneelle, M.A., and Lacave, C. 1991. Structure and stereochemistry of mycolic acids of *Mycobacterium marinum* and *Mycobacterium ulcerans*. *Res. Microbiol.* **142**: 397–403.

Daffe, M., Varnerot, A., and Levy-Frebault, V.V. 1992. The phenolic mycoside of *Mycobacterium ulcerans*: Structure and taxonomic implications. *J. Gen. Microbiol.* **138**: 131–137.

Eisen, J.A., Heidelberg, J.F., White, O., and Salzberg, S.L. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1**: research0011.

Etuafu, S., Carbone, B., Grosset, J., Lucas, S., Horsfield, C., Phillips, R., Evans, M., Ofori-Adjei, D., Klustse, E., Owusu-Boateng, J., et al. 2005. Efficacy of the combination rifampin–streptomycin in preventing growth of *Mycobacterium ulcerans* in early lesions of Buruli ulcer in humans. *Antimicrob. Agents Chemother.* **49**: 3182–3186.

Fortune, S.M., Jaeger, A., Sarracino, D.A., Chase, M.R., Sasseti, C.M., Sherman, D.R., Bloom, B.R., and Rubin, E.J. 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc. Natl. Acad. Sci.* **102**: 10676–10681.

Johnson, P.D., Stinear, T., Small, P.L., Pluschke, G., Merritt, R.W., Portaels, F., Huygen, K., Hayman, J.A., and Asiedu, K. 2005. Buruli ulcer (*M. ulcerans* infection): New insights, new hope for disease control. *PLoS Med.* **2**: e108.

Kana, B.D., Weinstein, E.A., Avarbock, D., Dawes, S.S., Rubin, H., and Mizrahi, V. 2001. Characterization of the *cydAB*-encoded cytochrome bd oxidase from *Mycobacterium smegmatis*. *J. Bacteriol.* **183**: 7076–7086.

MacCallum, P., Tolhurst, J., Buckle, G., and Ha, S. 1948. A new mycobacterial infection in man. *J. Pathol. Bacteriol.* **60**: 93–122.

Marsollier, L., Robert, R., Aubry, J., Saint Andre, J.P., Kouakou, H., Legras, P., Manceau, A.L., Mahaza, C., and Carbone, B. 2002. Aquatic insects as a vector for *Mycobacterium ulcerans*. *Appl. Environ. Microbiol.* **68**: 4623–4628.

Marsollier, L., Severin, T., Aubry, J., Merritt, R.W., Saint Andre, J.P., Legras, P., Manceau, A.L., Chauty, A., Carbone, B., and Cole, S.T. 2004a. Aquatic snails, passive hosts of *Mycobacterium ulcerans*. *Appl. Environ. Microbiol.* **70**: 6296–6298.

Marsollier, L., Stinear, T., Aubry, J., Saint Andre, J.P., Robert, R., Legras, P., Manceau, A.L., Audrain, C., Bourdon, S., Kouakou, H., et al. 2004b. 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.

Marsollier, L., Aubry, J., Coutanceau, E., Andre, J.P., Small, P.L., Milon, G., Legras, P., Guadagnini, S., Carbone, B., and Cole, S.T. 2005. Colonization of the salivary glands of *Naucoris cimicoides* by *Mycobacterium ulcerans* requires host plasmacytes and a macrolide toxin, mycolactone. *Cell. Microbiol.* **7**: 935–943.

Matsunaga, I., Bhatt, A., Young, D.C., Cheng, T.Y., Eyles, S.J., Besra, G.S., Briken, V., Porcelli, S.A., Costello, C.E., Jacobs Jr., W.R., et al. 2004. *Mycobacterium tuberculosis* pks12 produces a novel polyketide presented by CD1c to T cells. *J. Exp. Med.* **200**: 1559–1569.

Munoz-Elias, E.J. and McKinney, J.D. 2005. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nat. Med.* **11**: 638–644.

Onwueme, K.C., Vos, C.J., Zurita, J., Soll, C.E., and Quadri, L.E. 2005. Identification of phthiidiolone ketoreductase, an enzyme required for production of mycobacterial diacyl phthiocerol virulence factors. *J. Bacteriol.* **187**: 4760–4766.

Parkhill, J., Sebahia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., et al. 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**: 32–40.

Perez, E., Constant, P., Lemassu, A., Laval, F., Daffe, M., and Guilhot, C. 2004. Characterization of three glycosyltransferases involved in the biosynthesis of the phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. *J. Biol. Chem.* **279**: 42574–42583.

Portevin, D., De Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffe, M., and Guilhot, C. 2004. A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. *Proc. Natl. Acad. Sci.* **101**: 314–319.

Ramakrishnan, L., Tran, H.T., Federspiel, N.A., and Falkow, S. 1997. A *crfB* homolog essential for photochromogenicity in *Mycobacterium marinum*: Isolation, characterization, and gene disruption via

- homologous recombination. *J. Bacteriol.* **179**: 5862–5868.
- Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G., and Barry III, C.E. 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431**: 84–87.
- Rocha, E.P. and Danchin, A. 2003. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat. Genet.* **34**: 377–378.
- Rohdich, F., Bacher, A., and Eisenreich, W. 2004. Perspectives in anti-infective drug design. The late steps in the biosynthesis of the universal terpenoid precursors, isopentenyl diphosphate and dimethylallyl diphosphate. *Bioorg. Chem.* **32**: 292–308.
- Sassetti, C.M., Boyd, D.H., and Rubin, E.J. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proc. Natl. Acad. Sci.* **98**: 12712–12717.
- Stinear, T., Ross, B.C., Davies, J.K., Marino, L., Robins-Browne, R.M., Oppedisano, F., Sievers, A., and Johnson, P.D. 1999. Identification and characterization of IS2404 and IS2606: Two distinct repeated sequences for detection of *Mycobacterium ulcerans* by PCR. *J. Clin. Microbiol.* **37**: 1018–1023.
- Stinear, T.P., Jenkin, G.A., Johnson, P.D.R., and Davies, J.K. 2000. Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *J. Bacteriol.* **182**: 6322–6330.
- Stinear, T.P., Mve-Obiang, A., Small, P.L., Frigui, W., Pryor, M.J., Brosch, R., Jenkin, G.A., Johnson, P.D., Davies, J.K., Lee, R.E., et al. 2004. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc. Natl. Acad. Sci.* **101**: 1345–1349.
- Stinear, T.P., Hong, H., Frigui, W., Pryor, M.J., Brosch, R., Garnier, T., Leadlay, P.F., and Cole, S.T. 2005a. Common evolutionary origin for the unstable virulence plasmid pMUM found in geographically diverse strains of *Mycobacterium ulcerans*. *J. Bacteriol.* **187**: 1668–1676.
- Stinear, T.P., Pryor, M.J., Porter, J.L., and Cole, S.T. 2005b. Functional analysis and annotation of the virulence plasmid pMUM001 from *Mycobacterium ulcerans*. *Microbiol.* **151**: 683–692.
- Tekaia, F., Gordon, S.V., Garnier, T., Brosch, R., Barrell, B.G., and Cole, S.T. 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber. Lung Dis.* **79**: 329–342.
- Tian, J., Bryk, R., Itoh, M., Suematsu, M., and Nathan, C. 2005. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: Identification of α -ketoglutarate decarboxylase. *Proc. Natl. Acad. Sci.* **102**: 10670–10675.

Received September 12, 2006; accepted in revised form November 29, 2006.