2017

Cell wall peptidolipids of *Mycobacterium avium*: from genetic prediction to exact structure of a nonribosomal peptide

John Bannantine
USDA-ARS, john.bannantine@usda.gov

Giles Etienne
Université de Toulouse, CNRS, UPS, Toulouse

Francoise Laval
Université de Toulouse, CNRS, UPS, Toulouse

Judith R. Stabel
ARS-USDA, jstabel@nadc.ars.usda.gov

Anne Lemassu
Université de Toulouse, CNRS, UPS, Toulouse

See next page for additional authors

Molecular Microbiology (2017) 105(4), 525–539

doi:10.1111/mmi.13717
First published online 15 June 2017

Cell wall peptidolipids of *Mycobacterium avium*: from genetic prediction to exact structure of a nonribosomal peptide

John P. Bannantine,[1]* Gilles Etienne,[2,3] Françoise Laval,[2,3] Judith R. Stabel,[1] Anne Lemassu,[2,3] Mamadou Daffé,[2,3] Darrell O. Bayles,[1] Christelle Ganneau,[4,5] Frédéric Bonhomme,[4,5] Maxime Branger,[6] Thierry Cochard,[6] Sylvie Bay[4,5] and Franck Biet[6]**

[1]*National Animal Disease Center, USDA-Agricultural Research Service, Ames, IA, 50010, USA.*
[2]*Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, Toulouse, France.*
[3]*Université de Toulouse, UPS, IPBS, Toulouse, 31000, France.*
[4]*Institut Pasteur, Unité Chimie des Biomolécules, 75724 Paris Cedex 15, France.*
[5]*CNRS UMR 3523, 75724 Paris Cedex 15, France.*
[6]*Infectiologie et Santé Publique, INRA, Université de Tours, UMR1282, Nouzilly, F-37380, France.*

Summary

Mycobacteria have a complex cell wall structure that includes many lipids; however, even within a single subspecies of *Mycobacterium avium* these lipids can differ. Total lipids from an *M. avium* subsp. *paratuberculosis* (*Map*) ovine strain (S-type) contained no identifiable glycopeptidolipids or lipopentapeptide (L5P), yet both lipids are present in other *M. avium* subspecies. We determined the genetic and phenotypic basis for this difference using sequence analysis as well as biochemical and physico-chemical approaches. This strategy showed that a nonribosomal peptide synthase, encoded by *mps1*, contains three amino acid specifying modules in ovine strains, compared to five modules in bovine strains (C-type). Sequence analysis predicted these modules would produce the tripeptide Phe-*N*-Methyl-Val-Ala with a lipid moiety, termed lipotripeptide (L3P). Comprehensive physico-chemical analysis of *Map* S397 extracts confirmed the structural formula of the native L3P as D-Phe-*N*-Methyl-L-Val-L-Ala-OMe attached in *N*-ter to a 20-carbon fatty acid chain. These data demonstrate that S-type strains, which are more adapted in sheep, produce a unique lipid. There is a dose-dependent effect observed for L3P on upregulation of CD25+ CD8 T cells from infected cows, while L5P effects were static. In contrast, L5P demonstrated a significantly stronger induction of CD25+ B cells from infected animals compared to L3P.

Accepted 22 May, 2017. For correspondence. *E-mail John.Bannantine@ars.usda.gov; Tel. 515-337-7340; Fax 515-337-7458. **E-mail Franck.Biet@tours.inra.fr; Tel. 33 247 42 78 69; Fax 33 247 42 77 79.

Introduction

Mycobacterium avium subsp. *paratuberculosis* (*Map*) is the causative agent of Johne's disease in cattle and other ruminants, a chronic progressive intestinal disease that is difficult to accurately diagnose, especially in the early stages of disease. This problem has hindered efforts to control or eliminate the disease from dairy herds. Johne's disease is widely distributed on five continents and the most affected countries are in North America, Europe and Australia (Biet and Boschiroli, 2014). According to the most recent National Animal Health Monitoring System survey (NAHMS, 2008), Johne's disease prevalence has increased to over 90% of U.S. dairy herds. Four distinct stages of disease progression have been described as silent infection, subclinical disease, clinical disease and advanced clinical disease (Whitlock and Buergelt, 1996). The last two stages often develop after several years of infection. Typical advanced stage signs are weight loss, diarrhea, lethargy and increased weakness. However, the economic toll on the dairy industry is the primary motivation for efforts at disease control. A United States Department of Agriculture (USDA) study estimated the loss of approximately $200 per cow each year with an overall economic loss of between $200 million to $250 million dollars annually to the U.S. dairy industry.

The mycobacterial cell envelope is unique among prokaryotes in that it contains unusual lipid and carbohydrate compounds, such as lipoarabinomannan and mycolic acids. The mycomembrane, an unusual outer membrane, corresponds to the permeability barrier of which the inner leaflet is formed by a parallel arrangement of mycolic acids covalently linked to parietal backbone. Mycolic acids are the main components of this mycomembrane and constitute up to 60% of the lipid content of the cell wall (Marrakchi *et al*., 2014). The outer-most layers of the cell envelope, or capsule, are composed of several types of glycolipids embedded in the saccharidic matrix surrounding the bacillus. The three major classes of type- or species-specific glycolipids include the lipooligosaccharides, phenolic glycolipids and glycopeptidolipids (GPLs). It is not clear why different lipids exist among these species of mycobacteria, but these differences have been exploited to catalog and distinguish species in this genus. Rapid growing mycobacteria, including *Mycobacterium chelonae*, *Mycobacterium scrofulaceum*, *Mycobacterium abscessus* and slow growing mycobacteria such as *M. avium* subsp. *avium* produce GPLs in their cell envelope (Ripoll *et al*., 2007). Recently, *M. abscessus* strains showing rough colony morphology were found to lack GPLs and the majority of these variants could be linked to a detrimental mutation in one of eight genes in the GPL locus (Park *et al*., 2015).

Map is considered as a genetically homogenous subspecies of *M. avium*, especially among bovine, human and wildlife isolates (Wu *et al*., 2006; Paustian *et al*., 2008). However, two primary lineages have emerged following extensive phylogenetic analyses and comparative genomic studies (Biet *et al*., 2012). These lineages are classified as type I/III or S-type (ovine) and type II or C-type (bovine) strains. *Map* appears to have emerged from the common ancestor, *M. avium* subsp. *hominissuis*, to yield these two lineages (Turenne *et al*., 2008; Alexander *et al*., 2009). The *Map* C-type was first isolated from cattle and is the most commonly isolated type, while the *Map* S-type are typically isolated from sheep and are apparently less prevalent. The S-type isolates are readily distinguishable from C-type isolates based on genome sequencing studies (Li *et al*., 2005; Bannantine *et al*., 2012). But these two lineages can also be readily discriminated by genotyping methods due to single nucleotide polymorphisms (Marsh *et al*., 1999) as well as deletions/insertions of large DNA segments (termed large sequence polymorphisms or LSP) using phylogenetic techniques such as variable number tandem repeats (Lefrancois *et al*., 2013), single sequence repeats (Amonsin *et al*., 2004; Thibault *et al*., 2008), representational difference analysis (Dohmann *et al*., 2003) and *hsp65* sequencing (Turenne *et al*., 2006). Furthermore, genomic hybridization of S-type strains on a C-type microarray revealed a large 23-gene deletion in S-type strains (Marsh *et al*., 2006). However, in no case has a genetic difference been linked to a phenotypic difference between C- and S-type strains, until this study.

In addition to the genotypic distinctions between S- and C-type strains, phenotypic differences involving growth characteristics have been noted since the middle of the last century (Taylor, 1951). The S-type strains are more fastidious and have slower growth rates in laboratory media than C-type strains. In contrast to C-type, the S-type strains do not grow readily on Herold's egg yolk media or Middlebrook 7H9 media that is not supplemented with egg yolk, mycobactin J and ADC enrichment (Whittington *et al*., 2011). Nutrient limitation will kill S-type strains but it is only bacteriostatic for C-type (Gumber *et al*., 2009). Motiwala and coworkers have shown transcriptional changes in human macrophages infected with C-type, human and bison isolates, which induce an anti-inflammatory gene expression pattern, while the *Map* S-type isolates that were analyzed showed expression of pro-inflammatory cytokines (Motiwala *et al*., 2006), (Stevenson *et al*., 2002; Biet *et al*., 2012). Furthermore, many of the S-type strains are pigmented while C-type strains are not. On the transcriptional level, C- and S-type strains exposed to low iron or heat stress conditions had different mRNA expression patterns (Gumber and Whittington, 2009). Furthermore, iron storage in low iron conditions was only observed in the C-type but not S-type strains (Janagama *et al*., 2009) and virulence adhesin differences were characterized (Lefrancois *et al*., 2013). In this study, differences in a lipopeptide that is a component of the mycobacterial cell envelope were identified between C- and S-type strains.

Nonribosomally synthesized peptides include a diverse class of important metabolites such as antibiotics. Nonribosomal peptides (NRP) are usually 3–10 amino acids in length and are synthesized by large multi-modular enzymes called nonribosomal-peptide synthetases (NRPSs). As the name implies, these peptides are not assembled by ribosome, but rather are RNA template and ribosomal independent to allow for maximum biological flexibility by incorporating many unique amino acids. Although 10% of bacterial NRPS genes are non-modular (Wang *et al*., 2014), most have a modular organization where each module specifies the sequential addition of an amino acid. Several kilobases of DNA are needed for each module that consists of three domains termed the adenylation domain, peptidyl carrier domain and condensation domain (Mootz *et al*., 2002). The adenylation domain binds ATP, selects its cognate amino acid building block and performs substrate acyl adenylation. Amino acid translocation occurs with the peptidyl carrier domain. The largest NRPS yet

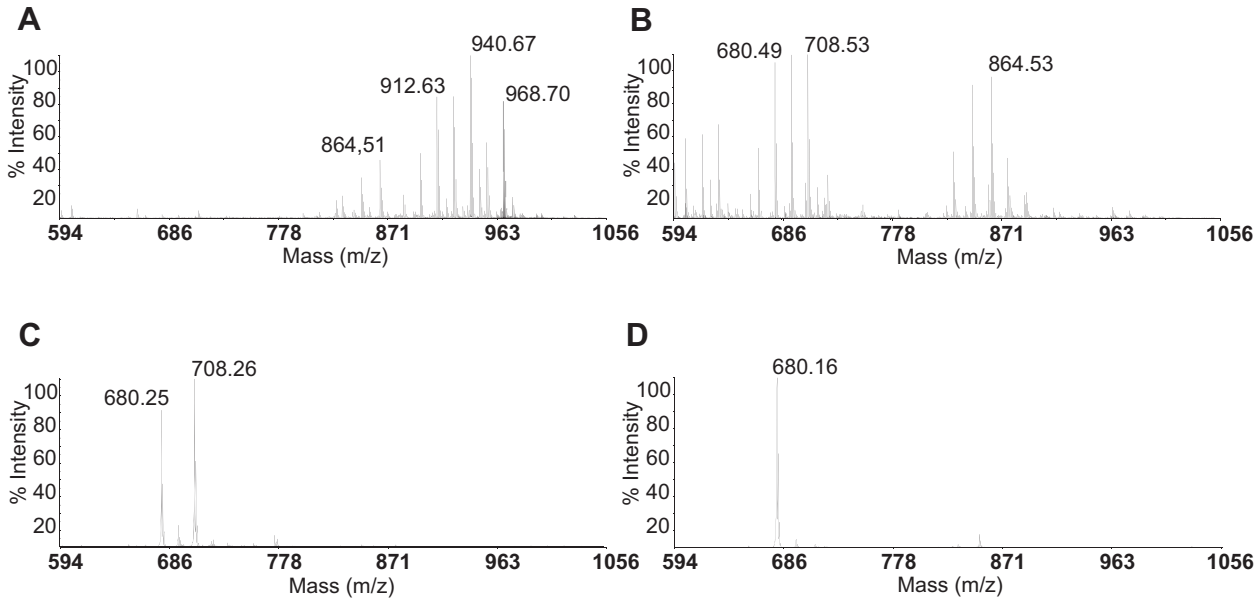

Fig. 1. Mass spectrometry analysis of the lipids from *Map*.
MALDI-TOF spectra of chloroform/methanol-extracted lipids from C-type *Map* K-10 (A), S-type *Map* S397 (B), purified native L3P (C) and synthetic L3P (D). The peak at 940 amu corresponds to L5P in the lipid extract of the C-type strain K-10, but is absent from the native lipids extracted from S397. The peak at 680 amu corresponds to the L3P.

discovered is from *Photorhabdus luminescens* (WP_011146892; 16,367 aa) and contains 15 modules (Wang *et al*., 2014). This may represent the upper limit of NRPSs. In C-type *Map* the *mps1* gene encodes a NRPS with five modules that have been previously shown to be involved in production of the pentapeptidic moiety of the lipopentapeptide (L5P) (Biet *et al*., 2008).

The objective of this study was to identify the composition of lipopeptides in the S-type strains of *Map* and determine if they are different from the C-type strains.

Genetic characterization allowed us to predict the production of different lipopeptide components, depending on the strain type. Synthesis of the predicted S-type lipopeptide together with thorough biochemical and physico-chemical analyses demonstrated that typical lipopeptides from *Map* are different in S-type (lipotripeptide) and C-type strains (lipopentapeptide). Overall, we reveal key elements of *Map* cell wall change, involving genes and lipopeptides, occurring on the patho-evolution of the subspecies *paratuberculosis*.

Results

The lipid composition differs between C- and S-type strains of Map

A panel of genetically diverse *Map* strains isolated from different animal species appears similar in their lipid profiles when analyzed by thin layer chromatography (TLC) in a single dimension (1-D) (Biet *et al*., 2008). However, the analysis of extracted lipids from both the S397 and K-10 *Map* (sequenced strains characteristic of S- and C-type respectively) revealed a striking difference by Matrix-Assisted Laser Desorption Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS). Only the C-type strain showed a major peak at a mass-to-charge ratio (*m/z*) of 940 atomic mass units (amu) (Fig. 1A), which corresponds to the $[M + Na]^+$ ion of the previously characterized L5P (Riviere *et al*., 1996). Additional minor peaks were also observed differing by 14 amu (including a peak at *m/z* 968 amu), all of which are present uniquely in the C-type strain, and assigned to variable lengths of the fatty acid moiety of the L5P (Eckstein *et al*., 2006; Biet *et al*., 2008). Instead of the ion peaks at *m/z* 940 ± 14 amu, the extracted lipids from the S397 strain show three major peaks at 680, 694 and 708 amu (Fig. 1B). The rest of the MS spectra were nearly identical between the two strains. These data indicate that the lipid composition of the S397 sheep strain is different from that of the C-type strains and does not include the L5P molecule.

The mps1 *gene is different between C- and S-type strains*

A comparative genomic study was performed to determine the genetic basis for the absence of L5P in S397. While approximately 28 genes are necessary for GPL

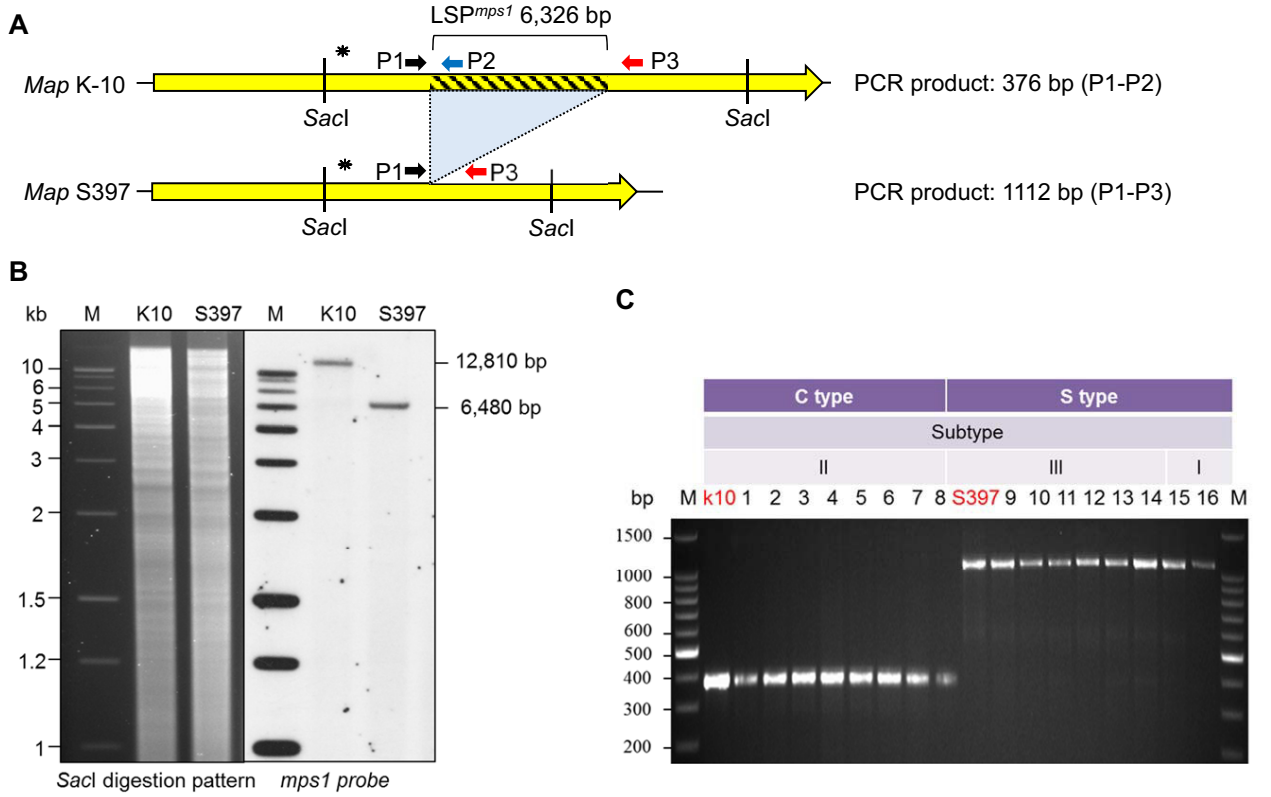

Fig. 2. DNA analysis of the *mps1* deletion in the S-type strains.
A. Schematic illustration showing the *mps1* coding region in K-10 and S397. The yellow arrow indicates the *mps1* coding sequence with the *Sac*I sites and primer binding locations shown. The asterisk shows the location of the labeled probe used in the experiment. The shaded area represents the 6.3 kb segment that is present uniquely in the K-10 strain. The same primers were used for the experiment in panel C.
B. Agarose gel and Southern blot of *Sac*I-digested genomic DNAs. The right half of panel B shows the respective sizes of the *Sac*I fragment after hybridization with the labeled probe. Molecular size standards (M) are indicated in kilobase pairs in the left margin. The K-10 fragment is over 10 kb and the S397 fragment is approximately 6.5 kb.
C. Amplification products from a panel of C- and S-type *Map* DNAs using a three-primer amplification approach where P1 is the forward primer and P2 and P3 are the reverse primers used in a single reaction. The primers were designed such that the resulting PCR products would be of different sizes depending on the presence or absence of the LSP^{mps1}. This experiment was performed using a collection of 18 C-type and S-type strains previously characterized and genotyped (Biet *et al.*, 2012). See Table S2 for details about these strains.

biosynthesis (Ripoll *et al.*, 2007), the peptide core of L5P in *Map* is assembled by the product of a single *nrps* gene, termed *mps1* (Biet *et al.*, 2008) (Table S1). The *mps1* gene of *Map* K-10 is also known by the locus tag MAP_1420 and has a size of 19.15 kb encoding 6,384 amino acids (Li *et al.*, 2005). This gene is under the control of the LuxR regulator and has shown increased transcription when exposed to cow's milk (Alonso-Hearn *et al.*, 2010). It has been suggested that the pentapeptide moiety is nonribosomally assembled by the modules encoded in this gene (Eckstein *et al.*, 2006), therefore, it was of interest to examine the homolog in the S-type strain. However, previous *de novo* whole genome assemblies of the *Map* S397 genome using the available Roche GS20, Roche FLX (i.e. 454), and Sanger sequence data (Bannantine *et al.*, 2012) were unsuccessful at producing a complete assembly of the *mps1* gene due to the large size and the presence of long, highly syntenic repeats in the amino acid specifying modules. Therefore two large sequence gaps were present in *mps1* in the S-type genome (Fig. S1).

While genome sequencing revealed that the *mps1* gene is present in the S-type strain, the question of why that strain does not produce L5P remained unanswered. To address this, additional sequence data were obtained to completely assemble the region containing *mps1* in the S397 genome (Table S1). Surprisingly, the *mps1* gene was only 12,822 bp in size compared to 19,148 bp in the K-10 genome, representing a difference of 6,326 bp. Southern blot analysis was used to confirm the 6.3 kb deletion (Fig. 2A and B). By taking advantage of two *Sac*I restriction sites that border the deletion (Fig. 2A), it was observed that the S397 *Sac*I fragment was

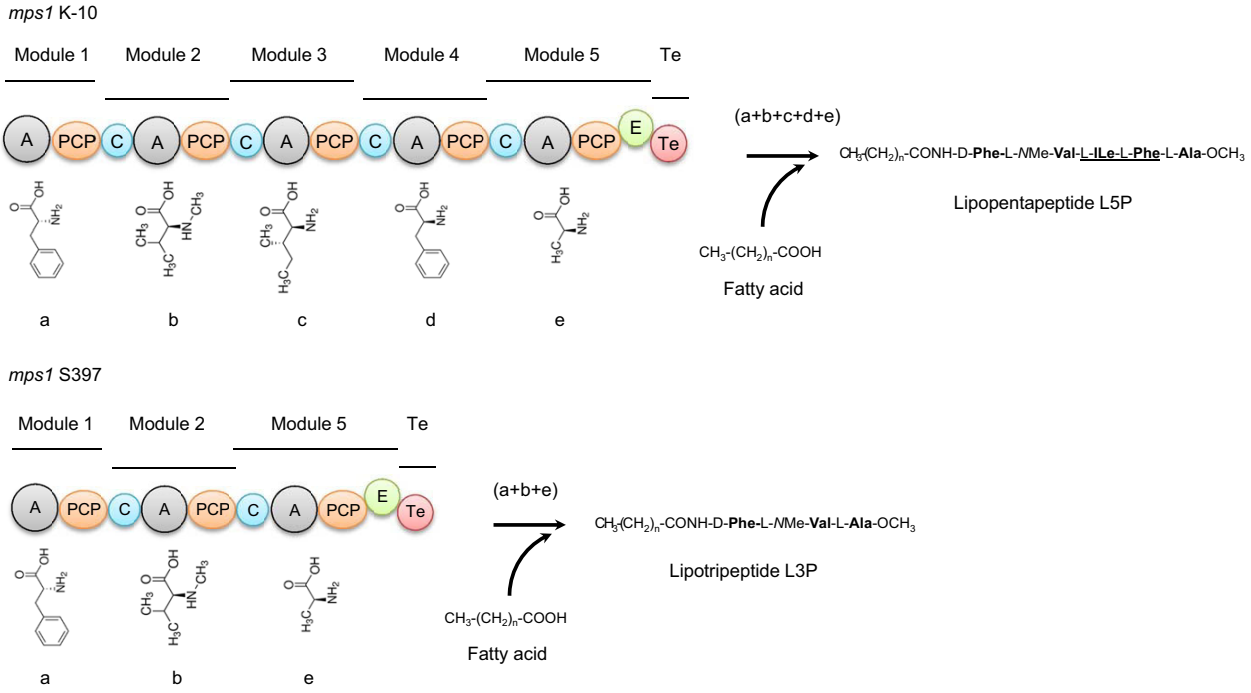

Fig. 3. Proposed model for NRP assembly of L3P and L5P in *Map*. Shown are the modules and domains predicted for Mps1 in K-10 and S397. Based on comparative sequence analysis, modules 3 and 4 are predicted to be absent in S397. When the 3- and 5-amino acid peptide moieties are combined with the fatty acid ($n = 18–20$), the lipopeptide emerges. The underlined amino acids in L5P are missing in L3P. C = condensation domain; A = adenylation domain; PCP = peptidyl carrier domain; E = epimerisation domain and Te = thioesterase domain to release the full-length peptide chain.

approximately 6 kb smaller than the corresponding fragment in K-10 (Fig. 2B).

The deletion was further characterized by PCR analysis and tested across multiple strains (Table S2). To verify that the difference in size of the *mps1* gene is characteristic of all sheep strains, a PCR test to detect this large sequence polymorphism (LSP^{mps1}) was developed based on the model described by Semret *et al.* (2006). From the *mps1* locus in K-10, three primers (P1, P2 and P3, Table S3) were designed and used in a single reaction to amplify DNA depending on the presence or absence of the 6.3 kb region (Fig. 2A). The primers were designed so that the size of the PCR product is different when using this 3-primer combination. The P1–P3 pair results in no amplification from C-type DNA due to the large distance between primers. However, P1–P2 results in successful amplification since they are only separated by 376 bp (Fig. 2C). Conversely, the P1–P2 primer combination does not work in S-type strains since P2 is located within the LSP^{mps1} (Fig. 2A). However, P1–P3 does amplify S-type DNA because they are only separated by 1,112 bp due to the LSP^{mps1} (Fig. 2C). Collectively, these results confirmed the boundaries of the deletion and showed it is consistent in all ten S-type strains tested including characteristic subtypes I and III.

NRPS encoded by mps1 *is missing modular domains in the S-type strains*

The NRPS of *mps1* is modular in its organization such that each module specifies the incorporation of one amino acid in the peptidic moiety of the lipopeptide. It became of interest to examine how the LSP^{mps1} deletion might have affected lipopeptide production in the S-type strains. Using bioinformatics (Rottig *et al.*, 2011), the functional modules and domains within each module of the NRPS were identified and this analysis established that the S-type NRPS is composed of three modules while the C-type has five modules. Furthermore, these analyses have established the nature and the position of the NRPS domains in S-type along with the domains present in C-type but missing in the S-type strain (Fig. 3). Comparison of the protein sequences corresponding to the three domains of Mps1 present in both strains shows a perfect homology suggesting a same functionality in terms of amino acid assembly (Fig. S2).

Altogether, the sequence analysis and bioinformatic predictions of NRPS module composition identified the tripeptide Phe-Val-Ala as the antigen backbone. By analogy with the known L5P, we therefore predicted that the S397 strain produces a lipotripeptide, named L3P, bearing the same structural formula as L5P but missing the two amino acids L-Ile and L-Phe (Fig. 3).

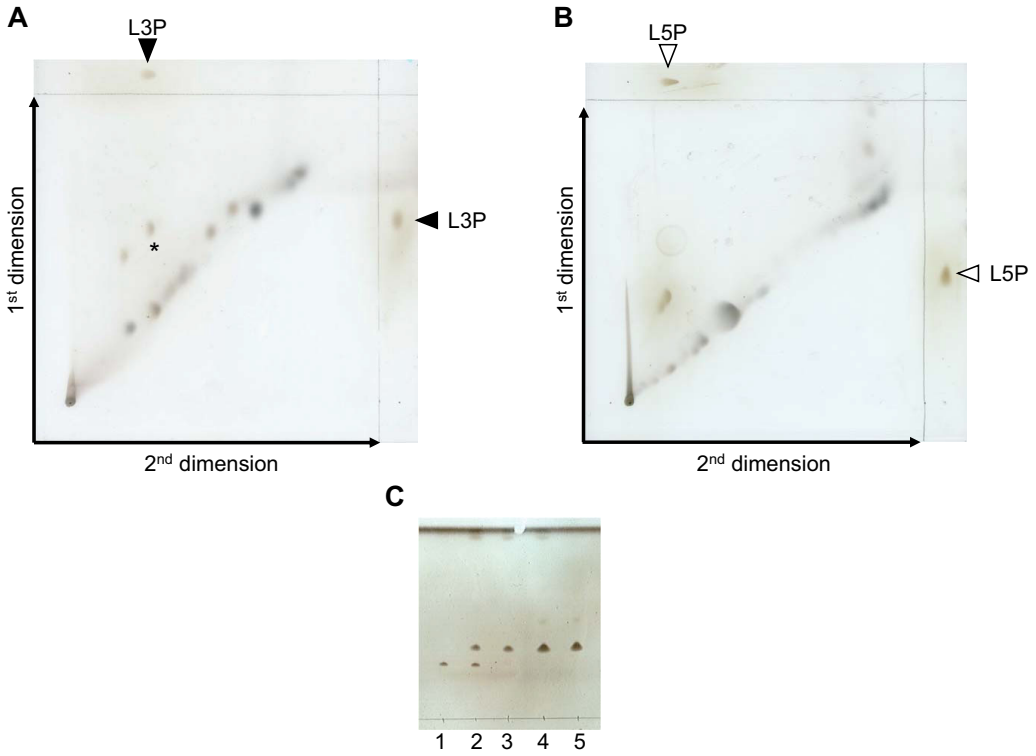

Fig. 4. S-type *Map* contains L3P.
2-D TLC of total S397 (A) and K-10 (B) lipids (500 (A) and 250 (B) μg spotted respectively), using chloroform/methanol (96:4) as first dimension and toluene/acetone (80:20) as second dimension, and chemically synthesized L3P or L5P (15 μg each) as markers (black and white arrows respectively). The asterisk indicates the position of native L3P. C. 1-D TLC of the purified native L3P (line 3), compared to the synthetic controls L5P (line 1) and L3P (line 5) using chloroform/methanol (95:5) as the solvent system. The samples are also loaded as mixtures of the adjacent spots: synthetic L5P and purified native L3P (line 2), purified native L3P and synthetic L3P (line 4). The TLC plates were sprayed with 10% copper sulfate in 8% phosphoric acid, and lipids were visualized by heating.

S-type strains produce a lipid antigen identical to the synthetic lipotripeptide L3P

To determine if S-type *Map* effectively produces this novel L3P antigen, the L3P molecule was chemically synthesized and compared with the native source of lipid (either the crude or the purified lipid extract from S397).

The synthetic L3P was obtained by solid-phase peptide synthesis using Fmoc chemistry and purified by chromatography on silica gel. It was then used as a control in a series of physico-chemical comparative analyses to formally identify the S-type lipid antigen.

Analysis and purification by TLC. The analytical two-dimensional (2-D) TLC of S397 lipid extracts shows a spot, not as prominent as L5P in C-type strains (Fig. 4A and B), co-migrating with the synthetic L3P (Fig. 4A). After loading a preparative 2-D TLC with 7 mg of crude extract obtained from 317 mg of cells (dry weight), the spot of interest was purified by scraping the silica gel and subsequent elution in CH_2Cl_2/methanol 95:5 (vol/vol). The resulting purified native antigen (approximately 50 μg) is clearly different from the C-type *Map* L5P, and it co-migrates with the synthetic L3P, as shown by the 1-D TLC (Fig. 4C).

Analysis by MALDI-TOF MS and MS/MS. The peak of the synthetic L3P at *m/z* 680 amu ($[M + Na]^+$ ion) matches that of the native antigen from the S397 strain, whether in the crude extract or in the purified lipids (Fig. 1B–D). The extra peaks differing by 14 amu (i.e. one methylene unit) observed for the native antigen (Fig. 1B and C) suggest the presence of different fatty acid chain lengths. In particular, the compound at *m/z* 708, which co-elutes with the L3P in 2-D TLC, may correspond to the L3P with a C22 acyl chain. The presumed L3P antigen is *O*-methylated at the *C*-terminus, as are the synthetic L3P and the L5P from C-type *Map* (Biet *et al.*, 2008). Indeed MALDI-TOF MS of both the synthetic and native L3P compounds showed, after saponification, a down-shift of 14 amu of the molecular ion, due to the hydrolysis of the *O*-methyl ester group from the *C*-terminus (data not shown).

Additional MS/MS analysis was conducted to confirm the structure of the putative L3P compound. Importantly,

A

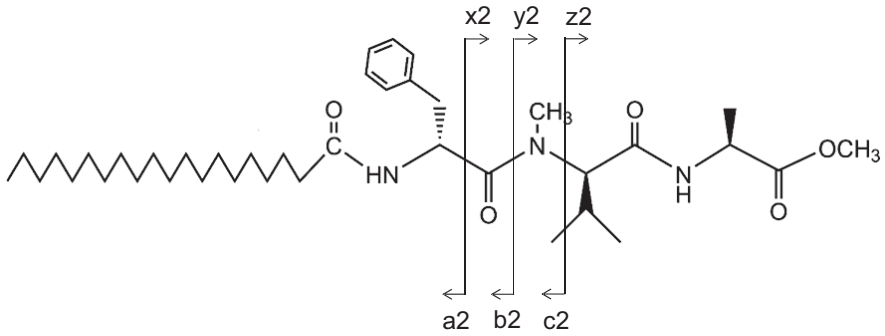

Fig. 5. Tandem MS spectra of purified native and synthetic L3P show identical fragmentation patterns. L3P purified from S397 lipids was analyzed by MALDI-TOF MS/MS and compared to synthetic L3P and L5P.
A. Structure of L3P and typical fragmentation at the Phe-*N*-Methyl-Val bond.
B. Ions originating from the fragmentation at the Phe-*N*-Methyl-Val bond.

B

Antigen	L3P	L3P	L3P	L3P	L5P
Source	native	synthetic	native	native	synthetic
Parental ion (*m/z*)	680.7	680.7	694.5	708.6	940.7
Fatty acyl chain	C20	C20	C21	C22	C20
a2	436.7	436.7		(464.5)*	436.5
x2	(267.3)	(267.3)			527.4
b2	464.6	464.6	(478.4)	492.5	(464.5)
y2	239.3	239.3	239.3	239.2	499.4
c2	495.7	495.7		523.4	495.5
z2	208.3	208.3	208.2	208.2	468.4

* in brackets : peak of low intensity

the purified S397 lipid and the synthetic compound displayed identical MS/MS spectra (Fig. S3). Probably because of the unusual structure of the lipopeptide, specifically the acylation at the *N*-terminus and the presence of an *N*-Methyl-Val residue, fragmentation of the peptide moiety of the synthetic L3P did not yield all the expected canonical couples of fragment ions, namely, the [a, b, c] ions from the *N*-terminus and the [x, y, z] ions from the *C*-terminus (Fig. 5A). Nevertheless, from the parental ion *m/z* 680 amu, the major observed fragmentation peaks were shared between synthetic and native L3P and were totally in agreement with the structural formula of the L3P. Representative fragment ions detected and corresponding to all the possible cleavages of the peptidic bond between the Phe and the *N*-Methyl-Val residues are shown in Fig. 5B. This bond has been chosen because (i) it gives the most complete sampling of the various fragment ions expected after cleavage of a peptide bond (Fig. 5A) and (ii) it is conserved in both the L3P and L5P lipopeptides and ideally located to discriminate between these two compounds. Indeed, L3P and L5P share a common structure at the *N*-terminus, from the C20 fatty acid to *N*-Methyl-Val (Fig. 3). MS/MS analysis of the parental ions at *m/z* 694 and 708 amu of the native L3P variants confirmed the identity of the [a, b, c] fragment ions (*N*-terminal moiety of the lipopeptide with variation of 14 or 28 amu for the fragment ions, according to the variation of the length of the fatty acyl chain) and of the [x, y, z] fragment ions (invariant *C*-terminal moiety regardless of the length of the fatty acyl chain) (Fig. 5B). Fewer expected fragment ions were detected with the 694 species of the L3P, probably due to the lower intensity observed with the corresponding parental ion.

Finally, MS/MS analysis of the L5P parental ion at 940 amu confirmed the assignment of these fragment ions: the [a, b, c] ions were identical between L3P and L5P, and the [x, y, z] ions increased in agreement with the presence of two additional amino acids (Fig. 5B). Collectively, these data are consistent with an identity of structure between the purified native S397 lipid and the synthetic L3P, i.e. a tripeptide sequence Phe-*N*-Methyl-Val-Ala with a *N*-ter C20 fatty acid and a *C*-ter methyl ester.

Analysis by nuclear magnetic resonance (NMR) spectroscopy. To confirm the structure of the native antigen, ^{1}H-NMR spectroscopy was performed on the presumed L3P purified from the lipid extract of S397 cells.

Results of the NMR analysis were in agreement with the structure proposed for the native L3P. ^{1}H-NMR spectra of the purified S397 lipid and the synthetic L3P are overlapping (Fig. S4A and B), showing all the characteristic peaks for Phe, *N*-Methyl-Val and Ala, including peak multiplicities, coupling constants and chemical shifts (Table 1). The spectra revealed three resonances

Table 1. Characteristic ^{1}H NMR data for the native purified L3P

Chemical shift (p.p.m.)	Peak multiplicity, Coupling constant	Assignment[a]
0.61	Doublet, $J = 6.8$ Hz	γ-C$\underline{H}_3$ Val
0.97	Doublet, $J = 6.4$ Hz	γ-C$\underline{H}_3$ Val
1.35	Doublet, $J = 7.4$ Hz	β-C$\underline{H}_3$ Ala
2.22	Multiplet	β-C$\underline{H}$ Val
2.92	Singlet	*N*-C$\underline{H}_3$ Val
2.95/3.06	2 Doublets of doublet, $J = 13.4$ Hz	β-C$\underline{H}_2$ Phe
3.52	Singlet	O-C$\underline{H}_3$
4.47	Doublet, $J = 10.9$ Hz	α-C$\underline{H}$ Val
4.50	Pentet	α-C$\underline{H}$ Ala
5.20	Multiplet	α-C$\underline{H}$ Phe
6.08	Doublet, $J = 7.9$ Hz	N$\underline{H}$ Phe
6.52	Doublet, $J = 7.6$ Hz	N$\underline{H}$ Ala

The synthetic L3P gives similar data.
a. The assigned protons are underlined.

characteristic of the alpha protons of Phe, Val and Ala at 5.20, 4.47 and 4.50 p.p.m. respectively. Two resonances typical of the amide region instead of three, between 6.0 and 7.0 p.p.m., confirm that one of the amino acids has no amide proton. The presence of a singlet at 2.92 p.p.m. is consistent with the presence of a *N*-Methyl group on this amino acid.

The assignments (Table 1) were determined by the ^{1}H-^{1}H-COSY NMR experiment where typical spin systems were observed for the three amino acids (Fig. S5).

^{1}H-NMR spectrum of the purified S397 shows additional peaks in comparison to the synthetic L3P (Fig. S4A and B). These peaks may originate from distinct contaminant compound(s) which partially co-elute with the L3P during the preparative 2-D TLC. Indeed, the ^{1}H-^{1}H-COSY NMR spectra show that spin systems of the extra peaks are not linked to any of the L3P peaks (Fig. S5). Moreover, when the preparative TLC silica gel was scraped in the zone adjacent to that of L3P, the resulting eluted compound unambiguously gave a 1H-NMR spectrum displaying all the peaks that could not be attributed to L3P in the characteristic range from 2 to 5 p.p.m. (Fig. S4C). Due to the resolution limit of the 2-D TLC and to the very low amount of native antigen, the complete purification of the antigen could not be achieved.

Nevertheless these results, together with the MS data highlighting the presence of L3P, demonstrate that the S397 strain produces a lipid content with, at least, the L3P compound.

Analysis of the optical purity. Finally, the optical purity of the individual amino acids within the native L3P was determined by gas chromatography coupled to MS after hydrolysis of the lipopeptide in 6N DCl in D_2O.

The results demonstrated the presence of the enantiomeric forms of D-Phe (91.4%), *N*-Methyl-L-Val (99.0%) and L-Ala (98.3%) (data not shown). Notably, in the course of this analysis, the identity of the three predicted amino acids was also confirmed based on their retention time and their mass spectra. Overall, the structure proposed for the L3P (Fig. 5A) produced by S-type *Map* from the sequence of the *mps1* gene has been confirmed: a peptidic core as D-Phe-*N*-Methyl-L-Val-L-Ala attached mostly to a 20-carbon fatty acid chain.

Lipopeptides are cell surface-exposed

It has been assumed for a long time that L5P is localized in the cell wall of *Map*, but to the best of our knowledge this has never been experimentally demonstrated. Analysis by MALDI-TOF MS of the lipids extracted from surface-exposed materials of *Map* K-10 showed that L5P is localized in the outer-most layers of the cell envelope (Fig. S6A and B). Control TLC established that cord factor, a lipid which is never exposed at the mycobacterial cell surface (Ortalo-Magne *et al.*, 1996) is indeed absent from the surface-exposed material analyzed here (data not shown), thus strengthening our conclusions.

Similarly, L3P was detected in surface-exposed materials prepared from *Map* S397 (Fig. S6D). MS/MS analysis of the compound at *m/z* 680 confirmed its identity as L3P, since all the representative fragment ions are present (data not shown). Minor amounts of cord factor were also detected in the surface extract of S397 (data not shown), suggesting a certain degree of cellular lysis for that strain. Nevertheless, the fact that the cell-bound and surface-exposed fractions displayed different lipid compositions (Fig. S6C and D) suggest that L3P should be present at the cell surface of the S–type strain. But additional experiments are needed to confirm this localization.

In both cases, detection of lipopeptides in the cell-bound lipidic fraction (Fig. S6A and C) implies that they are also present within deeper layers of the cell envelope.

L3P promotes a cell-mediated immune response whereas L5P promotes B cell responses

After 24 h culture, there was a dose-dependent proliferation of CD25+ CD8 T cells from infected cows stimulated with L3P. By contrast, L5P stimulated cells remained static over the range of lipopeptide concentrations (Fig. 6A). S397, which contains L3P, produced a slightly stronger response than K-10, which has L5P, although this difference was not significant (data not shown). In contrast, effects of lipopeptides on CD25+ B

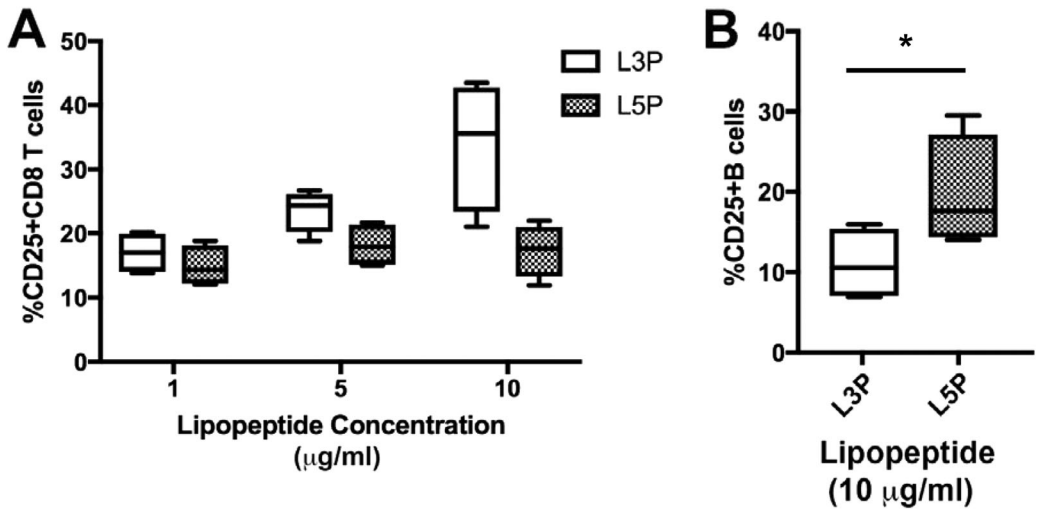

Fig. 6. T cell and B cell response to the *Map* lipopeptides.
Proliferation of CD25+ T cells (A) and B cells (B) after culture of PBMCs isolated from cows naturally infected with *Map* and exposed to lipopeptides. Results are presented as percentage of CD25+ T or B cells (mean ± SEM). Significant differences are designated by an asterisk ($p < 0.05$). Note the relative difference between the lipopeptides at the 10 μg ml^{-1} concentration for T cells and B cells.

cells were reversed as L5P promoted a significantly ($p < 0.0002$) stronger response compared to L3P (Fig. 6B). No significant differences were observed between L3P and L5P in CD25+ CD4 or CD25+ γδ T cell populations (data not shown). Both L3P and L5P elicited cytokine responses to IFN-γ, IL-1β and TNF-α with no significant differences between the L3P or L5P treatments (Fig. S7). However, significant differences were observed between infected and control cells ($p < 0.0001$ for IFN-γ and IL-1β, $p < 0.03$ for TNF-α; Fig. S7). Interestingly, a dose-dependent effect ($p < 0.0006$) of L5P concentration was observed on TNF-α secretion by peripheral blood mononuclear cells (PBMCs). IL-4 and IL-6 were not detected following stimulation with either lipopeptide (data not shown).

Discussion

In the process of characterizing the differences in lipids among C- and S-type strains of *Map*, we uncovered a new LSP not previously described. LSPs have been shown to distinguish *Map* from other *M. avium* subspecies, including *hominissuis* and *silvaticum* (Alexander *et al*., 2009). In addition, three S-type-specific LSPs were characterized by genomic hybridization to DNA microarrays (Marsh *et al*., 2006). While these LSPs usually span several genes and range in size from 4.5 kb to over 65 kb, the LSP reported here is located exclusively within the *mps1* gene and spans 6.3 kb of DNA present in C-type strains, but not in any of the S-type strains examined. It is likely that this LSP remained hidden, despite extensive genomic comparison studies, because it is entirely contained within a single gene. This newly discovered LSP now provides an additional target to distinguish S-type from C-type strains of *Map*.

Over 10% of the mycobacterial genome is coded for proteins involved in lipid metabolism. Large genes, including *mmpL/S*, *pks* and *nrp* are involved in lipid biosynthesis or transport (Ripoll *et al*., 2007), but the role of each of these needs to be determined by investigating genetic differences and correlating those to phenotypic differences as has been accomplished for lipooligosaccharides in *M. smegmatis* (Etienne *et al*., 2009). Although numerous genetic differences between C- and S-type *Map* strains have been reported, our results represent the first example of a genetic difference that has been phenotypically defined. It had been previously thought that all *Map* strains produce L5P since only one bovine strain had been evaluated by 2-D TLC (Eckstein *et al*., 2006) and several other *Map* strains examined by 1-D TLC (Biet *et al*., 2008); however, 1-D TLC did not resolve differences due to limits of the technique. The difference in lipid composition was discovered only through extensive biochemical and physico-chemical analysis of lipid extracts combined with detailed sequence and assembly of the large and highly repeated *mps1* gene in the S-type strain.

Based on TLC analysis, *Map* does not produce GPLs but instead contains a lipopeptide molecule (Biet *et al*., 2008) initially termed Lipopeptide-I (Riviere *et al*., 1996) and later Para-LP-01 (Eckstein *et al*., 2006). This nonpolar lipid, most recently termed L5P for lipopentapeptide, is an abundant molecule in *Map* and is not

detected in *M. avium* subsp. *avium* (Eckstein *et al*., 2006). Our group has previously demonstrated that L5P is antigenic in antibody-based tests (Biet *et al*., 2008; Verdier *et al*., 2013) with negligible cell-mediated immune responses (Holbert *et al*., 2015). In the present study, L5P preferentially resulted in the upregulation of activated B cells (CD25 + B cells), a finding that correlates with previous studies demonstrating this lipopentapeptide produces strong humoral responses in cattle and sheep (Biet *et al*., 2008). In contrast, L3P more distinctly upregulated T cell proliferation (CD25 + CD8 T cells) in a dose-dependent manner, suggesting more of a Th1 immune response to this cell wall component. These results suggest that genomic differences between L3P and L5P may translate to antigenic differences that present immunological diversity within the infected host.

We further show for the first time that L5P is clearly surface-exposed, i.e. localized in the outer-most layers of the cell envelope. The unique mycobacterial cell wall is important in the physiology of these bacteria (Kieser and Rubin, 2014) and has been studied for its immune stimulating adjuvant properties (Chedid *et al*., 1972; Lederer *et al*., 1975), resistance to stressors, physical disruption and increased virulence (Howard *et al*., 2006; Bernut *et al*., 2014). Considering that the lipopeptides and GPLs are surface-located in the cell envelope and that their presence or absence differs among a small subset of mycobacteria, their physiological properties may change greatly depending on the mycobacterial strains and their evolutionary history.

NRPSs create substantial biological flexibility because no ribosomes or RNA template are needed for peptide assembly. The ribosome recognizes only 20 naturally occurring amino acids for peptide assembly; however, NRPS can specify over 500 amino acids, creating unlimited peptides for highly specialized biological functions (Walsh *et al*., 2013). In this study we showed that the tri-peptide produced in S-type strains consists of only one naturally occurring amino acid, L-Ala, and two that are 'non-coded' amino acids. The C-type *mps1* has five modules encoding a lipopentapeptide, but there are examples of two NRP genes, arranged in tandem, that together encode a five module NRPS to construct the antibiotic nocardicin A (Gaudelli *et al*., 2015). Perhaps to further increase diversity in these NRP, known NRPSs can be classified into three groups, linear, iterative and nonlinear (Mootz *et al*., 2002). In linear NRPSs, the sequence of the resulting peptide chain is entirely determined by the number and order of the modules. Iterative NRPSs use their modules or domains more than once in the assembly of one single product. Nonlinear NRPSs involve complex scenarios with parallel nonlinear organization of domains and unusual arrangements such as internal cyclization or incorporation of small soluble molecules (Mootz *et al*., 2002). Data from this study show that *mps1* for both L3P and L5P NRPSs are linear in organization.

Could the defined change in peptide length described in this study be enough to account for host preferences in C- and S-type strains of *Map*? S-type has a substantial host preference for sheep, but not exclusively, since S-type has also recently been isolated from several Arabian camels (Ghosh *et al*., 2012) as well as goats, cattle and deer (Stevenson, 2015). However, C-type appears to have a broader host range since it has been isolated from a variety of ruminant species, including goat, deer and bison (Sibley *et al*., 2007; Biet *et al*., 2012), but has also been isolated from non-ruminant livestock, wildlife and also humans. Nonetheless, there is a clear host preference or adaptation among these strains. It may also be possible that this subtle change in peptide length could define the growth rates or other phenotypic differences between these types. However, we can exclude the fact that this NRP is responsible for pigment production reported in the S-type strains (Biet *et al*., 2012), since we observed that L3P is colorless (data not shown). Regardless, it is clear that both lipopeptides share common epitopes since D-Phe, *N*-Methyl-L-Val and L-Ala are conserved in both *Map* types. The two amino acids missing from the S-type strain L3P are L-Ile and L-Phe. Mutational studies may partially address the biological effect of these lipopeptides since it is unclear at the moment.

Rough and smooth colony appearance among *Mycobacterium* species is not only attributed to changes in their lipid composition (Wright *et al*., 1996) but also to virulence and drug resistance (Kansal *et al*., 1998; Howard *et al*., 2006). In fact L5P disappears when *Map* are cultured in cow's milk but is present in high abundance when cultured in Middlebrook 7H9 media (Alonso-Hearn *et al*., 2010), suggesting that the lipid profile of *Map* changes significantly when exposed to different environments. But there may be much more going on biologically that accounts for these lipid differences. Only recently were lipopeptides shown to interact with TLR2 receptors on key immune cells (Jimenez-Dalmaroni *et al*., 2015). Much research is still needed in this area to understand the biological significance of subtle lipid changes among mycobacterial species and isolates.

Experimental procedures

Culture of S-type Map

S397 is an S-type strain of *Map* that has been previously characterized by whole genome sequencing (Bannantine *et al*., 2012). It was initially isolated from a Suffolk breed of

sheep in Iowa in 2004. Both strains S397 and K-10 were cultured in Middlebrook 7H9 media (BD Biosciences, San Jose, CA) supplemented with 10% OADC, 0.05% tween 80 and 2 μg ml^{-1} Mycobactin J. The culture conditions were 37°C with no shaking in 2 L Erlenmeyer flasks each containing 500 ml volumes of media. Milligram quantities were obtained from multiple cultures for downstream analyses.

Sequencing and assembly of mps1

A combination of sequencing and assembly strategies were required to fully assemble the *mps1* gene from *Map* S397. The large size of this gene and the presence of long repeats resulted in incomplete *mps1* assembly regardless of the assembler employed (MIRA v. 3.9.9, Roche gsAssembler v. 2.6 and Velvet v. 0.7.09). Targeted *de novo* subassemblies of the *mps1* region were created by first extracting reads that mapped to the region via MIRA's mirabait functionality using the partial contigs that aligned with MAP_1420 from K-10 and the MAP4_2425 homolog (Bannantine *et al*., 2014) as targets, and then *de novo* assembling those reads with MIRA. This was done in an iterative fashion and was supplemented as needed with additional targeted subcloning, PCR and Sanger sequencing of the *mps1* gene region until full unambiguous assembly was obtained. The GenBank accession number for *mps1* in *Map* S397 is KP720596.

Southern hybridization analysis

Mycobacteria were grown to late log phase in Middlebrook 7H9 medium (10 ml) and harvested by centrifugation at 6,000 × *g* for 10 min. The bacteria were heat killed for 10 min at 95°C. The pellet was resuspended in 10 ml of TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) and centrifuged again at 6,000 × *g* for 10 min. The semidried mycobacterial pellet was resuspended into 1 ml TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). After the addition of 200 μl of lysozyme (200 mg ml^{-1}) and incubation overnight at 37°C, 100 μl of SDS 10% and 50 μl Proteinase K (20 mg ml^{-1}) (Macherey-Nagel) were added and incubated 4 h at 56°C. 100 μl of 10% CTAB were mixed and incubated for 1 h at 65°C. One volume of phenol-chloroform-isoamyl alcohol (25:24:1 (vol/vol)) was added and the solution was vigorously mixed and then centrifuged at 14,000 × *g* for 5 min in phase lock gel (Qiagen). The supernatant was mixed with 1 volume of chloroform-isoamyl alcohol (24:1 (vol/vol)) and centrifuged again. The DNA was precipitated by the addition of 0.8 volume of isopropanol and 0.3 M sodium acetate (final concentration). After centrifugation for 30 min at 14,000 × g, the DNA was air dried, dissolved in 50 μl of Tris buffer (5 mM Tris-HCl [pH 8.5]) and stored at −20°C until further use.

Southern blot of *Map* DNA was performed as previously described (Southern, 1975; van Soolingen *et al*., 1994) with some modifications. The *mps1* DNA probe was prepared by PCR amplification of a 491 bp fragment sequence specific for *Map* using the primers described in this study (Table S3). PCRs were performed starting from 10 ng of chromosomal DNA of *Map* strain K-10 by using a TC-512 thermal cycler (Techne). The PCR product was purified on Macherey-Nagel spin columns according to the manufacturer's instructions. The probe was biotin labeled with the NEBlot Phototope kit (New England Biolabs) by following the instructions of the manufacturer. Digestion was performed with 3 μg of DNA prepared as described above and 7 U of *Sac*I (Promega) at 37°C for at least 6 h. Fragments were resolved by agarose gel electrophoresis and transferred onto Immobilon-S nylon membranes (Millipore) by vacuum transfer with the VacuGene system (Pharmacia LKB Biotechnology). Detection of DNA fragments hybridizing with the biotinylated probe was performed with the Phototope-Star detection kit for nucleic acids (New England Biolabs), according to the manufacturer's instructions. The 2-Log DNA Ladder (New England Biolabs) was used as a molecular size marker.

Reaction conditions for LSP^{mps1} amplification

A panel of *Map* isolates described in Table S2 was tested for the presence or absence of the large sequences identified within the genes *mps1* of K-10 compared to S397. This was done with a multiplex PCR approach (Semret *et al*., 2006) using a set of three primers: two primers (forward and reverse) designed toward the flanking regions (bridging primers) of the LSP and a third primer designed to recognize a sequence internal to the LSP (internal primer). The primers were designed such that the resulting PCR products would be of different sizes depending on the presence or absence of the LSP under study. Primer sequences are provided in Table S3. The PCR mixture comprised 2 μl of DNA solution added to a final volume of 25 μl containing 0.1 μl of GoTaq Flexi DNA polymerase (5 U μl^{-1} Promega), 2 mM (each) dATP, dCTP, dGTP and dTTP (Promega); 5 μl of 5× PCR buffer supplied by the manufacturer; 1 μM of each primers; 1 μl of dimethyl sulfoxide (Sigma); 1.5 mM of $MgCl_2$ and 5 μl of 5M betaine solution (Sigma). The reactions were carried out using a TC-512 thermal cycler (Techne). PCR conditions were as follows: 1 cycle of 5 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 62°C and 30 s at 72°C; and 1 cycle of 7 min at 72°C. To detect presence or absence of each LSP, PCR products were analyzed by electrophoresis using 1.5% agarose gels.

Bioinformatic prediction of peptide composition from NRPS sequence

The peptide composition of the lipopeptides analyzed in this study were deduced from DNA sequence comparisons between K-10 and S397 strains as well as a bioinformatics approach using domain prediction software including the NCBI web tools http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi and the website of PKS/NRPS Analysis at http://nrps.igs.umaryland.edu/nrps/. Peptide composition was determined using the web-based server NRPSpredictor2 (Rottig *et al*., 2011).

Chemical synthesis of the lipopeptides

The control lipopeptides (L3P and L5P) were synthesized on solid phase using the standard Fmoc chemistry protocol,

as previously described (Biet *et al.*, 2008). After cleavage from the resin, the crude L3P product was purified on a silica gel column using CH_2Cl_2/methanol as eluent (from 98:2 to 97:3 (vol/vol)), and 80 mg of the lipopeptide were obtained (yield 80% based on the net peptide content).

Characterization of the synthetic L3P

The synthetic L3P was characterized by electrospray ionization MS (Q-Tof Micro Waters), quantitative amino acid analysis (AAA) (after hydrolysis with 6N HCl at 110°C for 48 h and using a Beckman 6300 analyzer) and NMR (Bruker 400 MHz instrument). The results detailed below demonstrate both the identity and the purity of the compound.

MS: $C_{39}H_{67}N_3O_5$ (calcd 657.5081) *m/z* 658.5155 $[M + H]^+$, 680.4994 $[M + Na]^+$.

AAA: Ala 1 (1), Phe 0.96 (1), *N*-Methyl-Val 0.82 (1).

^{1}H NMR ($CDCl_3$): δ 0.60 (d, 3H, $CH_3\gamma$ Val, $J = 6.68$ Hz), 0.90 (t, 3H, CH_3 lipid, $J = 7.05$ Hz), 0.96 (d, 3H, $CH_3\gamma$ Val, $J = 6.41$ Hz), 1.25–1.29 (m, 32H, 16 CH_2 lipid), 1.35 (d, 3H, $CH_3\beta$ Ala, $J = 7.21$ Hz), 1.53–1.59 (m, 2H, CH_2CH_2CO lipid), 2.15 (t, 3H, CH_2CO lipid, $J = 7.60$ Hz), 2.18–2.27 (m, 1H, CHβ Val), 2.93 (s, 3H, NCH_3), 2.97 (dd, 1H, $1CH_2\beta$ Phe, $J_{1CH2\beta,CH\alpha} = 8.16$ Hz), 3.08 (dd, 1H, $1CH_2\beta$ Phe, $J_{1CH2\beta,CH\alpha} = 8.04$ Hz $J_{1CH2\beta,1CH2\beta} = 13.36$ Hz), 3.74 (s, 3H, OCH_3), 4.45 (d, 1H, CHα Val, $J = 11.04$ Hz), 4.5 (p, 1H, CHα Ala, $J_{CH\alpha,NH} = 7.2$ Hz), 5.17–5.24 (dt, 1H, CHα Phe, $J_{CH\alpha,NH} = 6.09$ Hz), 6.13 (bd, 1H, NH Phe), 6.59 (bd, 1H, NH Ala), 7.18–7.30 (5H, Ph).

^{13}C NMR ($CDCl_3$): δ 14.08 (CH_3 lipid), 17.86 ($CH_3\beta$ Ala), 18.64, 19.65 ($CH_3\gamma$ Val), 22.67 (CH_2 lipid), 25.53 (CH_2CH_2CO lipid), 25.84 (CHβ Val), 29.21, 29.34, 29.45, 29.64, 29.68 (CH_2 lipid), 30.96 (NCH_3), 31.91 (CH_2 lipid), 36.44 (CH_2CO lipid), 38.97 ($CH_2\beta$ Phe), 47.89 (CHα Ala), 50.38 (CHα Phe), 52.28 (OCH_3), 63.12 (CHα Val), 127.11, 128.57, 129.33, 135.80 (Ph), 169.05 (CO Val), 172.55 (CO lipid), 172.94 (CO Ala), 173.41 (CO Phe).

Lipid extraction, 2-D TLC and 1-D TLC

The culture of the S-type strain of *Map* afforded 317 mg of cells (dry weight). Lipids were extracted with chloroform/methanol (1:2 then 2:1 (vol/vol)) resulting in 7.6 mg of product. For analytical purposes, 500 μg of this crude extract were loaded on 2-D TLC plates and eluted using chloroform/methanol (96:4 (vol/vol)) in the first dimension followed by toluene/acetone (80:20 (vol/vol)) in the second dimension. Control synthetic L3P was deposited at 15 μg in chloroform and served as a marker for each dimension. TLC plates were revealed by spraying 10% copper sulfate in 8% phosphoric acid, followed by charring.

For the L3P purification, 7 mg of the crude extract in 100 μl of CH_2Cl_2 were loaded on preparative silica gel 60 F_{254} 2-D TLC (20 × 20 cm, thickness 0.5 mm) (Merck) and eluted using the same solvent systems as above. After scraping the spot of interest (~7 mm diameter) off the silica plate, the L3P was eluted in batch with 4 times 500 μl of CH_2Cl_2/methanol 95:5 (vol/vol). The evaporation under argon afforded approximately 50 μg of purified native antigen. The adjacent silica gel zone below (~6 mm diameter spot) was treated using the same procedure for the NMR control.

This purified native L3P was analyzed by silica gel 60 F_{254} 1-D TLC in comparison to both synthetic controls L3P and L5P (approximately 2 μg of each). The TLC was eluted with CH_2Cl_2/methanol 95:5 (vol/vol) and revealed as described above.

Surface-exposed material preparation

The surface-exposed material was recovered from mycobacteria treated with 10 g of glass beads as previously described (Ortalo-Magne *et al.*, 1996). Chloroform and methanol were added to the filtrates derived from this treatment to obtain a partition mixture composed of chloroform/methanol/water (3:4:3 (vol/vol/vol)), then the organic phases were washed with water and evaporated to dryness to yield the cell surface-exposed lipids. The treated bacteria were extracted as described above to yield the cell-bound lipids. Presence of cord factor was monitored by TLC developed in chloroform/methanol (90:10 (vol/vol)) and revelation by spraying 0.2% anthrone in sulfuric acid, followed by charring.

PBMC isolation and stimulation for flow cytometry and cytokine measurements

PBMCs were isolated from control non-infected ($n = 4$) and cattle naturally infected with *Map* ($n = 4$) to determine if lipoproteins, L3P and L5P, can elicit immunological responses. Sixty milliliter of blood was collected via jugular venipuncture into a syringe containing 2× acid-citrate-dextrose to obtain PBMCs. PBMCs were resuspended at a final concentration of 8 × 106 ml^{-1} in complete medium consisting of RPMI-1640 with 2 m*M* l-glutamine and 25 m*M* HEPES (Gibco, Grand Island, NY) and supplemented with 10% fetal calf serum (Gibco), 100X penicillin-streptomycin (Gibco). Cells were plated in 24-well culture plates and incubated for 24 h at 39°C in 5% CO_2 in a humidified atmosphere with the following treatment groups, nonstimulated (NS; negative control), pokeweed mitogen (PWM, 10 μg ml^{-1}, positive control; Sigma, St. Louis, MO) and four antigens that included whole cell sonicated extracts of *Map* strains K-10 and S397 (10 μg ml^{-1}); lipoproteins L3P and L5P (1, 5, 10 μg ml^{-1} concentrations). The lipoproteins had to be solubilized in 100% methanol to 1 mg ml^{-1} concentrations and then diluted in the complete medium to final concentrations indicated above. This diluted solvent-lipopeptide mixture did not affect cell viability or response capabilities. After a 24 h stimulation, the supernatants were harvested by centrifugation at 400 × *g* for 5 min. Supernatants were removed without disturbing the cells in culture and stored at −20°C until cytokine measurement. Cytokines IFN-γ, IL-1, IL-2, IL-4, IL-6 and TNF-α were all measured using Ciraplex bovine multiplex cytokine arrays (Aushon Biosystems, Billerica, MA).

For flow cytometry, PBMCs were cultured in replicate 48-well flat-bottom plates (Nunc Technologies, Rochester, NY) as described above with the same culture conditions and in vitro treatments. After incubating for either 3 days (NS,

PWM) or 6 days (NS, antigens), cell populations were defined by labeling with 50 μl of a cocktail of primary antibodies to CD4, CD8 gamma delta T cell receptor (γδ TCR) and B cells, along with a CD25 activation marker (Washington State University Monoclonal Antibody Center, Pullman, WA). After a 15 min incubation at room temperature (RT), plates were centrifuged for 2 min at 400 × *g*, the supernatant was decanted, and 50 μl of a secondary antibody cocktail was added, which included APC/Cy7 anti-mouse IgG2a (Southern Biotech, Birmingham, AL), AF350 anti-mouse IgG2b (Invitrogen, Waltham, MA) and BUV395 anti-mouse IgG3 (BDBiosciences, San Diego, CA). Live/Dead populations were separated using Zombie Yellow™ Fixable Viability Dye (Biolegend, San Diego, CA). Cells were analyzed on a BDBiosciences LSRII Cytometer using FACSDiva V8.0.1 software. Further analysis was done using FlowJo® v10.2 (FLOWJO, LLC) software.

Analytical procedures

MALDI-TOF/TOF-MS and MS/MS analyses were conducted in the positive ionization and reflectron mode by accumulating 10 spectra of 250 laser shots, using the 5800 MALDI TOF/TOF Analyser (Applied Biosystems/Absciex) equipped with a Nd:Yag laser (349 nm wavelength). For MS and MS/MS data acquisitions, uniform, continuous and random stage motion was selected at a fixed laser intensity of 4,000 (instrument-specific units) and 400 Hz pulse rate and 6,000 (instrument-specific units) and 1,000 Hz respectively. For MS/MS data acquisition, the fragmentation of selected precursors ions was performed at a collision energy of 1 kV using air as collision gas. Lipid samples were dissolved in chloroform and were directly spotted onto the target plate as 0.5 μl droplets, followed by the addition of 0.5 μl of matrix solution (10 mg of 2,5-dihydroxybenzoic acid (Sigma-Aldrich).ml $^{-1}$ in $CHCl_3/CH_3OH$, 1:1 (vol/vol)). Samples were allowed to crystallize at room temperature. Spectra were externally calibrated using lipid standards.

For comparative NMR analyses, 1-D ^1H and ^1H-COSY $^1H/^1H$ (COrrelation SpectroscopY), compounds were dissolved in $CDCl_3/CD_3OD$ (1:1 (vol/vol), 99.8% purity, Eurisotop, CEA Saclay, France). Experiments were conducted using a 600 MHz Bruker NMR spectrometer equipped with cryosonde. ^1H chemical shifts are given in parts/million (p.p.m.) downfield from internal tetramethylsilane at 0 p.p.m. All experiments were recorded at 295°K without sample spinning. The Bruker pulse programs were used and optimized (pulse lengths and delays) for each 1-D or 2-D experiments. Data were analyzed using the TopSpin (Bruker BioSpin) software.

Acknowledgements

We thank Janis Hansen (USDA-ARS) for expert technical assistance. F.B., M.B. and T.C. were supported by the Institut National de la Recherche Agronomique (INRA) and the Cluster de recherche en infectiologie de la région Centre. NMR experiments were performed on the PICT-Genotoul platform of Toulouse and funded by CNRS, Université de Toulouse-UPS, Ibisa, European structural funds and the Midi-Pyrénées region. This study was financially supported by the EMIDA-EraNet MYCOBACTDIAGNOSIS project and the USDA-Agricultural Research Service.

Author contributions

F.Bi. and J.P.B. conceived the study, carried out experiments, interpreted the data and wrote the manuscript; S.B. and G.E. contributed to study design, carried out experiments, contributed to writing the manuscript and interpreted the data; T.C. and M.B. performed and assisted in interpreting the molecular biological experiments and bioinformatics analyses.; C.G. and F.Bo. performed the preparation and analysis of the synthetic antigen, as well as the purification of the native antigen; F.L. and A.L. performed and analyzed the MALDI-TOF and NMR experiments respectively. J.R.S. and J.P.B. designed and conducted the immune response experiments. D.O.B. analyzed sequencing data, performed sequence assemblies and provided bioinformatic support. All the authors discussed the results and commented on the manuscript.

References

Alexander, D.C., Turenne, C.Y., and Behr, M.A. (2009) Insertion and deletion events that define the pathogen *Mycobacterium avium* subsp. *paratuberculosis*. *J Bacteriol* **191**: 1018–1025.

Alonso-Hearn, M., Eckstein, T.M., Sommer, S., and Bermudez, L.E. (2010) A *Mycobacterium avium* subsp. *paratuberculosis* LuxR regulates cell envelope and virulence. *Innate Immun* **16**: 235–247.

Amonsin, A., Li, L.L., Zhang, Q., Bannantine, J.P., Motiwala, A.S., Sreevatsan, S., and Kapur, V. (2004) Multilocus short sequence repeat sequencing approach for differentiating among *Mycobacterium avium* subsp. *paratuberculosis* strains. *J Clin Microbiol* **42**: 1694–1702.

Bannantine, J.P., Wu, C.W., Hsu, C., Zhou, S., Schwartz, D.C., Bayles, D.O., *et al*. (2012) Genome sequencing of ovine isolates of *Mycobacterium avium* subspecies *paratuberculosis* offers insights into host association. *BMC Genomics* **13**: 89.

Bannantine, J.P., Li, L., Mwangi, M., Cote, R., Raygoza Garay, J.A., and Kapur, V. (2014) Complete genome sequence of *Mycobacterium avium* subsp. *paratuberculosis*, isolated from human breast milk. *Genome Announc* **2**: e01252–13.

Bernut, A., Herrmann, J.L., Kissa, K., Dubremetz, J.F., Gaillard, J.L., Lutfalla, G., and Kremer, L. (2014) *Mycobacterium abscessus* cording prevents phagocytosis and promotes abscess formation. *Proc Natl Acad Sci USA* **111**: E943–E952.

Biet, F., and Boschiroli, M.L. (2014) Non-tuberculous mycobacterial infections of veterinary relevance. *Res Vet Sci* **97**: S69–S77.

Biet, F., Bay, S., Thibault, V.C., Euphrasie, D., Grayon, M., Ganneau, C., *et al*. (2008) Lipopentapeptide induces a strong host humoral response and distinguishes *Mycobacterium avium* subsp. *paratuberculosis* from *M. avium* subsp. *avium*. *Vaccine* **26**: 257–268.

Biet, F., Sevilla, I.A., Cochard, T., Lefrancois, L.H., Garrido, J.M., Heron, I., *et al*. (2012) Inter- and intra-subtype genotypic differences that differentiate *Mycobacterium avium* subspecies *paratuberculosis* strains. *BMC Microbiol* **12**: 264.

Chedid, L., Parant, M., Parant, F., Gustafson, R.H., and Berger, F.M. (1972) Biological study of a nontoxic, water-soluble immunoadjuvant from mycobacterial cell walls. *Proc Natl Acad Sci USA* **69**: 855–858.

Dohmann, K., Strommenger, B., Stevenson, K., de Juan, L., Stratmann, J., Kapur, V., *et al*. (2003) Characterization of genetic differences between *Mycobacterium avium* subsp. *paratuberculosis* type I and type II isolates. *J Clin Microbiol* **41**: 5215–5223.

Eckstein, T.M., Chandrasekaran, S., Mahapatra, S., McNeil, M.R., Chatterjee, D., Rithner, C.D., *et al*. (2006) A major cell wall lipopeptide of *Mycobacterium avium* subspecies *paratuberculosis*. *J Biol Chem* **281**: 5209–5215.

Etienne, G., Malaga, W., Laval, F., Lemassu, A., Guilhot, C., and Daffe, M. (2009) Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. *J Bacteriol* **191**: 2613–2621.

Gaudelli, N.M., Long, D.H., and Townsend, C.A. (2015) beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520**: 383–387.

Ghosh, P., Hsu, C., Alyamani, E.J., Shehata, M.M., Al-Dubaib, M.A., Al-Naeem, A., *et al*. (2012) Genome-wide analysis of the emerging infection with *Mycobacterium avium* subspecies *paratuberculosis* in the Arabian camels (*Camelus dromedarius*). *PLoS One* **7**: e31947.

Gumber, S., and Whittington, R.J. (2009) Analysis of the growth pattern, survival and proteome of *Mycobacteriumavium* subsp. *paratuberculosis* following exposure to heat. *Vet Microbiol* **136**: 82–90.

Gumber, S., Taylor, D.L., Marsh, I.B., and Whittington, R.J. (2009) Growth pattern and partial proteome of *Mycobacterium avium* subsp. *paratuberculosis* during the stress response to hypoxia and nutrient starvation. *Vet Microbiol* **133**: 344–357.

Holbert, S., Branger, M., Souriau, A., Lamoureux, B., Ganneau, C., Richard, G., *et al*. (2015) Interferon gamma response to *Mycobacterium avium* subsp. *paratuberculosis* specific lipopentapeptide antigen L5P in cattle. *Res Vet Sci* **102**: 118–121.

Howard, S.T., Rhoades, E., Recht, J., Pang, X., Alsup, A., Kolter, R., *et al*. (2006) Spontaneous reversion of *Mycobacterium abscessus* from a smooth to a rough morphotype is associated with reduced expression of glycopeptidolipid and reacquisition of an invasive phenotype. *Microbiology* **152**: 1581–1590.

Janagama, H.K., Senthilkumar, T.M., Bannantine, J.P., Rodriguez, G.M., Smith, I., Paustian, M.L., *et al*. (2009) Identification and functional characterization of the iron-dependent regulator (IdeR) of *Mycobacterium avium* subsp. *paratuberculosis*. *Microbiology* **155**: 3683–3690.

Jimenez-Dalmaroni, M.J., Radcliffe, C.M., Harvey, D.J., Wormald, M.R., Verdino, P., Ainge, G.D., *et al*. (2015) Soluble human TLR2 ectodomain binds diacylglycerol from microbial lipopeptides and glycolipids. *Innate Immun* **21**: 175–193.

Kansal, R.G., Gomez-Flores, R., and Mehta, R.T. (1998) Change in colony morphology influences the virulence as well as the biochemical properties of the *Mycobacterium avium* complex. *Microb Pathog* **25**: 203–214.

Kieser, K.J., and Rubin, E.J. (2014) How sisters grow apart: mycobacterial growth and division. *Nat Rev Microbiol* **12**: 550–562.

Lederer, E., Adam, A., Ciorbaru, R., Petit, J.F., and Wietzerbin, J. (1975) Cell walls of Mycobacteria and related organisms; chemistry and immunostimulant properties. *Mol Cell Biochem* **7**: 87–104.

Lefrancois, L.H., Bodier, C.C., Cochard, T., Canepa, S., Raze, D., Lanotte, P., *et al*. (2013) Novel feature of *Mycobacterium avium* subsp. *paratuberculosis*, highlighted by characterization of the heparin-binding hemagglutinin adhesin. *J Bacteriol* **195**: 4844–4853.

Li, L., Bannantine, J.P., Zhang, Q., Amonsin, A., May, B.J., Alt, D., *et al*. (2005) The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci USA* **102**: 12344–12349.

Marrakchi, H., Laneelle, M.A., and Daffe, M. (2014) Mycolic acids: structures, biosynthesis, and beyond. *Chem Biol* **21**: 67–85.

Marsh, I., Whittington, R., and Cousins, D. (1999) PCR-restriction endonuclease analysis for identification and strain typing of *Mycobacterium avium* subsp. *paratuberculosis* and *Mycobacterium avium* subsp. *avium* based on polymorphisms in IS1311. *Mol Cell Probes* **13**: 115–126.

Marsh, I.B., Bannantine, J.P., Paustian, M.L., Tizard, M.L., Kapur, V., and Whittington, R.J. (2006) Genomic comparison of *Mycobacterium avium* subsp. *paratuberculosis* sheep and cattle strains by microarray hybridization. *J Bacteriol* **188**: 2290–2293.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3**: 490–504.

Motiwala, A.S., Janagama, H.K., Paustian, M.L., Zhu, X., Bannantine, J.P., Kapur, V., and Sreevatsan, S. (2006) Comparative transcriptional analysis of human macrophages exposed to animal and human isolates of *Mycobacterium avium* subspecies *paratuberculosis* with diverse genotypes. *Infect Immun* **74**: 6046–6056.

NAHMS (2008) *Johne's Disease on U.S. Dairies, 1991–2007*. USDA-APHIS-VS-CEAH Fort Collins, CO: Center for Epidemiology and Animal Health, pp. 1–4.

Ortalo-Magne, A., Lemassu, A., Laneelle, M.A., Bardou, F., Silve, G., Gounon, P., *et al*. (1996) Identification of the surface-exposed lipids on the cell envelopes of *Mycobacterium tuberculosis* and other mycobacterial species. *J Bacteriol* **178**: 456–461.

Park, I.K., Hsu, A.P., Tettelin, H., Shallom, S.J., Drake, S.K., Ding, L., *et al*. (2015) Clonal Diversification and Changes in lipid traits and colony morphology in *Mycobacterium abscessus* clinical isolates. *J Clin Microbiol* **53**: 3438–3447.

Paustian, M.L., Zhu, X., Sreevatsan, S., Robbe-Austerman, S., Kapur, V., and Bannantine, J.P. (2008) Comparative genomic analysis of *Mycobacterium avium* subspecies obtained from multiple host species. *BMC Genomics* **9**: 135.

Ripoll, F., Deshayes, C., Pasek, S., Laval, F., Beretti, J.L., Biet, F., *et al*. (2007) Genomics of glycopeptidolipid biosynthesis in *Mycobacterium abscessus* and *M. chelonae*. *BMC Genomics* **8**: 114.

Riviere, M., Puzo, G., Wright, E.L., and Barrow, W.W. (1996) A unique phenylalanine-containing lipopeptide isolated from a rough-colony variant of *Mycobacterium avium*. *Eur J Biochem* **241**: 682–690.

Rottig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2–a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* **39**: W362–W367.

Semret, M., Turenne, C.Y., de Haas, P., Collins, D.M., and Behr, M.A. (2006) Differentiating host-associated variants of *Mycobacterium avium* by PCR for detection of large sequence polymorphisms. *J Clin Microbiol* **44**: 881–887.

Sibley, J.A., Woodbury, M.R., Appleyard, G.D., and Elkin, B. (2007) *Mycobacterium avium* subspecies *paratuberculosis* in Bison (Bison bison) from Northern Canada. *J Wildl Dis* **43**: 775–779.

van Soolingen, D., de Haas, P.E., Hermans, P.W., and van Embden, J.D. (1994) DNA fingerprinting of *Mycobacterium tuberculosis*. *Methods Enzymol* **235**: 196–205.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503–517.

Stevenson, K. (2015) Genetic diversity of *Mycobacterium avium* subspecies *paratuberculosis* and the influence of strain type on infection and pathogenesis: a review. *Vet Res* **46**: 64.

Stevenson, K., Hughes, V.M., de Juan, L., Inglis, N.F., Wright, F., and Sharp, J.M. (2002) Molecular characterization of pigmented and nonpigmented isolates of *Mycobacterium avium* subsp. *paratuberculosis*. *J Clin Microbiol* **40**: 1798–1804.

Taylor, A.W. (1951) Varieties of *Mycobacterium johnei* isolated from sheep. *J Pathol Bacteriol* **63**: 333–336.

Thibault, V.C., Grayon, M., Boschiroli, M.L., Willery, E., Allix-Beguec, C., Stevenson, K., *et al*. (2008) Combined multilocus short-sequence-repeat and mycobacterial interspersed repetitive unit-variable-number tandem-repeat typing of *Mycobacterium avium* subsp. *paratuberculosis* isolates. *J Clin Microbiol* **46**: 4091–4094.

Turenne, C.Y., Semret, M., Cousins, D.V., Collins, D.M., and Behr, M.A. (2006) Sequencing of hsp65 distinguishes among subsets of the *Mycobacterium avium* complex. *J Clin Microbiol* **44**: 433–440.

Turenne, C.Y., Collins, D.M., Alexander, D.C., and Behr, M.A. (2008) *Mycobacterium* avium subsp. *paratuberculosis* and *M. avium* subsp. *avium* are independently evolved pathogenic clones of a much broader group of M. avium organisms. *J Bacteriol* **190**: 2479–2487.

Verdier, J., Deroche, L., Allez, M., Loy, C., Biet, F., Bodier, C.C., *et al*. (2013) Specific IgG response against *Mycobacterium avium paratuberculosis* in children and adults with Crohn's disease. *PLoS One* **8**: e62780.

Walsh, C.T., O'Brien, R.V., and Khosla, C. (2013) Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew Chem Int Ed In Engl* **52**: 7098–7124.

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci USA* **111**: 9259–9264.

Whitlock, R.H., and Buergelt, C. (1996) Preclinical and clinical manifestations of *paratuberculosis* (including pathology). *Vet Clin North Am Food Anim Pract* **12**: 345–356.

Whittington, R.J., Marsh, I.B., Saunders, V., Grant, I.R., Juste, R., Sevilla, I.A., *et al*. (2011) Culture phenotypes of genomically and geographically diverse *Mycobacterium avium* subsp. *paratuberculosis* isolates from different hosts. *J Clin Microbiol* **49**: 1822–1830.

Wright, E.L., Zywno-van Ginkel, S., Rastogi, N., and Barrow, W.W. (1996) Monoclonal infection involving *Mycobacterium avium* presenting with three distinct colony morphotypes. *J Clin Microbiol* **34**: 2475–2478.

Wu, C.W., Glasner, J., Collins, M., Naser, S., and Talaat, A.M. (2006) Whole-genome plasticity among *Mycobacterium avium* subspecies: insights from comparative genomic hybridizations. *J Bacteriol* **188**: 711–723.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.