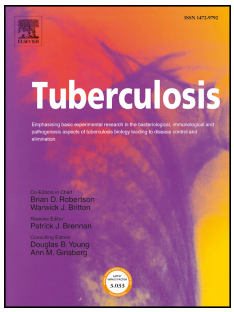
Accepted Manuscript



Ancient mycobacterial lipids: key reference biomarkers in charting the evolution of tuberculosis

Professor David E. Minnikin, Oona Y-C. Lee, Houdini H.T. Wu, Gurdyal S. Besra, Apoorva Bhatt, Vijayashankar Nataraj, Bruce M. Rothschild, Mark Spigelman, Helen D. Donoghue

PII: S1472-9792(15)00010-4

DOI: 10.1016/j.tube.2015.02.009

Reference: YTUBE 1270

To appear in: *Tuberculosis*

Please cite this article as: Minnikin DE, Lee OY-C, Wu HHT, Besra GS, Bhatt A, Nataraj V, Rothschild BM, Spigelman M, Donoghue HD, Ancient mycobacterial lipids: key reference biomarkers in charting the evolution of tuberculosis, *Tuberculosis* (2015), doi: 10.1016/j.tube.2015.02.009.

Ancient mycobacterial lipids: key reference biomarkers in charting the evolution of tuberculosis

David E. Minnikin[a*], Oona Y-C. Lee[a], Houdini H.T. Wu[a], Gurdyal S. Besra[a], Apoorva Bhatt[a], Vijayashankar Nataraj[a], Bruce M. Rothschild[b], Mark Spigelman[c] and Helen D. Donoghue[d]

[a]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK

[b]Biodiversity Institute and Departments of Anthropology and Geology , University of Kansas, Lawrence, KS 66045, USA

[c]Kuvin Center for the Study of Infectious and Tropical Diseases and Ancient DNA, Hadassah Medical School, Hebrew University, Jerusalem, Israel

[d]Centres for Clinical Microbiology and the History of Medicine, University College London, London, UK

Email addresses

d.e.minnikin@bham.ac.uk; o.y.lee@bham.ac.uk; h.wu.2@bham.ac.uk; g.besra@bham.ac.uk; a.bhatt@bham.ac.uk; VXN131@bham.ac.uk; bmr@ku.edu; spigelman@btinternet.com; h.donoghue@ucl.ac.uk

***Corresponding author**

Professor David E. Minnikin, School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Email: d.e.minnikin@bham.ac.uk Telephone: +44(0)1214158126

SUMMARY

Mycobacterium tuberculosis has a cell envelope incorporating a peptidoglycan-linked arabinogalactan esterified by long-chain mycolic acids. A range of "free" lipids are associated with the "bound" mycolic acids, producing an effective envelope outer membrane. The distribution of these lipids is discontinuous among mycobacteria and such lipids have proven potential for biomarker use in tracing the evolution of tuberculosis. A plausible evolutionary scenario involves progression from an environmental organism, such as *Mycobacterium kansasii,* through intermediate "smooth" tubercle bacilli, labelled *"Mycobacterium canettii"*; cell envelope lipid composition possibly correlates with such a progression. *M. kansasii* and *"M. canettii"* have characteristic lipooligosaccharides, associated with motility and biofilms, and glycosyl phenolphthiocerol dimycocerosates ("phenolic glycolipids"). Both these lipid classes are absent in modern *M. tuberculosis sensu stricto,* though simplified phenolic glycolipids remain in certain current biotypes. Dimycocerosates of the phthiocerol family are restricted to smaller phthiodiolone diesters in *M. kansasii*. Diacyl and pentaacyl trehaloses are present in *"M. canettii"* and *M. tuberculosis,* where they are accompanied by related sulfated acyl trehaloses. In comparison with environmental mycobacteria, subtle modifications in mycolic acid structures in *"M. canettii"* and *M. tuberculosis* are notable. The probability of essential tuberculosis evolution taking place in Pleistocene megafauna, rather than *Homo sapiens,* is reemphasised.

Keywords: Tuberculosis; Evolution; Lipids; Biomarkers; Zoonosis

1. Introduction

Tuberculosis is an ancient disease, whose pre-Holocene history is shrouded in mystery. Analysis of skeletal material has provided evidence for tuberculosis in *Homo sapiens* at up to 9,000 years (9 ka) before present (BP),[1] almost back to the start of the Holocene. Travelling back from the Holocene into the cyclical glacial times of the Pleistocene, human skeletal material becomes scarce and no direct evidence for any tuberculosis in *H. sapiens* has been demonstrated. In contrast, distinctive tuberculosis lesions have been recorded in a range of megafauna and other animals from that epoch. Typically, the lesions take the form of undermined articular surfaces, as exemplified by a metacarpal from *Bison antiquus* recovered from Natural Trap Cave, Wyoming.[2] In addition to the bison metacarpal, 19% of 1,002 125 ka to 8 ka BP bovid specimens[3] and 52% of 113 38 ka to 10 ka BP mastodon bones[4] had similar lesions indicative of tuberculosis. Bone lesions cannot be considered as complete proof of tuberculosis diagnosis, but the dearth of comparable lesions in bones from *H. sapiens*, over the same time period, is very striking. To resolve this conundrum it has been proposed[2] that *Mycobacterium tuberculosis* may have been principally an animal disease during its early evolution, with transmission to humans occurring later.

The use of amplified DNA sequences to diagnose tuberculosis in archaeological material has been developed during the past two decades.[5] Major advances in determining full genomic data have been recently provided by the application of so-called "Next Generation Sequencing"[6] and the more direct "Metagenomic" approach.[7] Informative genomic data have been obtained for specimens stretching back to 9 ka in *H. sapiens*[1] and 17 ka in extinct *Bison antiquus*[2] and these diagnoses have been supported by the use of robust lipid biomarkers.[1,2,5] The most diagnostic lipids have been mycolic, mycocerosic and mycolipenic acids and members of the phthiocerol family.[1,2,5] These, and a range of other lipids, are vital components in the integrity of the cell envelopes of the tubercle bacillus and related taxa.[8] Their distribution, however, is discontinuous and changes in lipid composition and structure may well be important factors in the evolution of effective pathogenic species. The aim of this paper is to outline a rational scenario for the evolution of the current *M. tuberculosis* complex from possible environmental candidates. The speculative focus will be on correlating changes in cell envelope lipid composition with developing pathogenicity, taking into account the suitability of particular animal

hosts along the way. Representative structures of the key lipids under consideration are shown in Figures 1 and 2.

2. An environmental opportunist to a perfect pathogen?

The challenge is to chart a pathway from ancestral environmental freely-circulating mycobacterial species to *M. tuberculosis sensu stricto*, an obligate pathogen with no environmental niche. Currently favoured hypotheses all point to an evolutionary bottle-neck, initiated around 35 ka BP.[9,10] Subsequent to this time period, the evolution of a range of particular clades follows an almost linear clonar evolutionary pattern, with key deletions leading to the well-defined modern *M. tuberculosis* complex (MTBC) causing tuberculosis in humans and various animals.[11-13]

There is increasing evidence that, before reaching the discontinuity of the bottle-neck, extensive horizontal gene transfer (HGT) was taking place in ancestral tuberculosis strains.[10,14] These strains may not necessarily have been obligate pathogens but opportunist mycobacteria with the ability to survive in the hostile environment of an animal stomach. The rich flora of multiple animal stomachs would provide plentiful opportunities for HGTs, eventually resulting in interim organisms with an enhanced potential to cause tuberculosis. Prime candidates for such a role are pre-bottle-neck ancestral strains, sometimes termed *"M. prototuberculosis"*,[9] which are associated with the "smooth" colony-forming "Canetti" variants of *M. tuberculosis*.[15,16] "*Mycobacterium canettii*" smooth strains continue to be encountered in isolated cases of tuberculosis, but they are usually confined to certain locations in the Horn of Africa.[15]

A case for *M. marinum* as the pivotal environmental source organism has been advanced,[17] but several key factors mitigate against such a selection. The stereochemistries of the *M. marinum* PDIM waxes and PGLs are completely different from those produced by *M. tuberculosis* and *M. kansasii*.[8,18] In addition, the oxygenated mycolic acids of *M. marinum* are not cyclopropanated, in contrast with those from *M. tuberculosis* and *M. kansasii*.[18,19] The environmental organism that phenotypically resembles *M. tuberculosis* most closely is *Mycobacterium kansasii* and this relationship has been supported by genomic comparisons.[20,21] Cogent arguments have been advanced to associate the evolution of tubercle bacilli with bacteria similar

to *M. kansasii*, including indications of HGTs between these taxa.[20,21] Key genes acquired by HGT include those coding for mycobacterial lipids, transferases and proteins related to adaptation to anaerobic conditions.[20,21] *M. kansasii* still causes pulmonary disease in Silesian and South African miners, the bacterium being contracted from water in showers.[21] In developing a coherent evolutionary route, the pathway from *M. kansasii*, through *"M. canettii"*, to *M. tuberculosis* is a good working hypothesis. Changes in lipid composition are potentially very significant, involving the mycolic acids, phthiocerol dimycocerosates (PDIMs), glycosyl phenolphthiocerol dimycocerosates ("phenolic glycolipids", PGLs) (Figure 1), diacyl trehaloses (DATs), pentaacyl trehaloses (PATs) and sulfated acyl trehalose glycolipids (SGLs) (Figure 2).[8]

3. Does lipid evolution parallel *M. tuberculosis* evolution?

3.1 *Mycolic acids*

Possibly the most deep-lying fundamental differences between the lipids from *M. kansasii* and members of the *M. tuberculosis* complex (MTBC), loosely including *"M. canettii"*, are subtle changes in mycolic acid structure. Mycolates from MTBC have characteristic 24-carbon chains in 2-position, whereas the mycolates from *M. kansasii* and the majority of mycobacteria have principally 22-carbon side chains.[8,19] In addition, the MTBC α-mycolates show a significant shortening of the size of the chain between carbon-3 and the proximal cyclopropane (17 → 13 carbons) and the lengthening of the terminal chain (18 → 20 carbons) beyond the distal cyclopropane unit, as compared with *M. kansasii* (Figure 1A).[19] The methoxymycolates and ketomycolates of *"M. canettii"* and *M. tuberculosis* (Figure 1A) conform to the general pattern of these components in related mycobacteria, such as *M. kansasii*, but, significantly, these oxygenated mycolates are slightly larger than any others.[19]

The balance of the three main types of mycolates is possibly significant; the ratios of the α-, methoxy- and ketomycolates are, respectively, ~10:5:8 for *M. kansasii*, ~10:6:8 for "*M. canettii"* and ~10:5:5 for *M. tuberculosis*.[19] Having half of the proportions as α-mycolates in *M. tuberculosis* may be quite significant. The major all *cis*-cyclopropyl α-mycolates of all three taxa are similar in size, centred around 80 carbons. Those from *M. kansasii* are restricted to 80 and 82 carbons overall, but there

are four detailed structural varieties of each giving a heterogeneous mixture of eight distinct α-mycolates.[19] In contrast, the four α-mycolates from "*M. canettii"* and *M. tuberculosis* are all very uniform, the two major C_{78} and C_{80} components being accompanied by minor C_{82} and C_{84} mycolates.[19] It is particularly notable that the central (14-carbon) and distal (20-carbon) meromycolate chains are invariable in the α-mycolates from "*M. canettii"* and *M. tuberculosis* (Figure 1A).[19] The complex methoxymycolates from *M. kansasii,* totalling eight *cis*- and four *trans*-components, have a *cis:trans* ratio of ~3:2,whereas both "*M. canettii"* and *M. tuberculosis* have an enhanced *cis:trans* ratio of ~3:1.[19] Somewhat simplified methoxymycolates, with six *cis*- and two *trans*-components are found in "*M. canettii"*, simplifying further to mainly a C_{85} and lesser C_{87} *cis*- and a single *trans*-methoxymycolate in *M. tuberculosis*.[19] The *trans*-ketomycolates predominate over the *cis*-forms in *M. kansasii* (~6:1), "*M. canettii"* (~4:1) and *M. tuberculosis* (~3:2, respectively); the latter two have mainly a C_{87} *trans*-ketomycolate accompanied by six very minor variants but this contrasts with a heterogeneous mix of ten in *M. kansasii*.[19]

The essence of the above seemingly complex changes is an apparent simplification and tightening up of mycolate composition. Mycolic acids are "cornerstones" of the mycobacterial outer membrane, providing a covalent hydrophobic inner leaflet, facilitating binding of the range of "free lipids" (Figures 1 and 2) that comprise the outer leaflet.[8] Physical studies indicate that ketomycolates appear to have a prime structural role in adopting tightly folded conformations to produce a solid foundation.[22] It is not surprising, therefore, that there is minimal variation in the general structure of ketomycolates between, for example, *M. kansasii* and MTBC (Figure 1A). In contrast, the structurally different α-mycolates from *M. kansasii* and MTBC (Figure 1A) behave quite distinctly in monolayer studies.[22] As noted above α-mycolates constitute half of the overall mycolates in *M. tuberculosis,* reinforcing the possibility of important modifications in cell envelope interactions with a special range of free lipids (Figures 1 and 2).

3.2 *Dimycocerosates of the phthiocerol family and glycosyl phenolphthiocerols*

The phthiocerol dimycocerosate (PDIM) waxes are important tuberculosis virulence factors and the *"M. canettii"* and *M. tuberculosis* examples are the largest (Figure 1B).[8,23] The *M. tuberculosis* complex phthiocerol family has major C_{34}/C_{36}

phthiocerol A components, minor C_{33}/C_{35} phthiocerol Bs and C_{33}/C_{35} phthiodiolones (Figure 1B).[2,8,23] In contrast, *M. kansasii* PDIM waxes were much smaller, having only C_{25}/C_{27} phthiodiolones (Figure 1B) and no methoxylated phthiocerol As and Bs.[23] Interestingly, the mainly $C_{29}/C_{30}/C_{32}$ mycocerosic acid composition of *M. kansasii* PDIM waxes is comparable with that of the *M. tuberculosis* complex.[23]

The glycosyl phenolphthiocerol dimycocerosates, the so-called "phenolic glycolipids" (PGLs) (Figure 1C), are related to the PDIMs (Figure 1B).[8,23] The PGLs from *"M. canettii"* comprise a 2-*O*-methyl rhamnosyl PGL and an extended triglycosyl PGL; interestingly, the main PGL produced by *M. kansasii* is extended further by an additional sugar (Figure 1C).[8,23] This close structural similarity in PGLs has been highlighted in schemes suggesting an evolutionary progression from *M. kansasii* to *"M. canettii"*.[21] The PGLs from *M. kansasii* do include a methoxylated phenolphthiocerol component (Figure 1C), in contrast to the situation for the PDIM waxes (Figure 1B).[23] The 2-*O*-methyl rhamnosyl PGL is also characteristic of modern ecotypes, such as *M. bovis, M. africanum* and some so-called "Beijing" lineages of *M. tuberculosis*[24] (Figure 1C). However, PGLs are not present in a large clade of modern TB lineages due to a decisive pks 15/1 gene frameshift.[25]

3.3 *Glycolipids based on trehalose*

In addition to PGLs, *M. kansasii* and *"M. canettii"* are characterised by the production of a range of highly polar antigenic lipooligosaccharides (LOSs) (Figure 2A, B).[26] The main *M. kansasii* LOS (Figure 2A) has an acylated thirteen sugar oligosaccharide, based on trehalose, but that from *"M. canettii"* is refined down to an unrelated acyl trehalose nonasaccharide (Figure 2B). It has been clearly demonstrated, for *M. marinum* and *M. kansasii*, that these relatively hydrophilic LOSs promote biofilm formation and motility.[27] These characteristics may have a useful survival role for free-living organisms and HGTs may be facilitated by such behaviour. However, once an opportunist *Mycobacterium* had evolved into an obligate parasite, the production of LOSs may no longer have conveyed a competitive advantage. Indeed, it has been shown that smooth variants of *M. kansasii*, containing LOSs, are rapidly cleared from the organs of infected animals, but rough variants, lacking all LOSs, produce chronic systemic infections.[28] It is possible to surmise that the loss of LOSs is a key event in the transition from a free-living opportunist *Mycobacterium* to a

transmissible obligate pathogen. This may be as significant a change as the pks 15/1 gene frameshift that resulted in the loss of PGLs.[25]

The diacyl trehaloses (DATs) and pentaacyl trehaloses (PATs) are two glycolipid classes also based on a trehalose scaffold (Figure 2C).[8,18] First characterised from *M. tuberculosis* H37Rv, DATs and PATs are present in representative modern clinical strains, as well as in "*M. canettii*",[8,18] but they have not been characterised from *M. kansasii.* The multimethyl-branched fatty acid components of DATs are C_{24} 2,4-dimethyl docosanoic ("mycosanoic") (Figure 2Ca) and C_{27} 3-hydroxy 2,4,6-trimethyl tetracosanoic ("mycolipanolic") (Figure 2Ca').[8] The related characteristic main fatty acid in PATs is C_{27} 2,4,6-trimethyl tetracos-2-enoic ("mycolipenic") acid (Figure 2Cb).[8,18] It is very significant that the absolute stereochemistry of the methyl-branched centres in all these acids (Figure 2C) is ***S*** in contrast to the ***R*** configuration in the mycocerosates (Figure 1C). The related sulfated acyl trehalose glycolipids (SGLs) also feature multimethyl-branched fatty acid components of the ***S*** series, examples being the C_{37} phthioceranic acids and C_{40} hydroxyphthioceranic (Figure 2Cc,d).[8] Sulfoglycolipids (SGLs) are restricted to modern *M. tuberculosis* and they have not been found in *"M. canettii"*.[29]

3.4 *Overall summary of lipid correlations*

In summary, it is clear that in charting a hypothetical progression from *M. kansasii*, via *"M. canettii"*, to *M. tuberculosis* there are identifiable changes in cell envelope lipid composition. The challenge is to pinpoint significant modifications that may have contributed to the undoubted success of the biotypes of modern tubercle bacilli. Scrutiny of the complex profiles of the α-, methoxy- and ketomycolates indicates a tightening up both in structural details and distribution of types, but the most significant change is undoubtedly the presence of a longer 2-alkyl chain and very specific alterations in the proximal and α-mycolate distal chain lengths (Figure 1A) for *"M. canettii"* and *M. tuberculosis*. It is enticing to speculate that the distinct α-mycolate (Figure 1A) may be influential in the outer membrane, enhancing links with the particular portfolios of free lipids (Figures 1 and 2) found in *"M. canettii"* and *M. tuberculosis*. The exceptionally long members of the whole phthiocerol family (Figure 1B), from the PDIMs of *"M. canettii"* and *M. tuberculosis*, also distinguish these taxa; in contrast, *M. kansasii* only has much smaller phthiodiolones (Figure 1B).

The apparent structural change in PGLs from *M. kansasii* to *"M. canettii"* has been advanced previously as evidence of a close evolutionary linking of these taxa.[21] The principal *M. kansasii* PGL is truncated by both one and three sugars to produce comparable proportions of triglycosyl and monoglycosyl PGLs in *"M. canettii"*. Modern post-bottleneck MTBC biotypes fall into two distinct categories, with respect to PGL production; monoglycosyl PGLs are retained in *M. africanum*, *M. bovis* and the *M. tuberculosis* "Beijing" and related clades, but a major group of modern *M. tuberculosis* lineages have lost the ability to produce PGLs.[25] *M. kansasii* and *"M. canettii"* both produce a range of highly polar LOSs,[26] associated with aquatic environments, motility and biofilms; modern MTBC organisms lack these lipids. LOSs are based on acylated trehaloses (Figure 2A, B) and several other classes of trehalose-based glycolipids are encountered (Figure 2C). The polar antigenic DATs and apolar PATs are characteristic of *"M. canettii"* and *M. tuberculosis sensu strictu*, but the closely-related, relatively polar, SGLs are limited to the latter category. A most significant feature of DATs, PATs and SGLs is the presence of multimethyl-branched fatty acids of the ***S*** series (Figure 2C), rather than the ***R*** methyl branches in the mycocerosates from the PDIMs and PGLs (Figure 1B, C). It would be interesting to determine if the genomic origin of these longer ***S*** series multimethyl-branched fatty acids, from DATs, PATs and SGLs (Figure 2C), correlates with the shorter ***S*** series fatty acids found in LOSs (Figure 2A). In view of the fact that the multimethyl-branched fatty acid components of the PDIMs and PGLs from *M. marinum* are also of the ***S*** series,[8,18,23] the possibility of HGT from this source should be considered.

A hypothesis is being advanced, here, for an outline model evolutionary pathway for modern tubercle bacilli from an environmental organism, such as *M. kansasii*, via the very diverse group of "smooth" isolates provisionally gathered together under the label *"M. canettii"*. Perceived significant lipid changes, in the taxa under consideration, are summarised in Figure 3, but such stylised changes would necessarily be much less compartmentalised than indicated. The tremendous diversity of extant smooth *"M. canettii"* strains[16] indicates that a complex labyrinth of pathways may have been followed to evolve these taxa. Indeed, extant *"M. canettii"* are not necessarily good representatives of the first mycobacteria that developed a preference for mammalian antibiosis/symbiosis rather than free-living; however they are the best existing signposts. As noted earlier, bovid metacarpal lesions point to the presence of tuberculosis back to 125 ka BP,[3] so productive mammalianisation and

adaptation of an environmental candidate could have taken place over a period of up to 100 ka at least. Modern tuberculosis was distilled out of such a melange, but again there may be no single definitive pathway. However, tuberculosis evolution was brought into sharper focus during passage through the clear bottleneck,[9,10,13] which preceded the relatively rapid evolution of all the modern biotypes.[9-13] The dearth of human skeletal material makes it difficult to delineate the true role of *H. sapiens* in accelerating tuberculosis evolution. Suffice it to say, however, that when settled human communities were established, modern tuberculosis found a convenient niche from which to expand and diversify.

4. Where on earth did tubercle bacilli evolve and who were the host vectors?

To return to the hypothesis of a possible key zoonotic origin of the *M. tuberculosis* complex, an environmental mycobacterial common ancestor might well have prospered in primeval waters. Aqueous suspensions of these bacteria could then have been repeatedly passaged through a range of prehistoric animals, such as bovids and mastodons.[3,4] The possible involvement of protozoa as a direct vector into humans has been suggested,[14] but such microorganisms could also be incidental surrogates facilitating transmission into animals and not specific direct vectors. It has been indicated[20,21] that tubercle bacilli have probably adopted genes favouring survival in anaerobic conditions, thereby encouraging growth in the reduced oxygen environment of animal cells. It has been argued that Pleistocene bovids and mastodons may have lived in almost symbiosis with ancestral tubercle bacilli for many aeons,[3,4] perhaps allowing the slow accumulation of variants that caused the widespread lesions indicative of tuberculosis. However, the cyclical changes during the Ice Ages may have eventually produced an unfavourable environment in which the presence of an infecting agent may have contributed significantly to the demise of these characteristic animal species.

The proposed scheme is not in general accord with a recent detailed examination of the parallel evolution of genomes from 186 members of the *M. tuberculosis* complex and 4,995 human mitochondria.[30] Extrapolation of the results were interpreted to suggest that TB and humans co-evolved “out of Africa”, commencing ~70 ka ago.[30] However, parallel evolution is not necessarily linked co-

evolution and no definitive evidence was advanced to substantiate the presence of any human tuberculosis going back as far as 70 ka BP.

The scenario, outlined in this communication, is just one of many zoonotic possibilities, but the general outcome could be an obligate animal pathogen causing a disease that is now recognised as tuberculosis. On the balance of evidence, a major ancient reservoir of this global disease would appear to be a range of prehistoric large animals, spread throughout the Northern Hemisphere.[3,4] The transmission to humans could well have been through ingestion of infected animal material, analogous to the way that modern bovine tuberculosis can be contracted by drinking raw milk or eating undercooked infected meat. The eventual gathering together of human communities could have facilitated a transition to the inter-person spread of modern human tuberculosis.

Funding: The Leverhulme Trust Project Grant F/00 094/BL (OY-CL, DEM, GSB)

Acknowledgements

GSB acknowledges support in the form of a Personal Research Chair from Mr. James Bardrick and the UK Medical Research Council.

References

1. Hershkovitz I, Donoghue HD, Minnikin DE, Besra GS, Lee OY-C, Gernaey AM, Galili E, Eshed V, Greenblatt CL, Lemma E, Kahila Bar-Gal G, Spigelman M. Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS ONE* 2008;**3**:e3426.
2. Lee OY-C, Wu HHT, Donoghue HD, Spigelman M, Greenblatt CL, Bull ID, Rothschild BM, Martin LD, Minnikin DE, Besra GS. *Mycobacterium tuberculosis* complex lipid virulence factors preserved in the 17,000-year-old skeleton of an extinct bison, *Bison antiquus*. *PLoS ONE* 2012;**7**:e41923.
3. Rothschild BM, Martin LD. Did ice-age bovids spread tuberculosis? *Naturwissenschaften* 2006;**93**:565-569.
4. Rothschild BM, Laub R. Hyperdisease in the late Pleistocene: validation of an early 20th century hypothesis. *Naturwissenschaften* 2006;**93**:557–564.
5. Minnikin DE, Lee OY-C, Wu HHT, Besra GS, Donoghue HD. Molecular biomarkers for ancient tuberculosis. In: Cardona P-J, editor. *Understanding Tuberculosis – Deciphering the Secret Life of the Bacilli.* Rijeka, Croatia: InTech - Open Access Publisher. 2012. pp. 1-36. http://www.intechopen.com/books/understanding-tuberculosis-deciphering-the-secret-life-of-the-bacilli ISBN-13: 978-953-307-946-2.
6. Bouwman AS, Kennedy SL, Muller R, Stephens RH, Holst M, Caffell AC, Roberts CA, Brown TA. Genotype of a historic strain of *Mycobacterium tuberculosis. Proc Natl Acad Sci USA* 2012;**109**:18511-18516.

7. Chan JZ-M, Sergeant MJ, Lee OY-C, Minnikin DE, Besra GS, Pap I, Spigelman M, Donoghue HD, Pallen MJ. Metagenomic analysis of tuberculosis in a mummy. *N Engl J Med* 2013;**369**:289-290.

8. Minnikin DE, Kremer L, Dover LG, Besra GS. The methyl-branched fortifications of *Mycobacterium tuberculosis. Chem Biol* 2002;**9**: 545-553.

9. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, Supply P, Vincent V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis. PLoS Pathog* 2005;**1**:e5.

10. Djelouadji Z, Raoult D, Drancourt M. Palaeogenomics of *Mycobacterium tuberculosis*: epidemic bursts with a degrading genome. *Lancet Infect Dis* 2011;**11**:641–650.

11. Wirth T Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T. Kremer K, van Soolingen D, Rüsch-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog* 2008;**4**:e1000160.

12. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Jared C, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 2008;**6**:e311.

13. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis. Nat Rev Microbiol* 2009;**7**:537-544.

14. Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O. Horizontally acquired genomic islands in the tubercle bacilli. *Trends Microbiol* 2008;**16**:303-308.

15. Koeck J-L, Fabre M, Simon F, Daffé M, Garnotel E, Matan AB, Gérôme P, Bernatas J-J, Buisson Y, Pourcel C. Clinical characteristics of the smooth tubercle bacilli '*Mycobacterium canettii'* infection suggest the existence of an environmental reservoir. *Clin Microbiol Infect* 2011;**17**:1013–1019.

16. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Médigue C, Parkhill J, Cruveiller S, Brosch R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis. Nat Genet* 2013;**45**:172-179.

17. Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, Johnson PDR, Abdellah Z, Arrowsmith C, Chillingworth T, Churcher C, Clarke K, Cronin A, Davis P, Goodhead I, Holroyd N, Jagels K, Lord A, Moule S, Mungall K, Norbertczak H, Quail MA, Rabbinowitsch E, Walker D, White B, Whitehead S, Small PLC, Brosch R, Ramakrishnan L, Fischbach MA, Parkhill J, Cole ST. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis. Genome Research* 2008;**18**:729–741.

18. Daffé M, Lacave C, Lanéelle M-A, Gillois M, Lanéelle G. Polyphthienoyl trehalose, glycolipids specific for virulent strains of the tubercle bacillus. *Eur J Biochem* 1988;**172**:579-584.

19. Watanabe M, Aoyagi Y, Mitome H, Fujita T, Naoki H, Ridell M, Minnikin DE. Location of functional groups in mycobacterial meromycolate chains; the

recognition of new structural principles in mycolic acids. *Microbiology* 2002;**148**:1881-1902.

20. Veyrier F, Pletzer D, Turenne C, Behr MA. Phylogenetic detection of horizontal gene transfer during the step-wise genesis of *Mycobacterium tuberculosis. BMC Evol Biol* 2009;**9**:196.
21. Veyrier FJ, Dufort A, Behr MA. The rise and fall of the *Mycobacterium tuberculosis* genome. *Trends Microbiol* 2011;**19**:156-161.
22. Villeneuve M, Kawai M, Watanabe M, Aoyagi Y, Hitotsuyanagi Y, Takeya K, Gouda H, Hirono S, Minnikin DE, Nakahara H. Differential conformational behaviors of α-mycolic acids in Langmuir monolayers and computer simulations. *Chem Phys Lipids* 2010;**163**:569-579.
23. Onwueme KC, Vos CJ, Zurita J, Ferreras JA, Quadri LEN. The dimycocerosate ester polyketide virulence factors of mycobacteria. *Prog Lipid Res* 2005;**44**:259–302.
24. Huet SG, Constant C, Malaga W, Lanéelle M-A, Kremer K, van Soolingen D, Daffé M, Guilhot C. A lipid profile typifies the Beijing strains of *Mycobacterium tuberculosis.* Identification of a mutation responsible for a modification of the structures of phthiocerol dimycocerosates and phenolic glycolipids. *J Biol Chem* 2009;**284**:27101–27113.
25. Constant P, Perez E, Malaga W, Lanéelle M-A, Saurel O, Daffé M, Guilhot C. Role of the *pks15/1* gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. *J Biol Chem* 2002;**277**:38148–38158.

26. Gilleron M, Puzo G. Lipooligosaccharidic antigens from *Mycobacterium kansasii* and *Mycobacterium gastri*. *Glycoconjugate Journal* 1995;**12**:298-308.

27. Ren H, Dover LG, Islam ST, Alexander DC, Chen JM, Besra GS, Liu J. Identification of the lipooligosaccharide biosynthetic gene cluster from *Mycobacterium marinum*. *Mol Microbiol* 2007;**63**:1345-1359.

28. Belisle JT, Brennan PJ. Chemical basis of rough and smooth variation in mycobacteria. *J Bacteriol* 1989;**171**:3465-3470.

29. Soto CY, Cama M, Gibert I, Luquin M. Application of an easy and reliable method for sulfolipid-I detection in the study of its distribution in *Mycobacterium tuberculosis* strains. *FEMS Microbiol Lett* 2000;**187**:103-107.

30. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;**45**:1176-1182.

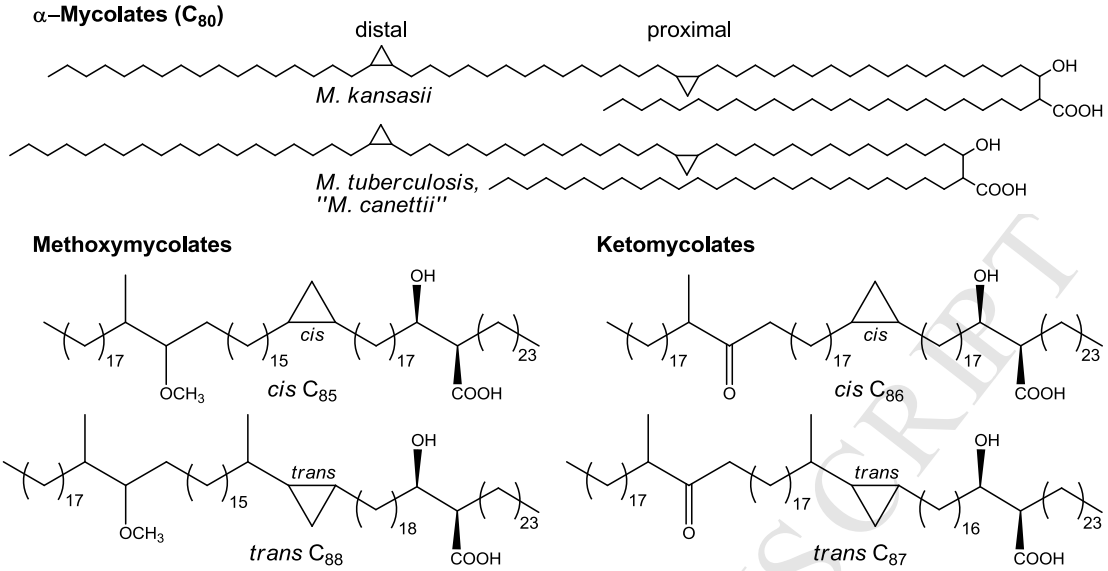
Figure legends

Figure 1. Representative structures of mycolic acids and phthiocerol and phenolphthiocerol-based lipids. (A) Mycolic acids, showing chain length differences between *M. kansasii* α-mycolates and those from *"M. canettii"* and *M. tuberculosis.* The main methoxy- and ketomycolates from *M. tuberculosis* are shown; those from *M. kansasii* are essentially similar. (B) Dimycocerosates of the main *M. kansasii* phthiodiolone and *M. tuberculosis* and *"M. canettii"* phthiocerol As. (C) Phenolic glycolipids (PGLs) from *M. kansasii, "M. canettii"* and *M. bovis*.

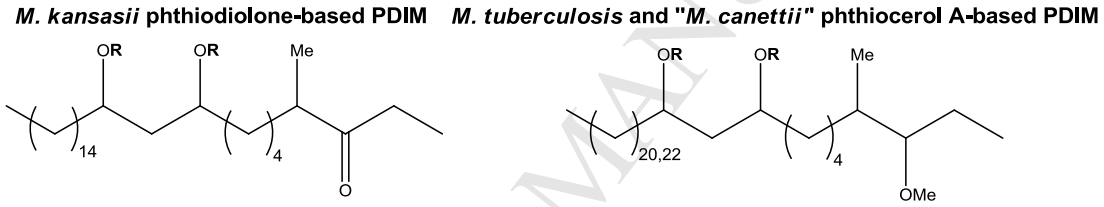
Figure 2. Representative structures of trehalose-based glycolipids. (A) The main lipooligosaccharide (LOS) from *M. kansasii*. (B) The main lipooligosaccharide (LOS) from *"M. canettii"*. (C) The main diacyl (DAT), pentaacyl (PAT) and sulfated acyl (SGL) trehalose glycolipids. The DAT with C_{24}-mycosanoate, a), is accompanied by a DAT with C_{27}-mycolipanolate, a').

Figure 3. Correlation of lipid composition of *M. kansasii, "M. canettii"* and *M. tuberculosis* in the context of hypothetical linked evolution. Cartoon structures aim to reflect the clear differences in the α-mycolates of *M. kansasii,* in comparison with those of *"M. canettii"* and *M. tuberculosis*; shown in detail in Figure 1A. The numbers of sugars in particular LOSs and PGLs are shown in brackets with an asterisk, e.g. LOS (13)*.

A. Mycolic acids



B. Dimycocerosates of phthiocerol family (PDIMs)



C. Glycosyl phenolphthiocerol dimycocerosates ('phenolic' glycolipids, PGLs)

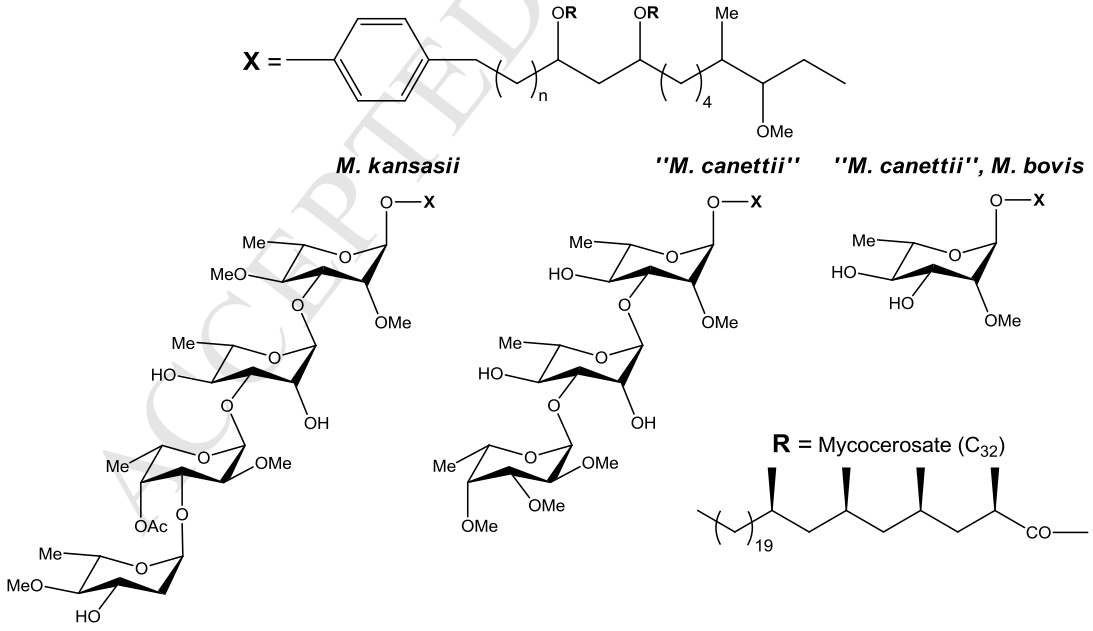
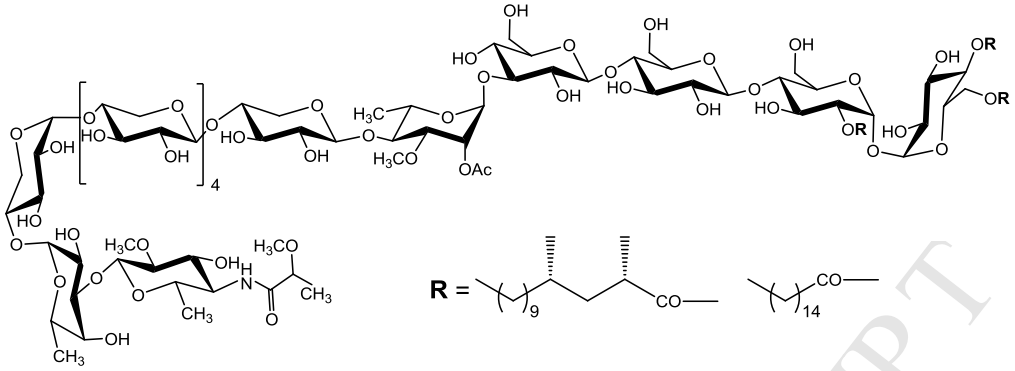
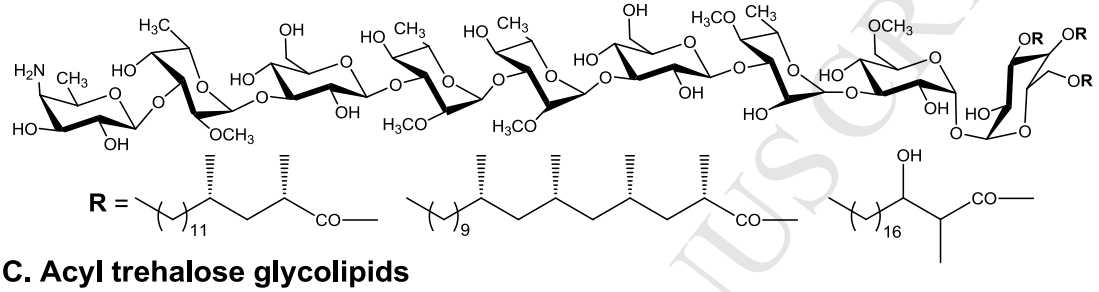


Figure 1

A. *M. kansasii* lipooligosaccharide (LOS)



B. "*M. canettii*" lipooligosaccharide (LOS)



C. Acyl trehalose glycolipids

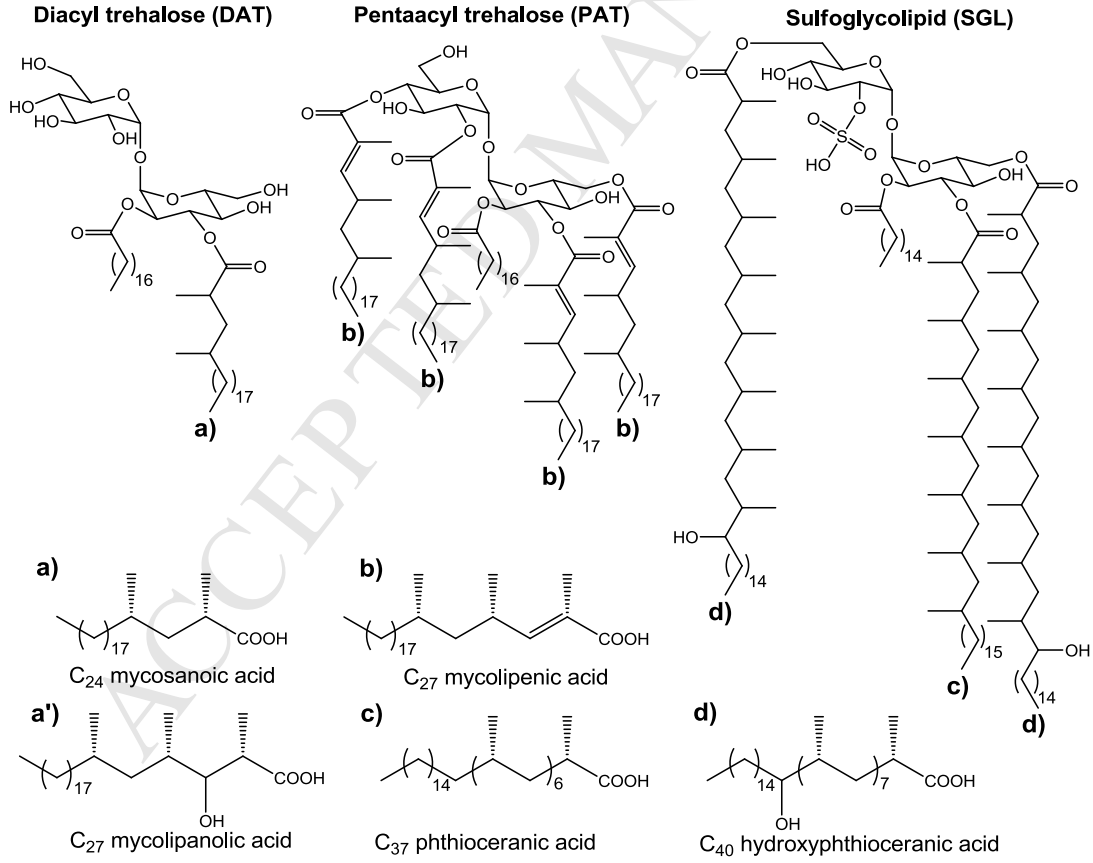


Figure 2

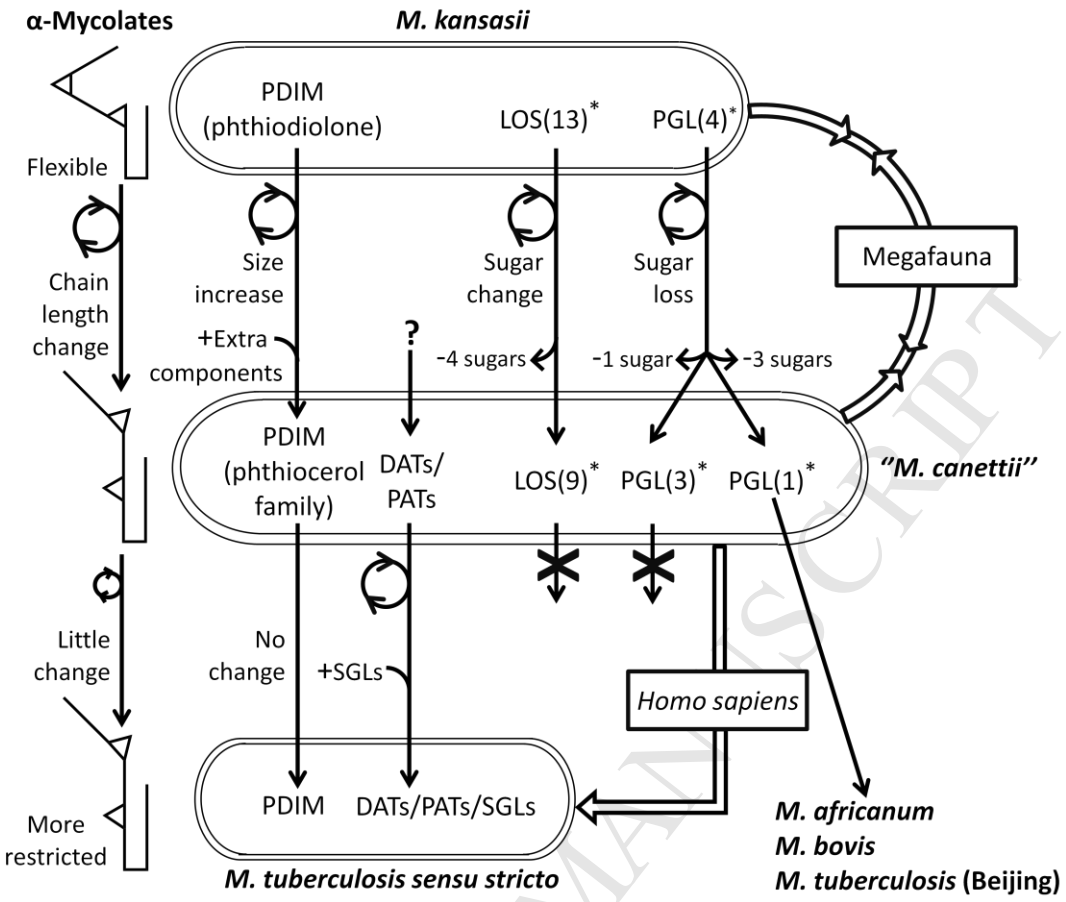


Figure 3