

Minireview

The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set

Varalakshmi D Vissa and Patrick J Brennan

Address: Department of Microbiology, Colorado State University, Fort Collins, CO 80523, USA.

Correspondence: Varalakshmi D Vissa. E-mail: Varalakshmi.Vissa@colostate.edu

Published: 3 August 2001

Genome Biology 2001, **2**(8):reviews1023.1–1023.8

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/8/reviews/1023>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

Comparison of the recently sequenced genome of the leprosy-causing pathogen *Mycobacterium leprae* with other mycobacterial genomes reveals a drastic gene reduction and decay in *M. leprae* affecting many metabolic areas, exemplified by the retention of a minimal set of genes required for cell-wall biosynthesis.

Mycobacterium leprae, 'Hansen's Bacillus', was the first human pathogenic bacterium to be identified, predating the discovery of the tubercle bacillus (*Mycobacterium tuberculosis*) by a decade. The genomes of both have now been decoded [1-4]. The genomes of other mycobacteria are also being sequenced, including those that cause opportunistic infections in people with AIDS (*Mycobacterium avium*) [5], bovine tuberculosis (*Mycobacterium bovis*) [6], and Johne's disease of cattle (*Mycobacterium avium* subsp. *paratuberculosis*) [7]. The sequencing of *Mycobacterium smegmatis* [5], the laboratory model strain used for studying mycobacterial physiology and genetics, and of the phylogenetically related *Corynebacterium glutamicum* [5] and *Corynebacterium diphtheriae* [8], are also under way. Although clinical aspects of the virulent mycobacterial strains vary, they are all intracellular pathogens that are transmitted by the respiratory route and occupy macrophages as their preferred niche [9]. A number of antibodies crossreact amongst these bacterial species, indicating similarities in protein composition, and the basic cell-wall architecture is the same [10]. Thus, comparative genomics is a useful tool for identifying common and divergent pathways.

Cole *et al.* [1,3] have found that, compared to the *M. tuberculosis* H37Rv genome of 4,411,529 base-pairs (bp), which can potentially encode 3,924 genes [3], the *M. leprae* genome of 3,268,203 bp encodes only 1,604 proteins and

contains 1,116 pseudogenes [1]. They have annotated and classified all these genes into various functional categories. Figure 1 depicts this drastic gene reduction and decay in *M. leprae* compared to *M. tuberculosis*, which affects nearly every aspect of metabolism.

Despite numerous experiments that demonstrated metabolic activity by labeling macromolecules such as phenolic glycolipid (PGL)-I, proteins, nucleic acids and lipids with radioactive precursors in bacteriological media or in macrophages infected with host-derived *M. leprae*, multiplication of *M. leprae* cells has not been achieved. The only sources of *M. leprae* are tissues from infected humans, armadillos or mouse footpads [11]. The failure to grow *M. leprae* cells *in vitro* may result from the combined effects of gene reduction and mutations in several metabolic areas (Figure 1b). Mutations are found in genes involved in regulation (encoding repressors, activators, two-component systems, serine-threonine kinases and phosphatases), detoxification (genes encoding peroxidases), DNA repair (the *mutT*, *dnaQ*, *alkA*, *dinX*, and *dinP* genes) and transport or efflux of metabolites such as amino acids (arginine, ornithine, D-alanine, D-serine and glycine), peptides, cations (magnesium, nickel, mercury, ammonium, ferrous and ferric ions and potassium), and anions (arsenate, sulfate and phosphate). In general, pseudogenes are found more frequently in degradative, rather than synthetic, pathways. Genes for the synthesis

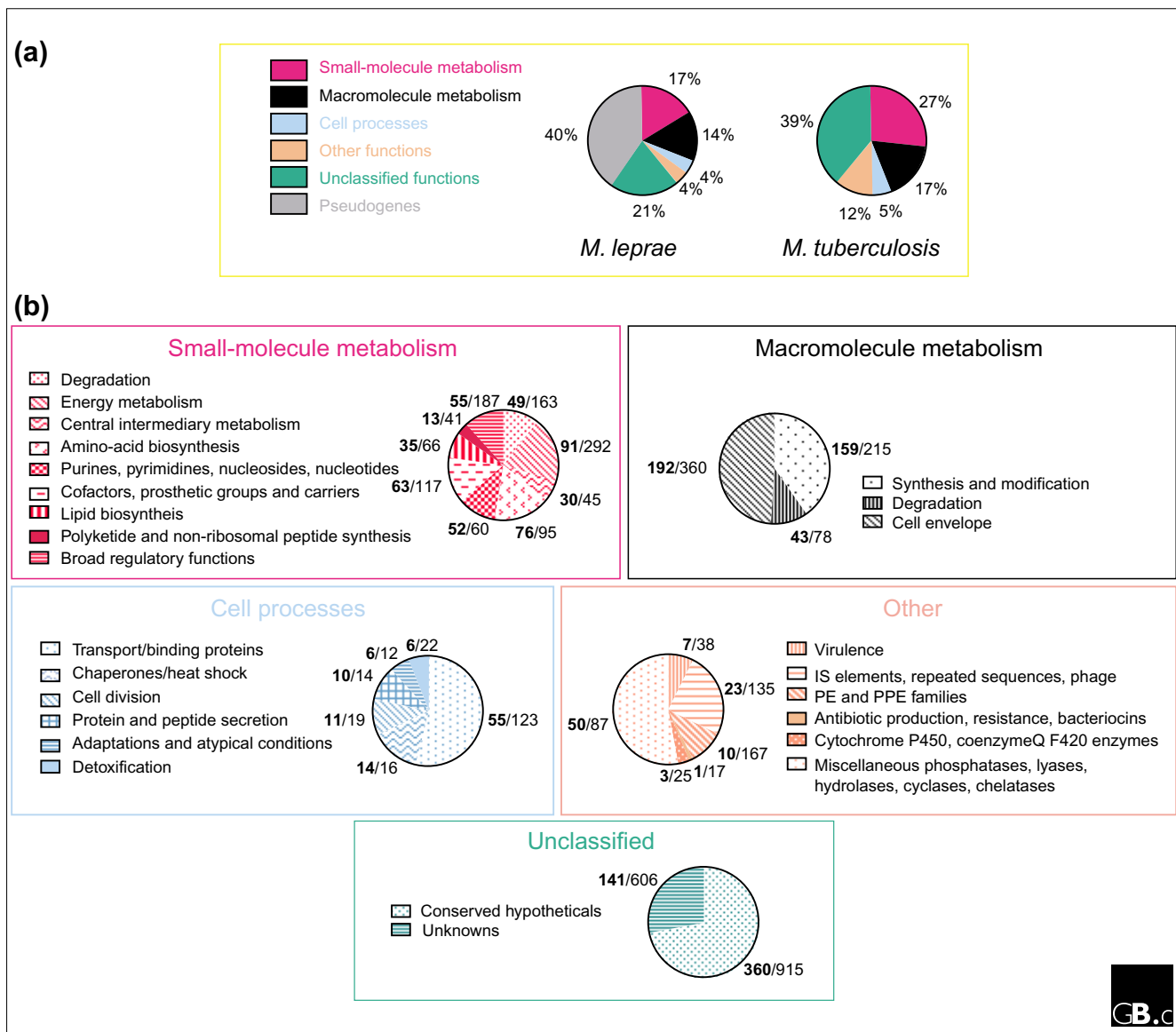


Figure 1
 The extent of gene reduction and decay in the genome of *M. leprae*. **(a)** The percentage of the total potential open reading frames assigned to major cellular functions are shown. **(b)** Each category has been sub-classified and the number of putative functional genes in *M. leprae* (after eliminating the pseudogenes) for each subclass are indicated by bold numbers, followed by the corresponding number in *M. tuberculosis*. The data were obtained from the databases of the *M. leprae* and *M. tuberculosis* genome projects [2,4] as annotated by Cole et al. [1,3].

of most small molecules, such as amino acids, purines, pyrimidines and fatty acids, and for the synthesis of macromolecules such as ribosomes, aminoacyl tRNAs, RNA and proteins, are reasonably intact.

In terms of gene reduction, there are fewer genes in almost every category, but notably affected are insertion sequences (IS) and the acidic, glycine-rich families of proteins that have proline-glutamic acid (PE) or proline-proline-glutamic acid (PPE) motifs at the amino terminus; these proteins may confer antigenic variation. Repressors, activators, oxidoreductases

and oxygenases are also affected. Thus, while preserving genes required for its transmission, establishment and survival in the host, *M. leprae* has discarded genes that can be compensated for by a host-dependent parasitic lifestyle. Analysis of the *M. leprae* genome therefore provides a useful paradigm for all mycobacteria, because of its smaller genome size, obligate intracellularism, and limited complement of genes. The availability of several completely or partially sequenced mycobacterial genomes allows us to dissect the genetics of conserved and dissimilar pathways, such as those for cell-wall biosynthesis.

Retention of the essence of mycobacterial cell walls in *M. leprae*

Extensive studies of the ultrastructure of the cell wall of *M. leprae*, both embedded in sections and as whole bacteria isolated from infected tissue in man, mouse, and armadillo, have shown properties common to all mycobacteria: beyond the plasma membrane is a rigid, moderately dense layer composed of an innermost electron-dense layer (probably consisting of peptidoglycan, PG, and arabinogalactan, AG), an intermediate electron-transparent zone (the mycolate layer), and an outermost electron-dense layer (probably composed of assorted lipoglycans, free polysaccharides, glycolipids, and phospholipids) [12,13] (see Figure 2).

The underlying framework or 'core' of all mycobacterial cell walls consists of PG, which is covalently attached through a linker unit (LU) (-Rha-GlcNAc-P-) to AG distinguished by furanose sugars (Gal_f and Ara_f) [10,14]; the abbreviations we use in the glycoconjugate and sugar names in this review are defined in Box 1. Attached to the terminal Ara_f units are the mycolic acids (mycolates - (Ara_f)₋₃₀-(Gal_f)₋₃₀-Rha-GlcNAc-P-PG), the lipophilicity of which provides the dominant physiological features of all mycobacteria [15]. Lipoarabinomannan (LAM), lipomannan (LM), the phosphatidylinositol-mannosides (PIMs), cord factor (trehalose dimycolate), sulfolipids, and proteins are associated with this framework in a physical arrangement that is poorly understood [10] (Figure 2).

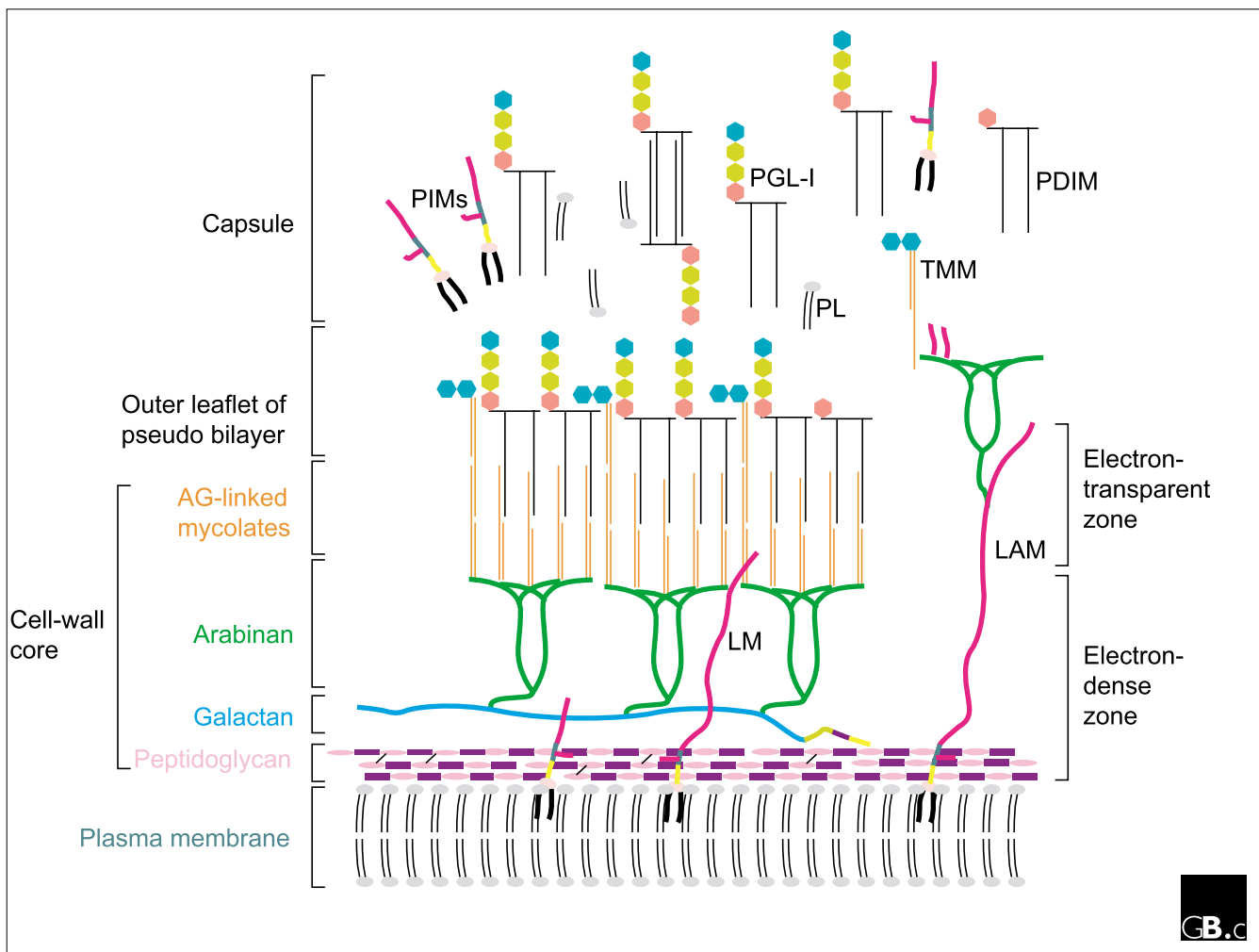


Figure 2

A schematic model of the cell envelope of *M. leprae*. The plasma membrane is covered by a cell-wall core made of peptidoglycan (chains of alternating GlcNAc and MurNGly, linked by peptide crossbridges) covalently linked to the galactan by a linker unit (-P-GlcNAc-Rha-) of arabinogalactan. Three branched chains of arabinan are in turn linked to the galactan. The peptidoglycan-arabinogalactan layer forms the electron-dense zone. Mycolic acids are linked to the termini of the arabinan chains to form the inner leaflet of a pseudo lipid bilayer. An outer leaflet is formed by the mycolic acids of TMM and mycoceroic acids of PDIMs and PGLs as indicated. The pseudo-bilayer forms the electron-transparent zone. A capsule presumably composed largely of PGLs and other molecules such as PDIMs, PIMs and phospholipids surrounds the bacterium. Lipoglycans such as PIMs, LM and LAM, known to be anchored in the plasma membrane, are also found in the capsular layer as shown. Abbreviations are as used in the text and Box 1.

Box 1**A list of abbreviations used in the glycoconjugate and sugar names**

ACP	Acyl carrier protein
AG	Arabinogalactan
Araf	Arabinofuranose
Galf	Galactofuranose
Glc	Glucose
GlcNAc	N-acetylglucosamine
LAM	Lipoarabinomannan
LM	Lipomannan
Me	Methyl
MurNGly	N-glycolylmuramic acid
P	Phosphate
PDIM	Phthiocerol dimycocerosate
PG	Peptidoglycan
PGL	Phenolic glycolipids
PIMs	Phosphatidylinositol mannosides
Rha	Rhamnose
TDP	Thymidine 5'-diphosphate
TMM	Trehalose monomycolate
UDP	Uridine 5'-diphosphate

The limited chemical analysis conducted on the *M. leprae* cell wall to date suggests that it conforms to this pattern, but with modifications [16]. Small amounts of trehalose monomycolate (TMM) are present, but there is no cord factor [17], and, apparently, *M. leprae* contains the full complement of PIMs but is devoid of the trehalose-based mycolipenic-acid-containing sulfolipids characteristic of virulent strains of *M. tuberculosis*. The application of freeze-etching techniques to *M. leprae* in phagolysosomes isolated from infected human, mouse, and armadillo cells showed large quantities of 'peribacillary substances', which appeared as 'spherical droplets', a feature unique to *M. leprae*-infected cells [18]. This material proved to be made up of the *M. leprae*-specific phenolic glycolipids (PGL-I, PGL-II and PGL-III) and the related phthiocerol dimycocerosate (PDIM) [19]. PGL-I consists of the basic phenol-PDIM with the *M. leprae*-specific trisaccharide (3,6-di-O-Me-Glc)-(2,3-di-O-Me-Rha)-(3-O-Me-Rha) in glycosidic link to the phenol component. Lepromatous leprosy is characterized by high titers of antibodies to the trisaccharide unit of PGL-I, and a synthetic derivative has proved useful for serodiagnosis of this condition [20]. Recently, the trisaccharide - notably the terminal 3,6-di-O-Me-Glc unit - was shown to be the *M. leprae*-specific ligand in the characteristic interaction of *M. leprae* and Schwann cells, the glial cells of the peripheral nervous system, which are invaded by *M. leprae in vivo* [21]. This discovery is important as it identified an *M. leprae* virulence factor that is involved in causing the characteristic nerve damage observed in some leprosy patients. The glycosyltransferases for the synthesis of PGL-I are therefore good candidate drug targets.

Comparative genomics of cell-envelope synthesis

Understanding of the biosynthesis of mycobacterial cell walls is still evolving, and our knowledge to date is confined to understanding individual components of the cell wall separately; the pathways and regulation of final assembly are not understood. The genetics of some of the pathways that have been elucidated in different mycobacterial species, such as *M. tuberculosis*, *M. smegmatis*, *M. avium* or *M. bovis*, have been compiled in reviews on the mycobacterial cell wall, and putative genes of *M. tuberculosis* have also been predicted on the basis of homology to genes in other bacteria [22,23]. In the complete, online version of this article, we update these analyses for various wall components - mycolic acids, polyprenyl phosphates, peptidoglycan, linker-unit arabinoglycan, mannans and PGL-I - by including and comparing the findings for the condensed genome of *M. leprae*. The genes for the synthesis of these molecules are largely preserved in *M. leprae*, with some variation relative to *M. tuberculosis*. Although *M. leprae* PG contains glycine rather than L-alanine in the peptide crosslinks, there is only one conserved ligase, MurC, for the addition of glycine or L-alanine to UDP-MurNGly rather than specialized ligases for the two amino acids. The bare minimum of genes for synthesizing polyprenyl phosphates are retained in *M. leprae*. The *pimB* gene, which in *M. tuberculosis* codes for a mannosyltransferase required for the synthesis of PIM₂, a precursor of LM and LAM, is a pseudogene in *M. leprae*. Similarly, *fabH*, encoding a β -ketoacyl-ACP synthase for mycolic-acid synthesis does not have a clear homolog in *M. leprae*, and the methoxymycolate synthase genes *mmaA2* and *mmaA3* are pseudogenes, which explains the lack of methoxymycolates in *M. leprae*.

As described in detail in the complete version of this article online, we know little about the glycosyltransferases involved in the synthesis of the mycobacterial cell wall such as mannosyltransferases for LM and LAM biosynthesis, rhamnosyl and glycosyltransferases for PGL-I and polyprenyl-P-glycosyltransferases (for C₅₀-P-Araf). By combining information from annotations in the genome databases of *M. tuberculosis* and *M. leprae* [2,4] with the results of BLAST and RPS-BLAST searches [24] and with what is known about some glycosyltransferases (such as *pimB* and *glfT*), we have compiled a list of glycosyltransferases from the genomes of *M. leprae* and *M. tuberculosis* and tentatively assigned certain functions to them (see Table 1, available with the full version of this article, online). Also included in the searches were the unfinished genomes of *M. avium*, *C. diphtheriae* and *M. bovis*. Such comparative genome analysis should also be helpful in identifying genes for species-specific pathways such as the pathway for sulfolipid found in virulent strains of *M. tuberculosis*.

Analysis of the genes involved in similar pathways across all mycobacterial genomes and *Corynebacterium* will facilitate a complete understanding of the physiology of *Mycobacterium*, *Corynebacterium* and *Nocardia*, including

knowledge about their cell walls, the most characteristic and yet most obscure features of these pathogens. This will allow identification of novel drug targets, formulation of vaccines, and development of new diagnostics. The sequencing of a *Rhodococcus* genome will be a welcome addition. In the case of *M. leprae*, recombinant-protein expression and proteomics will further our understanding, because, as of today, there are no genetic tools for manipulating this pathogen. It will be some time before the insights from comparative genomics of mycobacteria yields benefits to medicine, but we can be hopeful that they are guiding us in the right direction.

Acknowledgements

Research conducted in the authors' laboratory was supported by NIH, NIAID, DMID Contract NOI AI-55262 from the National Institute of Allergy and Infectious Diseases, NIH, and by the Heiser Program for Research in Leprosy and Tuberculosis, New York City, USA.

References

1. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, et al.: **Massive gene decay in the leprosy bacillus.** *Nature* 2001, **409**:1007-1011.
2. **The *Mycobacterium leprae* genome project** [http://www.sanger.ac.uk/Projects/M_leprae/]
3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, et al.: **Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.** *Nature* 1998, **393**:537-544.
4. **The *Mycobacterium tuberculosis* genome project** [http://www.sanger.ac.uk/Projects/M_tuberculosis/]
5. **The Institute for Genomic Research** [<http://www.tigr.org/>]
6. **The *Mycobacterium bovis* genome project** [http://www.sanger.ac.uk/Projects/M_bovis/]
7. **The *Mycobacterium paratuberculosis* genome project** [<http://www.cbc.umn.edu/ResearchProjects/AGAC/Mptb/Mptbhome.html>]
8. **The *Corynebacterium diphtheriae* genome project** [http://www.sanger.ac.uk/Projects/C_diphtheriae/]
9. Russell DG: ***Mycobacterium* and the seduction of the macrophage.** In *Mycobacteria. Molecular Biology and Virulence*. Edited by Ratledge C, Dale J: Oxford: Blackwell Science; 1999: 371-388.
10. Belisle JT, Brennan PJ: ***Mycobacteria*.** In *Encyclopedia of Microbiology*. San Diego: Academic Press, 2000; 312-327.
11. Hastings RC, Gillis TP, Krahenbuhl JL, Franzblau SG. **Leprosy.** *Clin Microbiol Rev* 1988, **1**:330-348.
12. Hirata T: **Electron microscopic observations of cell wall and cytoplasmic membrane in murine human leprosy bacilli.** *Int J Lepr* 1985, **53**:433-440.
13. Draper P: **The bacteriology of *Mycobacterium leprae*.** *Tubercle* 1983, **64**:43-56.
14. McNeil M, Daffe M, Brennan PJ: **Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls.** *J Biol Chem* 1990, **265**:18200-18206.
15. McNeil M, Daffe M, Brennan PJ: **Location of the mycolyl ester substituents in the cell walls of mycobacteria.** *J Biol Chem* 1991, **266**:13217-13223.
16. Daffe M, McNeil M, Brennan PJ: **Major structural features of the cell wall arabinogalactans of *Mycobacterium*, *Rhodococcus*, and *Nocardia* spp.** *Carbohydr Res* 1993, **249**:383-398.
17. Dhariwal KR, Yang YM, Fales HM, Goren MB: **Detection of trehalose monomycolate in *Mycobacterium leprae* grown in armadillo tissues.** *J Gen Microbiol* 1987, **133**:201-209.
18. Fukunishi Y: **Electron microscopic findings of the peripheral nerve lesions of nude mouse inoculated with *M. leprae* perineural lesions.** *Int J Lepr* 1988, **56**:501.
19. Hunter SW, Brennan PJ: **A novel phenolic glycolipid from *Mycobacterium leprae* possibly involved in immunogenicity and pathogenicity.** *J Bacteriol* 1981, **147**:728-735.
20. Brennan PJ, Chatterjee D, Fujiwara T, Cho S-N: **Leprosy-specific neoglycoconjugates: synthesis and application to serodiagnosis of leprosy.** *Methods Enzymol* 1994, **242**:27-37.
21. Rambukkana A: **Molecular basis for the peripheral nerve predilection of *Mycobacterium leprae*.** *Curr Opin Microbiol* 2001, **4**:21-27.
22. Baulard AR, Besra GS, Brennan PJ: **The cell-wall core of *Mycobacterium*: structure, biogenesis and genetics.** In *Mycobacteria. Molecular Biology and Virulence*. Edited by Ratledge C, Dale J. Oxford: Blackwell Science; 1999: 240-259.
23. Belanger AE, Inamine JM: **Genetics of cell wall biosynthesis.** In *Molecular Genetics of Mycobacteria*. Edited by Hatfull GF, Jacobs WR Jr. Washington, DC: ASM Press; 2000: 191-202.
24. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.

The complete version of this article, including additional references, is available in full online.

