Volume 72, Number 4, December 2001

LEPROSY REVIEW

SPECIAL ISSUE ON THE LEPROSY GENOME

Published Quarterly for Lepra: the British Leprosy Relief Association

ISSN 0305-7518

Leprosy Review

A journal contributing to the better understanding of leprosy and its control

LEPRA

Editorial Office: LEPRA, Fairfax House, Causton Road, Colchester CO1 1PU, England
Assistant Editor: Susan Boobis, 118 Malvern Avenue, South Harrow, Middlesex HA2 9EY, England. e-mail: sboobis@aol.com

Leprosy Review is published quarterly (Mar., June, Sept., Dec.) by LEPRA (2001, vol. 72), £34 for 4 issues or £8.50 per copy, inclusive of postage and packing, UK and abroad. Subscription orders or enquiries to LEPRA, Fairfax House, Causton Road, Colchester CO1 1PU, UK. At its own discretion, LEPRA may provide free issues of this journal to doctors working with leprosy who are unable to afford the above subscription, and to selected libraries covering tropical medicine. *Leprosy Review* welcomes original papers on all aspects of leprosy, including research. The journal also publishes educational and topical information of direct benefit to the control of leprosy under field conditions. The Editorial Board may invite special articles or editorials from expert authors, and may consider the production of supplements or special editions on a particular subject or theme of major importance.

LEPRA
Registered Offices: Fairfax House, Causton Road, Colchester CO1 1PU
website:http://www.lepra.org.uk

Lepr Rev (2001) **72,** 385–386

Editorial

MYCOBACTERIUM LEPRAE GENOME SEQUENCE; A LANDMARK ACHIEVEMENT

This edition of *Leprosy Review* is given over to the remarkable achievement, finally realized at the beginning of this year, of the sequencing of the genome of *Mycobacterium leprae.*[1]

M. leprae was one of the first organisms to be associated with a human disease. However, our inability to grow the organism in culture has made it extremely difficult to study in the laboratory. The pioneering work of Charles Shepard in the 1960s,[2,3] in which limited multiplication of *M. leprae* in the footpads of mice was demonstrated, provided new opportunities and eventually paved the way for testing and optimizing new drug regimens. This provided much of the rationale for the introduction of multiple drug treatment. However, we were still unable to address fundamental questions about the organism; why does it grow so slowly? Why has it proved impossible to grow in culture medium? What is its relationship to other pathogenic mycobacteria such as *Mycobacterium tuberculosis*?

The first step on the road to answering these questions was the demonstration that experimental infection of nine-banded armadillos could produce very large numbers of organisms;[4] enough to start looking at the biochemistry and physiology of the organism. Crucially, it was also enough to extract DNA and this marked the starting point of the genome sequencing project.

Now the challenge is to try and understand, as fully as possible, the language of the genes. As you will read in this issue of *Leprosy Review*, we now know that *M. leprae* has been steadily losing genes, and with it the ability to respond to different environments. The reason why it cannot be grown *in vitro* lies in this inability; it has come to rely on its host cell for essential nutrients. Our interpretation of information revealed in the genome sequence is discussed in the papers entitled 'The decaying genome of *Mycobacterium leprae*', 'DNA metabolism in *Mycobacterium leprae*', 'The microbial physiologist's guide to the leprosy genome' and 'Genomic evidence for the retention of the essential mycobacterial cell wall in the otherwise defective *Mycobacterium leprae*'. We can also use the genome sequence to develop new tools, such as rapid methods for detecting drug resistance, as illustrated by Honoré and colleagues ('A method for rapid detection of rifampicin-resistant isolates of *Mycobacterium leprae*'), or methods for differentiating strains of *M. leprae* which could help us to understand more about its transmission ('Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity'). We can also now understand the molecular basis of drug resistance and identify potential new targets for drug development ('Genomics and the chemotherapy of leprosy').

Correspondence: e-mail: jcolsto@nimr.mrc.ac.uk

0305-7518/01/072385+2 $1.00

Finally, the papers by Jones *et al.* ('Leproma: a *Mycobacterium leprae* genome browser') and Eiglmeier *et al.* ('The integrated genome map of *Mycobacterium leprae')* explain, in practical terms, how the information can be accessed, and how it has been possible to generate a renewable source of *M. leprae* DNA, which is available to researchers interested in pursuing the molecular biology of *M. leprae*.

This issue marks a historic achievement in the study of leprosy. I am grateful to all of the authors for sharing their ideas and results and hence making it possible. I am particularly grateful to Stewart Cole for his invaluable contribution to the field and for his help in making this special edition of *Leprosy Review* possible.

National Institute for Medical Research M. Jo Colston
The Ridgeway, Mill Hill
London NW7 1AA, UK

References

[1] Cole ST, Eiglmeier K, ParkhilI J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.
[2] Shepard CC. The experimental disease that follows the injection of human leprosy bacilli into footpads of mice. *J Exp Med*, 1960; **122:** 445–454.
[3] Shepard CC, Chang YT. Effect of DDS on established infections with *Mycobacterium leprae* in mice. *Int J Lepr Other Mycobact Dis*, 1967; **35:** 52–57.
[4] Kirchheimer WF, Storrs EE. Attempts to establish the armadillo *(Dasypus novemcinctus Linn)* as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int J Lepr Other Mycobact Dis*, 1971; **39:** 693–702.

Lepr Rev (2001) **72,** 387–398

The decaying genome of *Mycobacterium leprae*

KARIN EIGLMEIER*, JULIAN PARKHILL**, NADINE HONORÉ*, THIERRY GARNIER*, FREDJ TEKAIA***, AMALIO TELENTI[+], PAUL KLATSER[++], KEITH D. JAMES**, NICOLAS R. THOMSON**, PAUL R. WHEELER[+++], CAROL CHURCHER**, DAVID HARRIS**, KAREN MUNGALL**, BART G. BARRELL** & STEWART T. COLE*

**Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France*

***Sanger Centre, Wellcome Trust Genome Campus, Hinxton, CB10 1SA, United Kingdom*

****Unité de Génétique Moléculaire des Levures, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France*

[+]Microbiologie, Centre Hospitalier Universitaire Vaudois, Rue du Bugnon, 46, 1011 Lausanne, Switzerland

[++]Royal Tropical Institute (KIT) N.H. Swellengrebel Laboratorium voor Tropische Hygiene, Meibergdreef 39, 1105 AZ Amsterdam, The Netherlands

[+++]Veterinary Laboratories Agency- Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, United Kingdom

Summary Everything that we need to know about *Mycobacterium leprae,* a close relative of the tubercle bacillus, is encrypted in its genome. Inspection of the 3·27 Mb genome sequence of an armadillo-derived Indian isolate of the leprosy bacillus identified 1,605 genes encoding proteins and 50 genes for stable RNA species. Comparison with the genome sequence of *Mycobacterium tuberculosis* revealed an extreme case of reductive evolution, since less than half of the genome contains functional genes while inactivated or pseudogenes are highly abundant. The level of gene duplication was ~34% and, on classification of the proteins into families, the largest functional groups were found to be involved in the metabolism and modification of fatty acids and polyketides, transport of metabolites, cell envelope synthesis and gene regulation. Reductive evolution, gene decay and genome down-sizing have eliminated entire metabolic pathways, together with their regulatory circuits and accessory functions, particularly those involved in catabolism. This may

* Correspondence to: S. T. Cole, Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France; Phone: 33-1-45 68 84 46. Fax: 33-1-40 61 35 83. E-mail: stcole@pasteur.fr

0305-7518/01/072387+12 $1.00

explain the unusually long generation time and account for our inability to culture the leprosy bacillus.

Introduction

Determining the complete genome sequence of a strain of *Mycobacterium leprae* was one of the highest priorities defined for leprosy research and control programmes at the joint WHO/Sasakawa Memorial Health Fund meeting held in Bangkok in 1995. It was clear to all participants that genomics would not only provide understanding of the unusual biology of *M. leprae* but that the information thus obtained would underpin leprosy research in what are hopefully the final years of the elimination campaign. The choice of the strain to be sequenced was influenced by the disease burden and as a result we chose to work with a patient isolate from Tamil Nadu, India, one of the worst affected countries.

In order to produce sufficient bacilli to extract DNA for library construction this strain, referred to as TN, was passaged in the armadillo.[1] An immortalized source of DNA was then obtained in the form of a cosmid library and this will facilitate future functional studies.[2] The DNA sequence of the TN strain of *M. leprae* was obtained by a combined approach employing automated DNA sequence analysis of selected cosmids and whole genome shotgun clones.[2,3] After assembling the subsequences, the complete genome sequence was regenerated and this was subjected to bioinformatic analysis so that the genes, their control signals, repetitive elements and other genomic features could be identified. Comparison of the genome sequence and gene set of the TN strain with those of other organisms was then undertaken. Comparative genomic analysis with the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, was found to be a particularly powerful approach that enabled much of the biochemistry, physiology and genetics of *M. leprae* to be predicted and unraveled. The principal findings of this analysis were presented in the original landmark publication which the reader should consult for further details.[4] Here we present a brief summary and an update of some of the findings.

Results and discussion

ORGANIZATION AND SEQUENCE OF THE GENOME

The complete genome sequence of the TN strain of *M. leprae* was found to contain 3,268,203 bp, and to have an average G + C content of 57·8%. There is a single circular chromosome and no plasmids. Bioinformatics predicted the existence of 1605 genes encoding proteins and a further 50 that code for stable RNA molecules. These values are much lower than those reported for the genome of the H37Rv strain of *M. tuberculosis*, comprising ~4000 genes, at 4,411,532 bp and 65·6% G + C.[5]

On analysis of the first complete *M. leprae* cosmid sequence in 1993, only 50% of the sequence was found to contain functional genes.[3] At that time, meaningful comparisons were very limited as few other sequences were available and no clues could be found to the possible function of the remaining 50% of the DNA. However, the availability of the *M. tuberculosis* genome sequence in 1998 changed things radically and enabled detailed pairwise comparisons of the genome and proteome sequences of both pathogens to be undertaken.[5,6] These revealed that only 49·5% of the genome of *M. leprae* was occupied by protein-coding genes while at least 27% of the sequence contained recognizable

Figure 1. See enclosed poster. Organization of the genome of the TN strain of *M. leprae*. Only those genes predicted to be active are shown and their functions are colour coded as follows: lipid metabolism, black; intermediary metabolism and respiration, yellow; information pathways, pink; regulatory proteins, sky blue; conserved hypothetical proteins, orange; unknown, light green; insertion sequences and phage-related functions, blue; stable RNAs, dark blue; cell wall and cell wall processes, dark green; virulence, detoxification and adaptation, white; PE and PPE protein families, magenta. For the sake of clarity the pseudogenes are not indicated but these occur in all empty regions. The scale is shown with 10 or 100 kb intervals. Genes shown above the line are transcribed from left to right, those below from right to left.

pseudogenes, inactive reading frames with functional counterparts in the tubercle bacillus. The remaining 23·5% of the genome did not appear to be coding at all, and probably contains gene remnants mutated beyond recognition together with the regulatory sequences that usually occur in intergenic regions.

The distribution of the 1,114 recognizable pseudogenes was essentially random throughout the genome (Figure 1; see attached poster of *M. leprae* genome). However, the 1605 potentially active genes tend to occur in clusters often flanked on both sides by long stretches of seemingly non-coding DNA.

REDUCTIVE EVOLUTION

The process by which large-scale loss of gene function arises has been termed reductive evolution. It has been observed in a number of important human pathogens such as the obligate intracellular parasites *Rickettsia* and *Chlamydia* spp.[7] and this suggests that genes become inactivated once their functions are no longer required in these highly specialized niches. In some endosymbionts such as the *Buchnera* spp., which are related to the enteric bacteria and found in aphids, reductive evolution has proceeded so extensively that the genome size is thought to have been reduced from ~4·5 to 0·64 Mb.[8] There are few pseudogenes in this case, and deletion appears to have been the dominant means of genome downsizing. One hypothesis, known as Muller's ratchet, has been proposed to explain reductive evolution. This involves the stochastic loss of genetic material and results in decreased fitness and little genetic variability. In part, this is due to the inability of organisms with no sex-cycle to acquire DNA and hence to repair genetic lesions through acquisition of new genes or by recombination. Obviously, as a consequence of its highly specialized niche, the only organism with which *M. leprae* can exchange DNA is the human host.

Until the sequence of *M. leprae* became available, the most extensive genome degradation reported in a pathogen was in *Rickettsia prowazekii*, the typhus agent, where only 76% of the potential coding capacity was used.[9] In comparison with *M. leprae*, the level of gene loss detected in *R. prowazekii* was modest, and it is notable that elimination of pseudogenes by deletion lags far behind gene inactivation in both pathogens, in contrast to *Buchnera*.[8] Interestingly, the G + C content of the functional genes in the leprosy bacillus (60·1%) is higher than that of the recognisable pseudogenes (56·5%), which is in turn greater than that of the remainder of the genome (54·5%), which may have undergone the most extensive decay. This suggests that the relatively high G + C content of *M. leprae*, and by extension, the other mycobacteria, is driven by the codon preference of the active genes, while random mutation within the non-coding regions causes them to drift towards a more neutral G + C content that is closer to that of the host. Deamination of cytosine residues in the DNA is a possible mechanism to account for this trend. This process would account for the leprosy bacillus having the lowest G + C content of all mycobacteria and it is noteworthy that the genomes of organisms that have undergone reductive evolution are generally richer in A + T.[10]

When one examines closely the genes that have been lost or inactivated during the reductive evolutionary process, clear trends are observed that conform to Darwinian theory and testify to the importance of selective pressure. For instance, *M. tuberculosis* is capable of anaerobic respiration using nitrate as terminal electron acceptor in a reaction catalysed by nitrate reductase and using electrons from the quinone pool of the respiratory chain. Nitrate reductase comprises four subunits, encoded by *narGHIJ,* and uses a complex cofactor, molybdopterin, which is synthesized by at least nine *moe/moa* genes and requires molybdate to be taken up from the extracellular medium by the ABC-transporter encoded by *modABC*.[5] In the tubercle bacillus, the only other enzyme predicted to use molybdopterin as cofactor is formate dehydrogenase and many microbes are capable of growing on a defined medium containing formate and nitrate as sole carbon and energy sources. *M. leprae* has pseudogenes corresponding to both of these enzymes and for almost all of the proteins required to transport molybdate and to insert it into the pterin ring. Apparently, once the need to use the formate-nitrate pathway was lost, the genes for the entire system acquired mutations and decayed as none of their functions was required.

GENOME DOWNSIZING

Reductive evolution involves gene loss through mutational inactivation and deletion. If one makes the reasonable assumption that the genomes of *M. leprae* and *M. tuberculosis* were once topologically equivalent and roughly 4·4 Mb in size, as is the case for many other slow-growing mycobacteria,[11–13] then extensive downsizing must have occurred during evolution of the leprosy bacillus since its genome is <75% of the size of that of *M. tuberculosis*.[4] It seems likely that recombination events involving repetitive DNA were responsible for both a reduction in genome size and the rearrangement of chromosomal segments that led to the loss of global synteny between the genomes of the tubercle and leprosy bacilli.[4] This is discussed further in this issue.[14]

A 1·1 Mb reduction in the size of the genome would have eliminated ~1100 protein-coding sequences, and *M. leprae* should, therefore, produce 3000 proteins compared to the 4000 predicted in *M. tuberculosis*. On proteomic analysis of *M. leprae*, only 391 soluble protein species were detected,[15] compared to nearly 1800 in *M. tuberculosis*.[16] This is entirely consistent with the large number of pseudogenes in *M. leprae* and excludes the possibility of gene expression by a novel mechanism such as RNA editing. Since diverging from the last common mycobacterial ancestor, the leprosy bacillus may have lost over 2000 genes, and reductive evolution has probably defined naturally the minimal gene set for a pathogenic mycobacterium.

At present, little is known about the genomic diversity of different isolates of *M. leprae* as very few studies have been undertaken. No differences in genomic organization were uncovered by restriction fragment length polymorphism analysis[17] and the systematic sequence analysis of two regions of the genome of several different isolates of *M. leprae* revealed no single nucleotide polymorphisms.[3] These regions correspond to the pseudogenes ML1873, ML1874, ML1884 and ML1885, which are functionally inactive and should, therefore, be more prone to divergence as they are under less selective pressure. This sequence conservation suggests that the immediate ancestor of the present leprosy bacillus had already undergone reductive evolution and that a single clone then expanded and was disseminated globally. Sequence and micro-array analysis of more *M. leprae* isolates will shed further light on possible strain divergence and the emergence of the disease.

GENES ENCODING PROTEINS

Of the 1605 genes predicted in *M. leprae*, there are 1440 which are also found in tubercle bacilli and 165 genes that have no orthologue in *M. tuberculosis*. Of the latter, some functional information could be predicted for 29 genes by bioinformatics, while the remaining 136 show no similarity to known genes elsewhere. It is highly likely that many of these will also be pseudogenes as they are generally shorter than average and occur in regions of low gene density.[4] In consequence, *M. leprae* may produce as few as 1500 proteins. In most cases, these proteins have been assigned roles in housekeeping functions, information pathways, various metabolic pathways, and so on,[4] and they will not be discussed further here. A complete functional list may be found at http://www.pasteur.fr/recherche/unites/Lgmb/NATURE_DATA/ML_gene_list. From Figure 2, which presents the classification into the 11 functional categories used previously,[5] it can be seen that the major groups contain genes for central and intermediary metabolism, cell wall processes and conserved hypothetical proteins. Issues such as why *M. leprae* grows so slowly and resists attempts at in vitro culture are discussed elsewhere in this issue.[18] Instead, we will now concentrate on those proteins, and the principal protein families, that have not been described previously in detail.

MULTIGENE FAMILIES

The majority of the genes (52%) present in *M. tuberculosis* arose from gene duplication events and this process may have conferred extensive functional redundancy.[6] Many of these genes are involved in lipid metabolism or encode the novel glycine-rich proteins of the PE and PPE families. Slightly more than 34% of the proteins now found in *M. leprae* are the products of gene duplication events, or share common domains, as defined by the bioinformatic routines used here.[6,19] There are far fewer families of duplicated proteins than in the tubercle bacillus and they are much smaller in size as a result of reductive evolution. For instance, whereas *M. leprae* contains 87 families containing two members and 37 with three members, *M. tuberculosis* has 213 and 72, respectively.

The duplicated genes have been classified into partitions on the basis of their similarity and part of this information is presented in Table 1. As in the tubercle bacilli, the largest partition, P46.1, contains enzymes involved in polyketide synthesis and fatty acid metabolism (Table 1), and this again underlines the importance of these activities to slow growing mycobacterial pathogens. The second and third largest protein families in *M. tuberculosis* are those for the 167 PE and PPE proteins; their counterparts in *M. leprae* are considerably smaller accounting for only 12 members (P5.8, P7.2, Table 1) collectively, although about 30 pseudogenes also exist. None of the PE or PPE proteins predicted to be produced by *M. leprae* contain the multiple C-terminal repetitions that are suspected of being involved in antigenic variation.[5] Retraction of these gene families, which are exceptionally GC-rich, partly contributes to the difference in the respective genome sizes of the tubercle and leprosy bacilli, and to the much lower GC content of *M. leprae*.

From Table 1, it can be seen that the major protein families are involved in lipid or polyketide metabolism, modification and synthesis of cell envelope components (methyltransferases, glycosyltransferases), transport processes (ABC transporters, MmpL proteins), or in gene regulation (TetR, WhiB, two-component system response regulators).

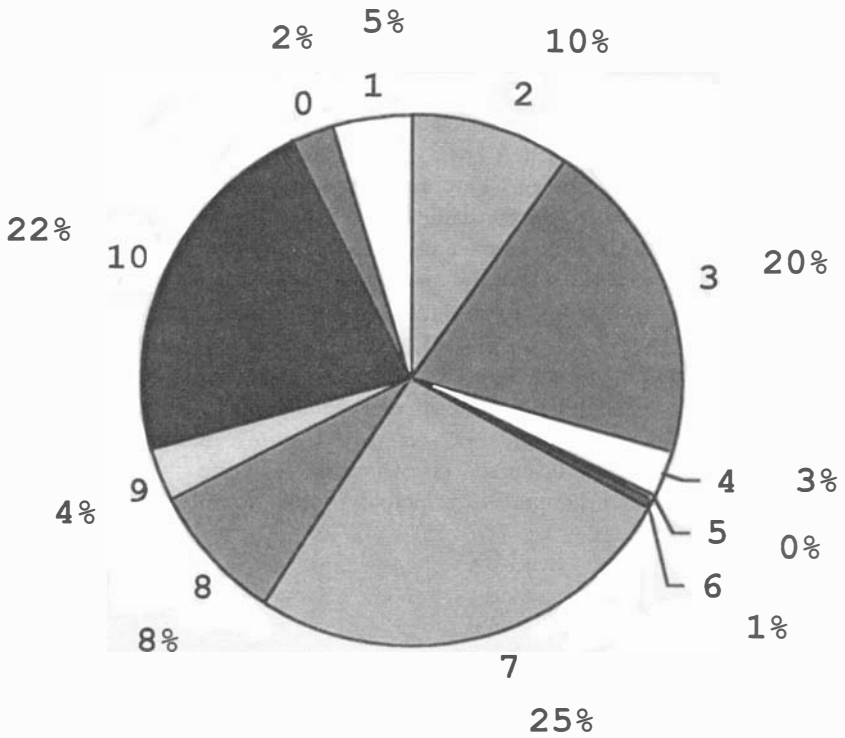

Figure 2. Broad functional classification of genes found in *M. leprae*. The functional groups and the percentage of total gene complement are shown, and correspond to: 1, lipid metabolism; 2, information pathways; 3, cell wall and cell wall processes; 4, stable RNAs; 5, insertion sequences and phage-related functions; 6, PE and PPE protein families; 7, intermediary metabolism and respiration; 8, unknown function; 9, regulatory proteins; 10, conserved hypothetical proteins; 0, virulence, detoxification and adaptation.

REGULATION AND SIGNAL TRANSDUCTION

Intracellular pathogens such as *Chlamydia* and *Rickettsia*,[10] as well as endosymbionts like *Buchnera*,[8] tend to have lost most of the genes involved in regulation, and *M. leprae* also displays this trend. The number of regulatory proteins predicted is roughly one-third of that recorded for *M. tuberculosis*, as a result of reductive evolution, and details of those remaining are presented in Table 2.

At the level of promoter recognition and transcription initiation controlled by the sigma factors of RNA polymerase, major differences exist between the tubercle and leprosy bacilli and these may explain some of the differences in their physiology. There are 13 sigma factors in *M. tuberculosis*, 10 of which belong to the extra-cytoplasmic function (ECF) family.[5] Only two of these ECF genes are still functional in *M. leprae, sigC* and *sigE,* while the eight others are present as pseudogenes. The sigma-70 family sigma factors SigA and SigB are predicted to be functional, however, whereas the 13th sigma factor gene, *sigF,* is inactive. It has been found recently in *M. tuberculosis* that the ability to survive at temperatures of 37°C and above is partly controlled by the ECF sigmas E and H, acting through the sigma 70 factor SigB.[20,21] In addition to the heat shock response, loss of *sigE* also affects resistance to SDS and oxidative stress, and survival in macrophages.[20,21] The

Table 1. Prominent protein families in *M. leprae*

Family	Description	Metabolic class
P5.1	Various methyl/cyclopropane mycolic acid synthases	Fatty acid metabolism
P5.10	WhiB transcriptional regulators	Gene regulation
P5.2	GTP-binding proteins	Regulation
P5.3	sugar-phosphate nucleotidyl transferases	Cell wall functions
P5.4	two-component system response regulators	Signal transduction
P5.5	acetohydroxyacid/acetolactate synthases	Central metabolism
P5.6	glycosyl transferases	Cell wall functions
P5.7	membrane/cell division proteins	Transport
P5.8	PE family	Unknown
P5.9	celldivision/anion transporting ATPase	Transport
P6.1	MmpL, conserved large membrane proteins	Transport
P6.2	phosphoserine phosphatase/acyl transferases	Fatty acid metabolism
P6.3	Mce proteins	Pathogenesis
P6.4	ABC-transport protein, inner membrane component	Transport
P7.1	methyltransferases	Fatty acid metabolism
P7.2	PPE-family (6)	Unknown
P8.1	conserved (membrane) protein	Transport
P8.2	phosphoglycerate mutases/mutT1	Central metabolism
P9.1	enoyl-CoA hydratase/isomerase	Fatty acid metabolism
P10.1	TetR-family transcriptional regulators	Regulation
P13.1	conserved hypothetical proteins	Unknown
P18.1	ABC-transport protein. ATP-binding component	Transport
P46.1	Polyketide synthesis/fatty acid metabolism	Fatty acid metabolism

optimal growth temperature of *M. leprae* is 32°C and this probably explains why the bacterium, which contains a full complement of heat shock proteins, multiplies principally in the extremities of the human body. It is conceivable that mutational inactivation of the sigma factor gene *sigH*, and possibly others, leads to lowered production of SigB at higher temperatures, and this may have led to *M. leprae* colonizing cooler regions of the body such as the skin and ears.

The interaction of RNA polymerase with the promoter regions of genes is often influenced by repressors or transcriptional activators that ensure expression under defined physiological conditions or in response to availability of a given substrate. In *M. tuberculosis* there are >110 proteins that have broad regulatory potential and in the leprosy bacillus this number has dwindled to 46, suggesting that the organism resides within a more stable niche than *M. tuberculosis (*Table 2). These regulatory proteins can be classed in the corresponding families on the basis of the characteristic motifs that they contain and, with one exception, all of these families are substantially smaller than those of the tubercle bacillus. The exception is the WhiB family[22] as there are five of these proteins in *M. leprae* compared to seven in *M. tuberculosis* (Table 2). The cysteine-rich WhiB proteins, also known as WhmA-G,[23] are confined to the *Actinomycetes* but in *Streptomyces* spp. they regulate developmental processes such as sporulation. Although mycobacteria do not sporulate, the change from exponential growth to persistancy or dormancy can be considered as a different state of development and the *whiB2* ortholog of *M. smegmatis, whmD*, has been found to be essential for septum formation and cell division.[24] The fact that so many WhiB proteins have been preserved in the face of reductive evolution strongly argues for their playing a major biological role in *M. leprae*.

Table 2. The regulatory repertoire of *M. leprae*

Sigma factors		
sigA (*rpoT*)	ML1022	RNA polymerase sigma-70 factor
sigB	ML1014	RNA polymerase sigma-70 factor
sigC	ML1448	ECF subfamily sigma factor
sigE	ML1076	ECF subfamily sigma subunit
Two component regulatory systems		
—	ML0174	response regulator
—	ML0175	sensor kinase
mtrA	ML0773	response regulator
mtrB	ML0774	sensor kinase
—	ML2123	response regulator
—	ML2124	sensor kinase
regX3	ML2439	response regulator
senX3	ML2440	sensor kinase
—	ML0803	sensor kinase
—	ML1286	response regulator
Serine-Threonine protein kinases and phosphoprotein phosphatases		
pknA	ML0017	serine-threonine protein kinase
pknB	ML0016	serine-threonine protein kinase
pknG	ML0304	serine-threonine protein kinase
pknL	ML0897	serine-threonine protein kinase
ppp	ML0020	probable phosphoprotein phosphatase
Repressors/activators		
TetR/AcrR-family		ML0064, ML0316, ML0717, ML0815, ML0949, ML1070, ML1733, ML2457, ML2568, ML2677
WhiB-family		ML0382 (*whiB3*), ML0639, (*whiB7*), ML0760 (*whiB2*), ML0804 (*whiB1*), ML2307 (*whiB4*)
MarR-family		ML0550, ML2696, ML2140
ArsR-family		ML0825
LysR-family		ML2041 (*oxyR*), ML2663 (*oxyS*)
NifR3-family		ML2186
Crp/Fnr-family		ML2302
Other families		ML0565 (*whiA*), ML0824 (*furB*), ML0988 (*recX*), ML1003 (*lexA*), ML1013 (*ideR*), ML1411 (*argR*), ML2188 (*phoY1*)
Possible others		ML0320, ML0592, ML0898, ML0919, ML1320, ML1328, ML1330, ML1367, ML1419, ML1652, ML1753, ML1783, ML2063, ML2156, ML2429, ML2530

Bacteria respond to changes in environmental conditions by means of diverse signal transduction systems that control gene expression through phosphorylation of regulatory proteins. These are of two types in mycobacteria, the classical two-component systems, comprising membrane-bound histidine protein kinases and phospho-aspartyl response regulators,[25] and the eukaryotic-like serine-threonine protein kinase phosphorelay system (STPK).[26,27] Of the 11 complete two-component systems present in the tubercle bacillus, only four have been retained by *M. leprae* together with isolated genes coding for a single histidine protein kinase and a response regulator. One of these systems, *mtrAB,* is essential for *M. tuberculosis* and this is almost certainly true of *M. leprae* as well. Indeed, one can

speculate with confidence that orthologs of the remaining two-component system genes of *M. leprae* will also prove to be essential in the tubercle bacillus.

Although formal evidence is still lacking, it is highly likely that the STPK, are also involved in signal transduction given their similarity to many other enzymes of the STPK superfamily that play this role. Only four of the 11 STPK genes of *M. tuberculosis* and the putative phosphoprotein phosphatase (Table 2), have functional orthologues in *M. leprae*. Indeed, it has been suggested that three of these genes *pknA*, *pknB* and *ppp* may control the timing of cell division or septation as they occur in an operon with other division genes.[26]

PATHOGENICITY

In some microbes, the combination of genomics and bioinformatics has been of great value in identifying genes for potential virulence factors that augment the degree of pathogenicity, and an excellent example is provided by the genome of the plague bacillus, *Yersinia pestis*.[28] When bioinformatic tools such as the GC skew[29] or dinucleotide bias[30] were applied to this genome, several new pathogenicity or adaptation islands of atypical base composition were uncovered. Using similar approaches to investigate the *M. leprae* genome, no such islands were detected, although they did provide evidence for recent chromosomal rearrangements.[4] Likewise, when database searches were performed few hits to genes for known virulence factors were obtained. Similar observations have been made previously for the tubercle bacillus,[31] and it now seems unlikely that any virulence genes were acquired. Instead, it seems more probable that the ability to survive in the macrophage or Schwann cell and hence to persist in the body represent the major determinants of pathogenicity. Although our understanding of the initial steps in infection of Schwann cells by *M. leprae* has improved considerably, thanks to the definition of the roles of laminin-binding protein and phenolic glycolipid 1 in this process,[32–36] we know little about the ensuing events or mycobacterial persistance in either Schwann cells or macrophages. Nevertheless, although genomics has not pinpointed a handful of potential candidate virulence genes, the sample size has been reduced to a tangible level by comparative genomics. A set of ~120 genes of unknown function has been defined that are common to both the leprosy and tubercle bacilli but no other sequenced pathogens[4] and this includes the *mce* genes encoding the cell entry factors[37,38] (Table 1). Testing their role in pathogenesis can now be undertaken by means of surrogate functional genomics.

NOVEL FUNCTIONS, HORIZONTAL GENE TRANSFER AND IMMUNODIAGNOSTICS

As outlined above, while most of the *M. leprae* genes have orthologues in *M. tuberculosis*, there are several that appear to be unique and may have novel activities. These include hypothetical proteins of unknown function and a number of potential enzymes such as the inorganic pyrophosphatase encoded by *ppa*, prolyl-tRNA synthetase, a eukaryotic-like uridine phosphorylase, phospho-*enol*-pyruvate carboxylase, adenylate cyclase, cytochrome P450 and enoyl-CoA hydratase. Furthermore, there are two transport systems that may play significant physiological roles: an ABC-transporter for sugars, and a second Nramp1-like protein, possibly involved in divalent metal ion uptake[39] that may offset the apparent absence of a siderophore system. It is probable that the phospho-*enol*-pyruvate (PEP) carboxylase replaces the pyruvate carboxylase of *M. tuberculosis*, as this enzyme is missing from *M. leprae*, and intervenes in the anaplerotic pathways. There is only one cytochrome

P450 (ML2088) present in the leprosy bacillus, compared to 20 in *M. tuberculosis* and, as this enzyme has no counterpart in *M. tuberculosis,* its function might be specific.

There is evidence that some of these enzymes have been acquired as a result of horizontal gene transfer and this is best illustrated by the prolyl-tRNA synthetase, ProS, which is the sole aminoacyl-tRNA synthetase of *M. leprae* with no counterpart in *M. tuberculosis.* Surprisingly, ProS is more similar to the enzymes of *Borrelia burgdorferi* and to eukaryotes such as Drosophila, humans and yeast. It has been proposed that horizontal transfer of tRNA synthetase genes occurs frequently, and that the pathogen *B. burgdorferi* may have acquired *proS* from its host.[40] Comparison of the genetic neighbourhood provides further support for this hypothesis as the *M. leprae proS* is both displaced and inverted with respect to the *M. tuberculosis* genome,[4] consistent with recent acquisition. In this case, the domain structure of the enzyme is indicative of a eukaryotic origin, and the human host appears the most likely candidate. Another example is found in the case of uridine phosphorylase, an enzyme that is not common in bacteria, as the closest relative of the *M. leprae* protein is that of the mouse.

Ensuring the elimination of leprosy as a public health problem will require both continued implementation of multidrug therapy and improved detection of infected individuals. Diagnosis is difficult in patients with few lesions and accurate information about subclinical infection is often rare. The identification of proteins that may be specific for the leprosy bacillus opens up new avenues for the development of immunodiagnostic tests possibly of the transdermic kind. In the post-genomic era it is important that the immunogenicity of these polypeptides be appraised and, if the initial findings are promising, attempts should be made to produce batches suitable for field testing. Determining the genome sequence of *M. leprae* has taught us much about the biology of the pathogen but it is the application of this new knowledge to disease control that should be prioritized now.

Acknowledgements

We are grateful to B. R. Bloom, P. J. Brennan, M. J. Colston, J. Grosset, and B. Ji for their advice, reagents and encouragement. This work was supported by the New York Community Trust, ILEP, the Association Française Raoul Follereau, the Wellcome Trust, the Institut Pasteur, and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

References

[1] Kirchheimer WK, Storrs EE. Attempts to establish the armadillo (*Dasypus novemcinctus Linn.*) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int J Lepr*, 1971: **39:** 693–702.

[2] Eiglmeier K, Honoré N, Woods SA *et al.* Use of an ordered cosmid library to deduce the genomic organisation of *Mycobacterium leprae*. *Mol Microbiol*, 1993: **7:** 197–206.

[3] Honoré N, Bergh S, Chanteau S *et al.* Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol Microbiol*, 1993: **7:** 207–214.

[4] Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001: **409:** 1007–1011.

[5] Cole ST, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998: **393:** 537–544.

[6] Tekaia F, Gordon SV, Garnier T *et al*. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle Lung Dis*, 1999: **79:** 329–342.
[7] Andersson JO, Andersson SGE. Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev*, 1999: **9:** 664–671.
[8] Shigenobu S, Watanabe H, Hattori M *et al*. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 2000: **407:** 81–86.
[9] Anderssen SGE, Zomorodipour A, Andersson JO *et al*. The complete genome sequence of the obligate intracellular parasite *Rickettsia prowazekii*. *Nature*, 1998: **396:** 133–140.
[10] Tamas I, Klasson LM, Sandstrom JP, Andersson SG. Mutualists and parasites: how to paint yourself into a (metabolic) corner. *FEBS Lett*, 2001: **498:** 135–139.
[11] Brosch R, Gordon SV, Eiglmeier K *et al*. Comparative genomics of the leprosy and tubercle bacilli. *Res Microbiol*, 2000: **151:** 135–142.
[12] Philipp W, Schwartz DC, Telenti A, Cole ST. Mycobacterial genome structure. *Electrophoresis*, 1998: **19:** 573–576.
[13] Stinear TP, Jenkin GA, Johnson PDR, Davies JK. Comparative genetic analysis of *Mycobacterium ulcerans* and *Mycobacterium marinum* reveals evidence of recent divergence. *J Bacteriol*, 2000: **182:** 6322–6330.
[14] Cole ST, Supply P, Honoré N. Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. *Lepr Rev*, 2001, **72:** 387–398.
[15] Marques MAM, Chitale S, Brennan PJ, Pessolani MCV. Mapping and identification of the major cell-wall associated components of *Mycobacterium leprae*. *Infect Immun*, 1998: **66:** 2625–2631.
[16] Jungblut PR, Schaible UE, Mollenkopf H-J *et al*. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol*, 1999: **33:** 1103–1117.
[17] Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol*, 1990: **4:** 1653–1659.
[18] Wheeler P. The microbial physiologists guide to the leprosy genome. *Lepr Rev*, 2001, **72:** 399–407.
[19] Tekaia F, Dujon B. Pervasiveness of gene conservation and persistence of duplicates in cellular genomes. *J Mol Evol*, 1999: **49:** 591–600.
[20] Manganelli R, Dubnau E, Tyagi S *et al*. Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol*, 1999: **31:** 715–724.
[21] Manganelli R, Voskuil MI, Schoolnik GK, Smith I. The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol*, 2001: **41:** 423–437.
[22] Hutter B, Dick T. Molecular genetic characterisation of whiB3, a mycobacterial homologue of a *Streptomyces* sporulation factor. *Res Microbiol*, 1999: **150:** 295–301.
[23] Soliveri JA, Gomez J, Bishai WR, Chater KF. Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene whiB are present in *Streptomyces* and other actinomycetes. *Microbiology*, 2000: **146:** 333–343.
[24] Gomez JE, Bishai WR. whmD is an essential mycobacterial gene required for proper septation and cell division. *Proc Natl Acad Sci USA*, 2000: **97:** 8554–8559.
[25] Grebe TW, Stock JB. The histidine protein kinase superfamily. *Adv Microb Physiol*, 1999: **41:** 139–227.
[26] Fsihi H, De Rossi E, Salazar L *et al*. Gene arrangement and organisation in a ~76 kilobase fragment encompassing the oriC region of the chromosome of *Mycobacterium leprae*. *Microbiology*, 1996: **142:** 3147–3161.
[27] Av-Gay Y, Davies J. Components of eukaryotic-like protein signaling pathways in *Mycobacterium tuberculosis*. *Microb Comp Genom*, 1997: **2:** 63–73.
[28] Parkhill J *et al*. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, 2001: **412:** 777–999.
[29] Lobry JR. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol*, 1996: **13:** 660–665.
[30] Karlin S. Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr Opin Microbiol*, 1998: **1:** 598–610.
[31] Cole ST. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett*, 1999: **452:** 7–10.
[32] Ng V, Zanazzi G, Timpl R *et al*. Role of the cell wall phenolic glycolipid-1 in the peripheral nerve predilection of *Mycobacterium leprae*. *Cell*, 2000: **103:** 511–524.
[33] Rambukkana A, Salzer JL, Yurchenco PD, Tuomanen EI. Neural targeting of *Mycobacterium leprae* mediated by the G domain of the laminin-α2 chain. *Cell*, 1997: **88:** 811–821.
[34] Rambukkana A, Yamada H, Zanazzi G *et al*. Role of alpha-dystroglycan as a Schwann cell receptor for *Mycobacterium leprae*. *Science*, 1998: **282:** 2076–2079.
[35] Rambukkana A. Molecular basis for the peripheral nerve predilection of *Mycobacterium leprae*. *Curr Opin Microbiol*, 2001: **4:** 21–27.
[36] Shimoji Y, Ng V, Matsumura K *et al*. A 21-kDa surface protein of *Mycobacterium leprae* binds peripheral nerve laminin-2 and mediates Schwann cell invasion. *Proc Natl Acad Sci USA*, 1999: **96:** 9857–9862.

[37] Arruda S, Bomfim G, Knights R *et al.* Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science*, 1993: **261:** 1454–1457.
[38] Chitale S, Ehrt S, Kawamura I *et al.* Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. *Cell Microbiol*, 2001: **3:** 247–254.
[39] Makui H, Roig E, Cole ST *et al.* Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. *Mol Microbiol*, 2000: **35:** 1065–1078.
[40] Wolf YI, Aravind L, Grishin NV, Koonin EV. Evolution of amino-acyl-tRNA synthetases – analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res*, 1999: **9:** 689–710.

Lepr Rev (2001) **72,** 399–407

The microbial physiologist's guide to the leprosy genome

PAUL R. WHEELER
Tuberculosis Research Unit, Veterinary Laboratories Agency
Weybridge, New Haw, Surrey, UK, KT15 3NB, UK

Introduction

One of the most exciting things about living in the genome sequencing era is that hitherto intractable pathogens can be understood for the first time in unimaginable detail. Many bacterial genomes have been sequenced and one can now enquire about their physiology. It is possible not just to enquire about *Mycobacterium leprae*, but make comparisons with other host-dependent and axenically uncultivable pathogens[1] such as *Treponema pallidum*, that have also had their genomes sequenced.

Previously, to study the metabolism and biochemistry of leprosy bacilli one had to purify them from the heavily infected tissues of armadillos or nude mice. Some data were obtained from less heavily infected sources including even human biopsies but from such limited material only very limited data could be obtained. To be able to start to deduce metabolic pathways in operation, greater numbers of leprosy bacilli from more heavily infected tissue were needed.[2] Regardless of the source, it was always essential to prove that any metabolic activity was from the bacteria, and not contamination from the host. For example, catalase activity is readily detected in suspensions and extracts of leprosy bacilli but biochemical[3] and genetic[4,5] analysis show that this was host-derived, and there was no mycobacterial KatG (see Box 1) catalase activity. All this was difficult, but perhaps on balance easier than having to extract bacteria from rabbit testicles as is the case for the aforementioned agent of syphilis, *T. pallidum*.

Now there is genome data to mine. The truth is in there, or rather predictions of the truth are. It is important to keep in mind that deducing a function for a gene (that is, to give an example, referring to ML2323 as *ask*, the gene for aspartokinase) are predictions of function made on the basis of similarities to raw sequence information elsewhere in nature. Many of the similarities are so high, and so consistent throughout a metabolic pathway, that the attributions are essentially incontestable. These have been compiled into metabolic pathways in the KEGG database (http://www.genome.ad.jp/kegg/) where, for example, the nucleotide biosynthetic pathways referred to in Box 1 can be viewed. Other areas of metabolism are still

*Correspondence to: P. R. Wheeler (e-mail: pwheeler.vla@gtnet.gov.uk)

0305-7518/98/072399+09 $1.00

Box 1. The physiology of *Mycobacterium leprae* as deduced from its genome

- Uniquely for a host-dependent pathogen, the biosynthetic pathways for purine and pyrimidine nucleotides are complete.
- Biosynthetic pathways for lipids and amino acids are also complete with the exception that *M. leprae* is a methionine auxotroph.
- Only central pathways of carbon and energy metabolism are complete but alternative pathways are degenerate.
- The co-enzyme central to the most universal metabolic pathways – NADH cannot be recycled to NAD by the usual oxidative respiratory route.
- Redundancy seen in *M. tuberculosis* is often lost in *M. leprae*, as most paralogues seen in *M. tuberculosis* are pseudogenes (= non-functional) in *M. leprae*.
- Defence against toxic radicals is severely degenerate, as neither *katG* nor the *narGHJI* cluster is functional.
- None of the few (142) additional genes only found in *M. leprae* appear to confer additional metabolic pathways.

far from clear. Some of these are explored in this review, such as whether *M. leprae* is capable of the biosynthesis of siderophores for the vital activity of acquisition of iron from the environment. This is but one example of how the genomic data generates hypotheses to be tested. There are yet others to be generated. The 'take-home message' is that the genome sequence is not the conclusion of research into any particular microbe, but merely the 'end of the beginning', a beginning started with rational research on microbial metabolism and the action of antimicrobial agents by Pasteur, Entner and Doudoroff, Stamp and Green amongst others.[6]

Anabolic and catabolic pathways – what's different about *M. leprae?*

The big questions in bacterial physiology are: how do bacteria obtain their nutrition, use it for production of energy (catabolism) and make all their components – proteins, lipids, cell envelope, nucleic acids and so on (anabolism). The first four points in Box 1 summarize how the genome project has shown how *M. leprae* differs from other microbes in general, while the fifth point shows how *M. leprae* differs from *Mycobacterium tuberculosis* in particular.

Microbes of medical interest vary considerably in their metabolic capabilities. At one extreme, some are autotrophs so that they can make everything they need from ammonium ions and carbon dioxide. At the other extreme, there are highly host-dependent organisms such as *Rickettsia* and treponeme species.[1] A classical example of their dependency upon the host for preformed intermediates is that they are unable to synthesize the structures on which nucleotides are based – purine and pyrimidine rings that include four to five carbon atoms and two to five nitrogen atoms. Thus, like many parasitic microbes, they have to scavenge purines and pyrimidines from their host. Intuitively, *M. leprae* would be expected to lie on the host-dependent end of this scale, since it cannot be grown axenically and it grows almost by stealth when it is in the host, dividing no more rapidly than every 12 days (growth rate $\sim 0.0035\ h^{-1}$).[7] However, a surprising finding from the genome sequence is that *M. leprae* has the capacity to make most of its own components. For example, although it is able to obtain lipids, amino acids, purines and pyrimidines directly from the host,[4] it can also make all these from the simplest components. In this respect its metabolism is no different from *E. coli* or *M. tuberculosis*. The one exception is that *M. leprae* cannot make methionine and must

Table 1. Distribution of genes by selected functions

Category	*M. tuberculosis* genes	*M. leprae* genes	*M. leprae* pseudogenes
Energy metabolism	300	95	120
Nucleotide and nucleoside metabolism (includes purine and pyrimidine biosynthesis)	65	57	5
Amino acid biosynthesis	91	78	5
Lipid biosynthesis	68	37	12
PE/PPE	167	8	8
Virulence determinants	39	16	4

depend upon the host for this one amino acid. This one piece of degeneracy in the major anabolic pathways is evidently something that pathogenic mycobacteria can tolerate, since experimentally obtained methionine auxotrophs of *M. tuberculosis* show little or no attenuation.[8,9]

Compare now in Table 1 the difference between the distribution of genes for the anabolic pathways I have just mentioned and those for energy metabolism. The former are virtually complete, with losses in *M. leprae* compared with *M. tuberculosis* being down to loss of redundancy more often than function. Thus, whereas *M. tuberculosis* may have two or more genes capable of performing a particular function, *M. leprae* frequently has just one. Even in the case of lipid biosynthesis, central pathways are complete with the losses reflecting the more limited, but still extensive, repertoire of lipids produced by *M. leprae*. However, for energy metabolism (catabolism) *M. leprae* has less than one-third the functional genes possessed by *M. tuberculosis*. Here, alternative pathways are lost and *M. leprae* has a very limited aerobic respiratory metabolism. It is possible for *M. leprae* to use glucose or glycerol or even pyruvate, catabolizing them to carbon dioxide to generate ATP oxidatively (Figure 1). However, the rate at which pyruvate is converted to acetyl-CoA to feed carbon into the oxidative Krebs cycle (Figure 1) is limited by a loss of redundancy. The *pdh* genes found in *M. tuberculosis* are lost in *M. leprae* and the function – the corresponding enzymatic activity – is markedly low.[10] This remaining function is apparently encoded by *aceE* and *lpd*, two separate genes that are ML1651 and ML2387, respectively. Disregarding the rate of catabolism of glucose and glycerol to carbon dioxide, their oxidation is predicted to generate only about 35% the ATP per mole carbon source oxidized that analogous pathways in *M. tuberculosis* generate, since all the NAD is predicted to be recycled in an oxygen-independent way (Figure 1). This failure to couple NADH oxidation to electron transport appears to be an adaptation to low oxygen tension yet the anaerobic respiratory systems seen in *M. tuberculosis* are all gone in *M. leprae*. Finally, the inability to recycle acetate (Figure 1) is predicted to lessen opportunities for axenic growth by limiting the number of carbon sources that *M. leprae* can use.[11]

Other notable differences between *M. leprae* and *M. tuberculosis* occur in the anaplerotic pathways in which 3-carbon (pyruvate, phospho-*enol*-pyruvate or PEP) and 4-carbon compounds (e.g. oxaloacetate, malate) can be interconverted to be used in either a catabolic or anabolic direction (Figure 1). Here, *M. leprae* lacks the *pca* gene. This is predicted to be required for pyruvate to be used in gluconeogenesis in groups of organisms that posses *pca*, such as actinomycetes and yeasts; thus another component of carbon metabolism is lacking in

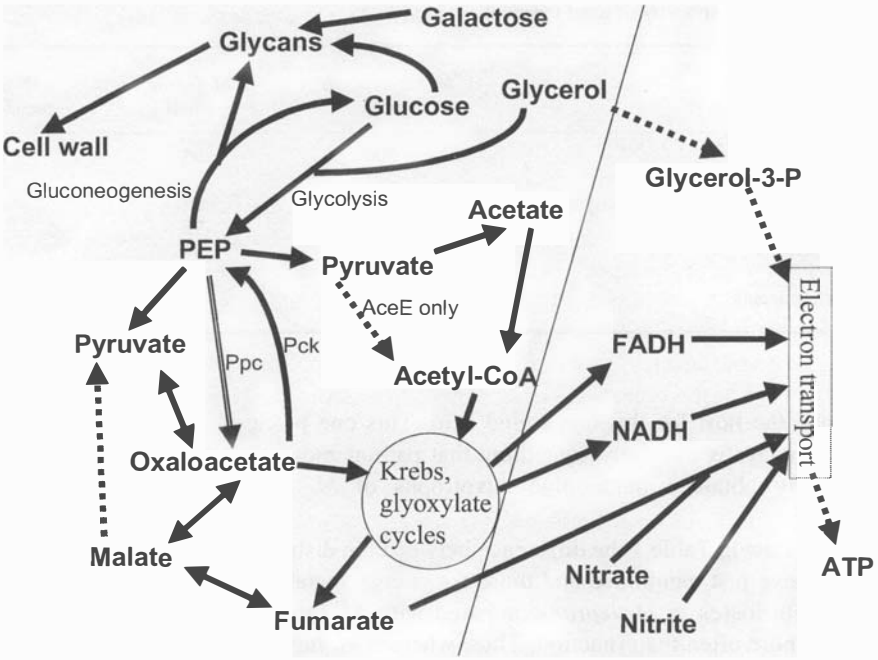

Figure 1. Carbon and energy metabolism in *M. leprae*; a comparison with a pathogen that can be cultured on simple media, *M. tuberculosis*. This simplified figure shows how glucose, glycerol, pyruvate, acetate and galactose can be metabolized in leprosy bacteria. *Black arrows* show reactions that can take place in both bacteria, gray ones only in *M. tuberculosis* and the *open arrow* shows the reaction that can only occur in *M. leprae*. *Dashed arrows* are for reactions predicted to occur at a limited rate. *Left of the line arrows* show interconversion of substrates, *right of the line arrows* show electron flow, in essentially respiratory pathways. The conversion of PEP (phospho-*enol*-pyruvate) to pyruvate is shown twice for clarity. For *M. leprae*, it follows that glucose and glycerol can be used for both ATP generation and gluconeogenesis to lead to cell wall intermediates as black arrows can be followed throughout the pathways. Pyruvate can only be used to generate ATP in the pathways shown in this figure, and acetate and galactose cannot be used at all. All substrates can be used in all directions by tubercle bacilli.

M. leprae. Intriguingly, *M. leprae* has a gene not present in *M. tuberculosis, ppc* (Figure 1). It is hard to predict the function of *ppc,* with certainty. Possibly it would allow oxaloacetate to be topped up so that the citric acid cycle could continue through dismutation of PEP (Figure 1). The effects of these differences should be ascertained by comparing the metabolic fate of pyruvate, glycerol, and perhaps other carbon sources (see Figure 1 for examples) in *M. leprae* and other actinomycetes.

A further enzyme in interconverting 3- and 4-carbon compounds is malic enzyme (decarboxylating), which converts malate to pyruvate. The dichotomy between fast and slow growing mycobacteria was once thought to be based, with few exceptions, on the presence or absence of this enzyme, with its activity found only in fast growers (mycobacteria that have a growth rate of above about 0.15 h^{-1}). Intriguingly the corresponding gene, *mez,* is actually found in *M. tuberculosis* (Rv2332), but is a pseudogene in *M. leprae*. Together with the apparent 'pinch-point' for conversion of pyruvate to acetyl-CoA mentioned above, this may reflect an irreversible commitment of *M. leprae* to a lifestyle characterized by slow growth and necessarily slow central metabolism.

The overall conclusion is that loss of alternative pathways in catabolism and also the loss

of 'redundant' genes (Box 1) has resulted in *M. leprae* occupying, and being severely restricted to a specialized niche. But there is worse to come!

Defence against toxic radicals

Intracellular pathogens such as *M. leprae* and *M. tuberculosis* have to protect themselves from exposure to toxic products produced by host macrophages. *M. leprae* lacks major components of its defences against the highly toxic and damaging oxygen- and nitrogen-free radicals. Against oxygen radicals there is no functional *katG* to encode catalase. This is evident from the genome sequence and has been proved biochemically[3] and genetically.[4,5] The *narGHJI* gene cluster,[12] which is lacking in *M. leprae,* may be a virulence factor, and could be involved in resistance against nitrogen oxide radicals. Their deletion, together with the lack of a functional *katG* to encode catalase as a defence against oxygen-free radicals, makes *M. leprae* look very vulnerable to the host's cell defence mechanisms. This reduction in the genes for respiration and to deal with the consequences of an oxygen-rich environment may therefore reflect a niche with relatively constant conditions (intracellular, little toxicity, a fairly constant microaerophilic environment) for *M. leprae*. In contrast, *M. tuberculosis* exists in a variety of tissues, from intracellular when it has been engulfed by alveolar macrophages, to caseated, probably almost anaerobic, lesions. Extreme specialization of the leprosy bacillus in terms of its oxygen environment may account for the deletion or inactivation in the *M. leprae* genome of 145 of the 170 remaining oxidoreductases and oxygenases found in *M. tuberculosis.*

Iron acquisition and metabolism

Classical studies show that the ability to acquire iron from the host is essential for the survival and growth of intracellular pathogens. Yet in comparison with *M. tuberculosis M. leprae* appears to have a major deficiency in its ability to acquire iron. Its entire *mbt* operon is deleted, rendering it unable to make either the membrane-associated or excreted form of mycobactin T. In *M. tuberculosis,* such a deletion would lead to loss of virulence as deletion of the *mbtB* gene results in restriction of growth in iron-limited media and in macrophages.[13] However, suspensions of *M. leprae* do not take up iron from the excreted mycobactin siderophores of slow-growing mycobacteria including *M. tuberculosis* and *Mycobacterium avium.* Instead, they acquire iron from the exochelin siderophores of *Mycobacterium neoaurum*, one of the most rapid-growing and least fastidious mycobacteria, and an unclassified, cultivable, armadillo-derived mycobacterium. Although they share the function of iron acquisition, these two types of siderophores are chemically distinct. While mycobactin T is derived from salicylate, the exochelins are hydroxymate-derived siderophores.[14] Thus, maybe it is the genes for a hydroxymate-based siderophore that should be searched for in the *M. leprae* genome. The genes for the biosynthesis, and probably the export and uptake of the exochelin of *M. smegmatis* are known.[15] The *fxb* genes for biosynthesis have no strong matches in the *M. leprae* or *M. tuberculosis* genomes even though one would predict that *M. leprae* would synthesize the class of siderophore that it can transport. However, there are important differences between the exochelins of *M. smegmatis* and *M. neoaurum* (for example the use of D-ornithine and N-methylation in the *M. noeaurum* in

contrast to L-ornithine and N-formylation in *M. smegmatis*) (see [14] for a comparison of the full structures) so all the similarity one might expect to see in the genes could be peptide synthases. Thus it is difficult to deduce whether *M. leprae* possesses functional genes for siderophore biosynthesis. However, it might be more feasible to search for genes involved in their transport. In *M. smegmatis, fxuD* is regarded as the gene encoding the ferri-siderophore receptor, but while this has 36% identity to the *M. tuberculosis fecB2* gene for an iron transport protein (Rv0265c), the corresponding gene (ML2548) in the *M. leprae* genome is a pseudogene. A cluster of three orfs immediately downstream of the *M. smegmatis* exochelin biosynthesis locus have strong similarity to clusters of three orfs (in both *M. tuberculosis* (Rv1458c to Rv1456c) and *M. leprae* (ML0589 to ML0591) but in both pathogens these appear neither to be under iron regulation nor flanked by genes likely to be involved in iron metabolism.

While the genes involved in iron acquisition by the leprosy bacillus remain obscure, there is no doubt that *M. leprae* does utilize iron. The presence of genes for cytochrome c (*ccsA,B)*, a haemoglobin-like oxygen carrier (*glbO),* ferredoxin (*fdxC,D*), biosynthesis of the haem group (*hem* genes) and for the iron-storage bacterioferritin *bfrA* (in one of the many examples of loss of redundancy, a second, *bfrB* gene present in *M. tuberculosis* is deleted from *M. leprae*) are all examples of genes encoding iron-containing molecules. Finally, IdeR, the key iron regulation protein dependent on intracellular iron, is highly conserved, with 84% identity to the *M. tuberculosis* IdeR. As in *M. tuberculosis*, *ideR* is just downstream of another important regulator, *sigB*, which is involved in response to stress and growth in macrophages. The capacity of *M. leprae* to tightly regulate intracellular iron would be anticipated to be vital since its *katG*, the gene that encodes the major catalase activity of pathogenic mycobacteria, is a pseudogene. Thus an excess of iron that would catalyse the formation of peroxides and oxygen-free radicals would be especially disastrous for the leprosy bacillus, and just as harmful to the bacterium as iron deficiency.

Do the genes found only in *M. leprae* explain its neurotropism?

Two distinctive properties of *M. leprae* are its ability to invade and grow in peripheral nerves, and its production of a unique phenolic glycolipid that forms a capsule around the bacillus. These properties are now thought to be linked, as the phenolic glycolipid binds to peripheral nerves. This lipid binds to the basal lamina of Schwann cell axons in a specific way that is inhibited by both (neural) laminin fragments and the trisaccharide moiety of the glycolipid.[16] Another product that is involved in binding to Schwann cells is the laminin binding protein, Lbp, though this is found in other mycobacteria.[17] The phenolic glycolipid is unique because it has a unique trisaccharide moiety, whereas other similar glycolipids, such as the one found in *M. bovis*, have a single sugar.[18] The genes for adding the additional sugars have not been identified yet,[4] but the 142 genes specific to *M. leprae* are not candidates since they are too short to encode for the required glycosyltransferases enzymes. However, some of the apparent gene fragments amongst the 142 specific genes could be large enough to encode neuroactive peptides.

Genes found in *M. leprae* but not in *M. tuberculosis*, for which it has been possible to attribute functions, are shown in Table 2. It is hard to implicate any of them in neurotropism but they do appear to be involved in uptake and metabolism. The adenyl cyclase and nramp homologues (ML2341 and ML2667, respectively) may suggest unusual regulatory functions,

Table 2. Genes with attributed functions in *M. leprae* but not in *M. tuberculosis*

Gene	Function	Paralogues in . . .
ML0397	ABC transporter involved in ribose transport?	*Bacillus subtilis*, other bacteria
ML0398	D-Ribose binding protein	*Bacillus subtilis*, other bacteria
ML0458	Possible oxidoreductase	Bacteria and eukaryotes, e.g. *Schizosaccharomyces pombe*
ML0578	Phospho-*enol*-pyruvate carboxylase (Ppc)	Bacteria, e.g. *Rhodopseudomonas* palustris and streptomycetes
Between ML1092 and ML1093	Similarities to malate oxidoreductases	Other bacteria
ML1553	ProS, prolyl-tRNA synthetase	Eukaryotic
ML1795	hsp18	Other actinomycete hsps, e.g. in *Mycobacterium intracellulare* and *Streptomyces albus*
ML2177	Uridine phosphorylase	Eukaryotic
ML2341	Adenylate cyclase – regulatory function	Many, e.g. *Stigmatella aurantiaca* and *Streptomyces coelicolor*
ML2667	Mn2+ transporter, similar to nramp family – regulatory function	Bacteria, e.g. *Pseudomonas aeruginosa*

perhaps related to a predilection of the leprosy bacillus for adenosine, a property predicted by classical biochemical studies.[19] A possible link that emerges from the *M. leprae* genome is that while most of the genes for making cobalamin and related vitamins have been lost in *M. leprae*, most of the few that remain use adenosine or its derivatives as one of their substrates. While the genes for making cobalamin are in *M. tuberculosis,* another of the genes found only in *M. leprae* (uridine phosphorylase) encodes an enzyme in pyrimidine nucleotide metabolism. There is something unusual about this whole area of purine and pyrimidine metabolism in *M. leprae.* Why should the biosynthetic pathways have been retained in such a host-dependent pathogen? Maybe it is part of the mechanism that allows the leprosy bacillus to survive and grow within the rather metabolically inert Schwann cells. Most leprosy bacilli that have been used for biochemical work were obtained from established infections in macrophages (spleens, lymph nodes and livers of experimentally infected armadillos[2]). In such established macrophage infections, biochemical studies showed that *M. leprae* only uses scavenging pathways typical of obligate parasites and not the *de novo* purine biosynthesic pathways[20] of which it is capable according to its genome. It is possible that the *de novo* purine biosynthetic pathway is required to establish infection if the early stages are analogous to *M. tuberculosis* infections in which *purC* mutants in the *de novo* pathway are cleared by innate immunity.[21]

In this section, I have raised a number of speculative ideas. However, it should be possible to test these ideas in the post-genomic era. It may be possible, for example to compare the transcripts of mRNA produced by leprosy bacilli in macrophages and in Schwann cells. Thus one could identify genes that are preferentially expressed during growth in Schwann cells and hence might account for this unusual property.

A few conclusions and many more questions

Obtaining a genome sequence should be regarded as a springboard to research on the microbe

in question. I hope in this short review I have shown that this is a time to push forward with research into *M. leprae*, a paradigm of obligate intracellular host-dependency. Fundamental questions for an intracellular pathogen about iron metabolism can be addressed. How the loss of redundancy throughout the genome in comparison with the tubercle bacilli has resulted in a specialized pathogen in contrast to the adaptable *M. tuberculosis* complex is another basic issue in mycobacteriology. Does the apparent limitation in virulence determinants and cell entry genes (Table 1) commit *M. leprae* to gaining access only to its narrow niche? While two of the three haemolysin genes found in *M. tuberculosis*, including *tlyA*, persist in *M. leprae*, all four *plc* genes are lost. Will we find out the role of PPEs and PEs in tubercle bacilli by making comparisons with *M. leprae*, a 'natural mutant' for most of them? What else can we learn about tubercle bacilli by comparative genomics now we have the leprosy genome? Why cannot *M. leprae* be grown axenically; do the lesions in energy metabolism only allow interrupted growth when conditions are just right in the host? Are media too toxic, at least in aerobic conditions? With a massive loss of regulatory functions[4] have those that would allow *M. leprae* to adapt to axenic culture been lost? We are now in a position to generate better defined hypotheses. Soon the *Mycobacterium ulcerans* genome will be sequenced; will comparisons with this difficult to grow mycobacterium help us to formulate new hypotheses for *M. leprae*? Finally, we can design experiments that will provide a better understanding of the interaction between *M. leprae* and Schwann cells. With neurological reactions still a major clinical issue in the treatment of leprosy these are urgent experiments. My conclusion would be to recommend the leprosy research community develops post-genomic research and investigates the expression of *M. leprae* genes as a means of addressing the many biological questions that still remain.

References

[1] Wheeler PR. Understanding the physiology of difficult, pathogenic bacteria from analysis of their genome sequences. *J Med Microbiol*, 2001; **51:** 1–4.

[2] Wheeler PR. Recent research into the physiology of *Mycobacterium leprae*. *Adv Microb Physiol*, 1990; **31:** 71–124.

[3] Wheeler PR, Gregory D. Superoxide dismutase, peroxidatic activity and catalase in *Mycobacterium leprae* purified from armadillo liver. *J Gen Microbiol*, 1980; **121:** 457–464.

[4] Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.

[5] Nakata N, Matsuoka M, Kashiwabara Y *et al.* Nucleotide sequence of the *Mycobacterium leprae* katG region. J Bacteriol, 1997; **179:** 3053–3057.

[6] Beck RW. *A chronology of microbiology in historical context.* American Society for Microbiology Press, Washington DC, USA, 2000.

[7] Levy L. Studies of the mouse foot pad technique for cultivation of *Mycobacterium leprae*. 3. Doubling time during logarithmic multiplication. *Lepr Rev*, 1976; **47:** 103–106.

[8] McAdam RA, Weisbrod TR, Martin J *et al. In vivo* growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect Immun*, 1995; **63:** 1004–1012.

[9] Smith DA, Parish T, Stoker NG, Bancroft GJ. Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates. *Infect Immun*, 2001; **69:** 1142–1150.

[10] Wheeler PR. Oxidation of carbon sources through the tricarboxylic acid cycle in *Mycobacterium leprae* grown in armadillo liver. *J Gen Microbiol*, 1984; **130:** 381–389.

[11] Wheeler PR, Ratledge C. Use of carbon sources for lipid biosynthesis in *Mycobacterium leprae:* a comparison with other pathogenic mycobacteria. *J Gen Microbiol*, 1988; **134:** 2111–2121.

[12] Weber I, Fritz C, Ruttkowski S *et al.* Anaerobic nitrate reductase (narGHJI) activity of *Mycobacterium bovis* BCG *in vitro* and its contribution to virulence in immunodeficient mice. *Mol Microbiol*, 2000; **35:** 1017–1025.

[13] De Voss JJ, Rutter K, Schroeder BG *et al.* The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc Natl Acad Sci USA*, 2000; **97:** 1252–1257.

[14] Ratledge C. Iron metabolism. In: Ratledge C, Dale J (eds) *Mycobacteria. Molecular biology and virulence.* Blackwell Science, Oxford, 1999, pp. 260–286.

[15] Yu S, Fiss E, Jacobs WR, Jr. Analysis of the exochelin locus in *Mycobacterium smegmatis:* biosynthesis genes have homology with genes of the peptide synthetase family. *J Bacteriol*, 1998; **180:** 4676–4685.
[16] Ng V, Zanazzi G, Timpl R *et al*. Role of the cell wall phenolic glycolipid-1 in the peripheral nerve predilection of *Mycobacterium leprae*. *Cell*, 2000; **103:** 511–524.
[17] de Melo Marques MA, Mahapatra S, Nandan D *et al*. Bacterial and host-derived cationic proteins bind alpha2-laminins and enhance *Mycobacterium leprae* attachment to human Schwann cells. *Microbes Infect*, 2000; **2:** 1407–1417.
[18] Dermateau-Ginsburg H, Lederer E. Sur la structure chemique du mycoside B. *Biochim Biophys Acta*, 1963; **70:** 442–451.
[19] Wheeler PR. Enzymes for purine synthesis and scavenging in pathogenic mycobacteria and their distribution in *Mycobacterium leprae*. *J Gen Microbiol*, 1987; **133:** 3013–3018.
[20] Wheeler PR. Biosynthesis and scavenging of purines by pathogenic mycobacteria including *Mycobacterium leprae*. *J Gen Microbiol*, 1987; **133:** 2999–3011.
[21] Jackson M, Phalen SW, Lagranderie M *et al*. Persistence and protective efficacy of a Mycobacterium tuberculosis auxotroph vaccine. *Infect Immun*, 1999; **67:** 2867–2873.

Lepr Rev (2001) **72,** 408–414

DNA metabolism in *Mycobacterium leprae*

STEPHANIE S. DAWES & VALERIE MIZRAHI*
MRC/SAIMR/WITS Molecular Mycobacteriology Research Unit, South African Institute for Medical Research and School of Pathology, University of the Witwatersrand, P.O. Box 1038, Johannesburg 2000, South Africa

Introduction

Understanding the molecular basis underlying the hallmark features of *Mycobacterium leprae*, such as its parasitism and extraordinarily slow growth rate, has stimulated research into the biology of this pathogen for decades, but it is through the completion of its genome sequence that a quantum leap of progress has been achieved.[1] In this review, we analyse the genes in *M. leprae* that are involved in the synthesis and salvage of purines and pyrimidines and in DNA replication and repair in an attempt to uncover the relationship between the massive gene decay observed in the *M. leprae* genome and its DNA metabolic capacity. This analysis has provided insights into possible mechanisms for the genomic deterioration in the leprosy bacillus and supplements the sparse biochemical data hard won from this organism.

Nucleotide biosynthesis and salvage

Purine and pyrimidine compounds are central to nucleic acid synthesis and byproducts of the purine pathway are essential components of energy metabolism and protein synthesis. Cellular levels of purines also affect induction of the stringent response and are important for the regulation of many other genes.

DE NOVO SYNTHESIS OF PURINES

Purine biosynthesis proceeds in a stepwise fashion that is conserved across prokaryotes and eukaryotes. A complete pathway for the production of purines has been preserved in *M. leprae* but it has a minimal gene complement for the conversion of ribose-5-phosphate to inosine monophosphate (IMP); it does not possess *purT* or *purU* (encoding the formate-dependent phosphoribosylglycinamide formyltransferase II and formyltetrahydrofolate

* Corresponding author: V. Mizrahi, MRC/SAIMR/WITS Molecular Mycobacteriology Research Unit, SAIMR, P.O. Box 1038, Johannesburg 2000, South Africa. Tel. +27 11 4899370; Fax +27 11 4899001; E-mail: 075val@chiron.wits.ac.za

 0305-7518/01/072408+07 $1.00

deformylase, respectively) which restricts the addition of the C-8 position to 5′-phosphoribosyl-1-glycinamide to the action of phosphoribosylglycinamide formyltransferase (*purN*). In contrast to *E. coli*, where phosphoribosylformylglycinamide synthase activity comprises PurL alone,[2] mycobacteria follow the *B. subtilis* model[3] with the *M. leprae* activity comprising PurL, PurQ and PurS (ML2219A).

For a genome so peppered with pseudogenes, it is surprising that *M. leprae* has retained three *guaB* genes (*guaB1-B3*) encoding IMP dehydrogenase, which converts IMP to xanthine monophosphate (XMP). By analogy with *Mycobacterium tuberculosis*,[4] GuaB2 probably provides most of the IMP dehydrogenase activity in *M. leprae*. A candidate for the *M. tuberculosis purR* regulator[5,6] is Rv3575c,[7] the counterpart of which is a pseudogene in *M. leprae* (ML0338), suggesting that regulation of this pathway may differ between mycobacterial species. The presence of a complete pathway for *de novo* purine biosynthesis deduced from the genome sequence is not consistent with the conclusion that *M. leprae* is dependent upon the host as a source for the purine ring.[8,9] However, regulatory impairment and low levels of constitutive gene expression might explain the observed purine auxotrophy of *M. leprae*. Levels of phosphoribosylamine glycine ligase (*purD*) and phosphoribosylaminoimidazole carboxylase ATPase (*purK*) in *M. leprae* were undetectable compared to adenylosuccinate lyase (*purB*).[10]

DE NOVO SYNTHESIS OF PYRIMIDINES

M. leprae also possesses a complete gene set for *de novo* pyrimidine synthesis, in accordance with experimental observations.[11,12] Although the biosynthetic components of the pathway are conserved with *M. tuberculosis*, the *M. leprae pyrR*, which would be responsible for UMP-dependent attenuation of the pyrimidine operon, is a pseudogene. The context of the biosynthetic genes is largely conserved between these two species, but the genes encoding the first three steps in the conversion of glutamine to dihydroorotate differ in their operonic structure, suggesting that this biosynthetic cluster may be differentially regulated.

BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM RIBONUCLEOTIDES

The reduction of ribonucleotides to deoxyribonucleotides is catalysed by ribonucleotide reductases (RNRs), which are classified on the basis of their O_2 dependence. The O_2 dependent, class I family of RNRs is divided into subclasses I(a) and I(b), represented by NrdAB and NrdEF respectively, which utilize NDPs as substrates. Class II RNRs (NrdJ) are indifferent to O_2, catalyse the reduction of NTPs and utilize adenosylcobalamin as a co-factor, and the anaerobic class III (NrdDG) enzymes generate a glycyl radical from S-adenosylmethionine and an iron-sulfur cluster to reduce NTPs. The RNR gene repertoire of *M. leprae* is simpler than that of *M. tuberculosis*,[4] consisting only of *nrdE* and *nrdF*. As in *M. tuberculosis*, *nrdE* is operonic with *nrdH* and *nrdI*, which encode the electron transport component of NrdEF and a protein with stimulatory properties for ribonucleotide reduction, respectively.[13] dADP, dGDP and dCDP are derived directly by reduction of the corresponding NDPs and all other enzymes involved in the production of dNTPs are conserved (nucleotide diphosphate kinase, dUTPase and thymidylate synthase).

SALVAGE PATHWAYS

Salvage pathways control the intracellular levels of nucleosides and bases produced from catabolism of nucleic acids and nucleotides and those taken up from the extracellular milieu.

Analysis of the genome sequence suggests that the purine salvage pathway of *M. leprae* is largely intact, whereas the pyrimidine pathway is fragmented, consistent with experimental evidence suggesting that salvage of purines plays a more important role than that of pyrimidines.[14] The uptake and utilization of adenosine and hypoxanthine has been demonstrated in macrophage-maintained leprosy bacilli,[15] and dephosphorylation of nucleotides and subsequent importation of nucleosides has also been shown to occur in *M. leprae*.[8,14]

The purine salvage pathway converting adenosine to AMP via adenine by *pnp* or *deoD* (purine nucleoside phosphorylases) and *apt* (adenine phosphoribosyltransferase) is not complete in *M. leprae*; although the *pnp-apt* pathway is probably functional in *M. tuberculosis, pnp* and *apt* are pseudogenes in *M. leprae*. ML0707 is annotated as *deoD*, a guanosine/inosine/adenosine phosphorylase of the PNP family I of phosphorylases, but domain analysis indicates that it belongs instead to the PNP family II and would therefore catalyse the degradation of all purine nucleosides except adenosine and deoxyadenosine. However, conversion of adenosine to adenine is a minor pathway in *E. coli* and the major pathway, which cycles adenosine to IMP is conserved in *M. leprae* (*add*, ML0707 and *hpt*). Guanosine can be cycled to guanosine monophosphate (GMP) in a similar fashion (ML0707, *hpt*). In *Escherichia coli* and *Bacillus subtilis*, xanthine and guanine are only cycled to IMP via GMP. Neither xanthine phosphoribosyltransferase (*xpt/gpt*), which recycles xanthine to XMP, nor GMP reductase (*guaC*) was identified in the *M. leprae* genome, but since IMP dehydrogenase and GMP reductase are closely related enzymes with similar tertiary structures, GuaB1 and/or GuaB2 may function as GMP reductase. Finally, although high levels of adenosine kinase activity were detected in *M. leprae*,[10] we could not correlate a coding region(s) with this activity, suggesting that a non-orthologous enzyme may be present.

M. leprae has been shown to incorporate exogenous pyrimidines as bases or nucleosides, with uracil alone being able to supply all of its pyrimidine requirements.[14] Incorporation of ^{3}H-thymidine has also been reported in *M. leprae* isolates.[15] However, we could not identify genes for thymidine kinase (*tdk*), thymidine phosphorylase (*deoA*) or uridine/cytidine kinase (*udk*). Moreover, the genes encoding uracil phosphoribosyltransferase (*upp*, *pyrR*) are non-functional, which would preclude the direct conversion of uracil to UMP. *M. leprae* has also acquired a eukaryotic-like uridine phosphorylase (ML2177) that is not found in *M. tuberculosis*, for the interconversion of uracil and uridine (and possibly also thymine and thymidine, and uracil and deoxyuridine). Cytidine may also be converted to uridine by cytidine deaminase (*cdd*) and deoxycytidine to deoxyuridine by deoxyctidine deaminase (ML2507). In the absence of an enzyme to produce UMP from uracil or uridine, it is surprising that an intact *cdd* gene still remains, as this would shuttle cytidine to a metabolic dead-end. These observations, in conjunction with experimental evidence,[14,15] argue in favour of the presence of non-orthologous uridine and/or thymidine kinases.

DNA replication

The highlights of our comparative analysis of the DNA replication and repair gene complements of *M. leprae* and *M. tuberculosis* are summarized below.

INITIATION/ELONGATION/PROCESSING PROTEINS

Proteins involved in initiation (DnaA), accessory proteins (mIHF, HU, RNA-P), those in the pre-priming complex (SSB, DnaB), DnaG primase and DNA gyrase are all conserved

in *M. leprae*, but DnaC is absent, suggesting that the replicative complex does not require DnaC to load DnaB. Like *M. tuberculosis*, *M. leprae* lacks RNase HI, suggesting that TopA and RNase HII provide initiation specificity. The RNA primers for replication could be processed by the 5′–3′ exonuclease activity of PolI, but as previously noted, the RLEP elements flanking the *polA* gene may affect its expression and hence compromise its processing function.[16]

REPLICATIVE COMPLEX

Although most information on the bacterial replication complex (DNA PolIII holoenzyme) has been gained from studies in *E. coli*,[17] recent work has elucidated the subunit composition of a Gram-positive PolIII[18] against which that of *M. leprae* can be compared. *M. leprae* PolIII comprises τ/γ (*dnaZX*), δ' (*holB*) and δ (ML0603) subunits. As in *Streptomyces coelicolor*, the polymerase activity of *M. leprae* PolIII is provided by the Gram-negative type α subunit encoded by *dnaE1*.[19] This structure differs from that of low G+C Gram positive organisms, which contain a PolC subunit.[18] Mycobacteria also possess a second, Gram-negative type α subunit gene (*dnaE2*), but in *M. leprae*, this is a pseudogene. Studies in *E. coli* have revealed that in the core holoenzyme, α is tightly associated with ϵ, the *dnaQ*-encoded proofreading subunit, which serves to markedly increase the fidelity of replication.[20,21] In addition to its 3′–5′ exonuclease activity, ϵ is required for maximal DNA synthesis[22] and also plays an important structural role within the core.[23] We therefore consider the fact that *M. leprae dnaQ* is a pseudogene to be highly significant.[4] Apart from less efficient (slower) polymerase action, the absence of a proofreading subunit of the replicative polymerase has major implications for the maintenance of genome integrity particularly as *M. leprae* also lacks the genes involved in *mutLS*-based mismatch repair.[4] The absence of two of the three pillars of replication fidelity in *M. leprae* places the full burden of fidelity on the α subunit, which led us to speculate that enhanced base selectivity by α might be the only mechanism available to avoid error catastrophe[21,24] in this organism.[4]

The loss of *dnaQ* may be reflected in the astonishing number of base substitutions and frameshifts in the *M. leprae* genome. *M. leprae* isolates also display variable numbers of TTC repeats that are not found in *M. tuberculosis* or *Mycobacterium avium*.[25] Since expansion and contraction of triplet repeats map to mutations in *dnaQ*,[26] this form of genomic diversification argues against the existence of an alternate proofreading subunit in *M. leprae*. This conclusion is also consistent with the absence of intact 3' exonuclease-encoding genes that could serve this function. Finally, as replication in *M. leprae* may be inherently inefficient and error-prone, it is not surprising that this organism has also dispensed with its UmuC-like, error-prone polymerases.

DNA repair

As in other metabolic processes, considerable decay is evident in the repertoire of DNA repair genes in *M. leprae*. However, as highlighted below, the decay has mainly served to reduce the number of genes encoding redundant or overlapping biochemical functions while leaving the major repair pathways intact.

DAMAGE REVERSAL

Although *M. tuberculosis* is well equipped to reverse bases damaged by O-alkylation and pyrimidine dimers, *M. leprae* has retained only the suicidal DNA methyltransferase, Ogt, which is required for mutation avoidance.

MISMATCH REPAIR

All mycobacterial genomes sequenced to date lack the highly conserved, *mutLS*-based mismatch repair pathway,[4,27] implying that as in *M. tuberculosis*, base substitution and simple frameshift mutations may become fixed in *M. leprae* due to the lack of strand discrimination inherent in repair by enzymes such as MutY and UvrABC.[27]

EXCISION REPAIR

M. leprae possesses at least one copy of each of the genes required for the base excision repair (BER) of lesions created by hydrolytic, alkylative and oxidative and damage, namely uracil DNA glycosylases (*ung* and possibly ML1105), 3-methylpurine DNA glycosylases (*tagA*, ML1351), glycosylase/lyases (*fpg*, *nth*) and an apyrimidinic/apurinic endonuclease (*nfo*), but its restricted gene repertoire reflects the functional redundancy of BER activities. As such, *M. leprae* possesses neither *nei* nor *xthA*, but as these genes play minor roles in the presence of functional *fpg*, *nth* and *nfo*,[28,29] their absence is tolerable. Both *M. tuberculosis* and *M. leprae* possess AlkB (ML0190), which acts on alkylated single-stranded (ss) DNA at replication forks.[30] Since AlkB is not widely distributed amongst the genomes sequenced to date, its presence suggests that mycobacteria may be particularly prone to alkylation damage.[31]

M. leprae also possesses a complete gene complement for nucleotide excision repair (NER) of bulky lesions and adducts that cannot be repaired by BER. The NER process is initiated by UvrABC, which recognizes and binds to the lesions, with UvrD and PolI being responsible for further processing.

RECOMBINATIONAL REPAIR

The recombinational proficiency of *M. leprae* is evident in the structure of its genome, which shows that multiple homologous recombination events have occurred between repetitive sequences.[1] It is therefore not surprising that *M. leprae* contains a complete set of genes for homologous recombination. As in other mycobacteria, the *recA* gene, which plays a pivotal role in the recombinational repair of strand breaks, is probably *lexA*-regulated and damage-inducible in *M. leprae*.[32–34] The function of *M. tuberculosis* RecA is well described[35] and *M. leprae* RecA is assumed to have a similar activity. In *E. coli*, a number of pathways exist for the initial processing of double-stranded DNA breaks to ss substrates for recombination, each featuring the action of exonucleases and helicases. In *E. coli*, 3′ invasive ends for recombination can be generated by RecBCD or by the exonuclease/helicase pairs, RecE/RecT or RecQ/RecJ. Mycobacteria do not contain *recE*, *recT*, *recO* or *recJ*, and are unique amongst the Gram-positive organisms sequenced thus far in that they possess a Gram-negative type RecBCD, rather than the analogous AddAB found in low G + C Gram-positives.[36] *M. leprae* possesses neither of these systems, but it does possess homologues of

an archaeal exonuclease (ML1155) and helicase (ML1312) belonging to the *recB* family of exonucleases/helicases,[37] in addition to several other helicases (*helY*, ML2157 and ML1624) and exonucleases (*sbcD* (ML1119), *xseAB*) which could provide the break-processing function.

In *E. coli,* only one pathway has been shown to process the resulting ss DNA substrate further, and this RecFOR pathway is conserved in *M. leprae*. Helicase II (UvrD) is proposed to play an important part in RecBCD$^-$ cells in the RecF pathway of recombination, and in this context it is notable that *M. leprae* has retained both *uvrD* and *uvrD2* genes. Strand exchange is subsequently promoted by the actions of TopA and DNA Pol I, both of which are present in *M. leprae,* and as in other organisms, RecA, together with RuvAB and RecG would carry out branch migration and RuvC and DNA ligase would finally resolve the heteroduplex.

Conclusions

While acknowledging the limitations inherent in deducing gene function from 'sequence gazing', our analysis has nonetheless provided some intriguing insights into the genetics of DNA metabolism in *M. leprae*, which could guide future studies. High on our list of priorities would be to investigate the functional relationship between defects in components of replication fidelity and the process of gene decay.

Acknowledgements

We thank Digby Warner and Steven Durbach for critically reviewing the manuscript. VM is supported by an International Scholars grant from the Howard Hughes Medical Institute.

References

[1] Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.
[2] Sampei G, Mizobuchi K. The organization of the *purL* gene encoding 5′-phosphoribosylformylglycinamide amidotransferase of *Escherichia coli. J Biol Chem*, 1989; **264:** 21230–21238.
[3] Saxild HH, Nygaard P. The *yexA* gene product is required for phosphoribosylformyl-glycinamidine synthetase activity in *Bacillus subtilis*. *Microbiology*, 2000; **146:** 807–814.
[4] Mizrahi V, SS Dawes, Rubin H. DNA Replication. In: Hatfull GF, Jacobs WR Jr (eds) *Molecular genetics of mycobacteria.* ASM Press, Washington, D.C., 2000, pp. 159–172.
[5] He B, Smith JM, Zalkin H. *Escherichia coli purB* gene: cloning, nucleotide sequence, and regulation by *purR*. *J Bacteriol*, 1992; **174:** 130–136.
[6] Weng M, Nagy PL, Zalkin H. Identification of the *Bacillus subtilis pur* operon repressor. *Proc Natl Acad Sci USA*, 1995; **92:** 7455–7459.
[7] Cole ST, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998; **393:** 537–544.
[8] Wheeler PR. Biosynthesis and scavenging of purines by pathogenic mycobacteria including *Mycobacterium leprae*. *J Gen Microbiol*, 1987; **133:** 2999–3011.
[9] Wheeler PR. Biosynthetic pathways in *Mycobacterium leprae*. *Acta Leprol*, 1989; **7:** 21–24.
[10] Wheeler PR. Enzymes for purine synthesis and scavenging in pathogenic mycobacteria and their distribution in *Mycobacterium leprae*. *J Gen Microbiol*, 1987; **133:** 3013–3018.
[11] Wheeler PR. Pyrimidine biosynthesis in *Mycobacterium leprae* and other intracellular mycobacteria. *Acta Leprol*, 1989; **7:** 33–35.
[12] Wheeler PR. Pyrimidine biosynthesis de novo in *M. leprae. FEMS Microbiol Lett*, 1989; **48:** 185–189.

[13] Jordan A, Aslund F, Pontis E *et al.* Characterization of *Escherichia coli nrdH*. A glutaredoxin-like protein with a thioredoxin-like activity profile. *J Biol Chem*, 1997; **272:** 18044–18050.
[14] Wheeler PR. Pyrimidine scavenging by *Mycobacterium leprae*. *FEMS Microbiol Lett*, 1989; **48:** 179–184.
[15] Harshan KV, Mittal A, Prasad HK *et al.* Uptake of purine and pyrimidine nucleosides by macrophage-resident *Mycobacterium leprae:* ^{3}H-adenosine as an indicator of viability and antimicrobial activity. *Int J Lepr Other Mycobact Dis*, 1990; **58:** 526–533.
[16] Gordhan BG, Mizrahi V. The RLEP-flanked *polA* gene from *Mycobacterium leprae* is not transcribed in *Mycobacterium smegmatis*. *Gene*, 1997; **187:** 63–66.
[17] Kelman Z, O'Donnell M. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem*, 1995; **64:** 171–200.
[18] Bruck I, O'Donnell M. The DNA replication machine of a gram-positive organism. *J Biol Chem*, 2000; **275:** 28971–28983.
[19] Flett F, de Mello Jungmann-Campello D, Mersinias V *et al.* A 'gram-negative-type' DNA polymerase III is essential for replication of the linear chromosome of *Streptomyces coelicolor* A3(2). *Mol Microbiol*, 1999; **31:** 949–958.
[20] Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J Biol Chem*, 1993; **268:** 23762–23765.
[21] Fijalkowska IJ, Schaaper RM. Mutants in the Exo I motif of *Escherichia coli dnaQ:* defective proofreading and inviability due to error catastrophe. *Proc Natl Acad Sci USA*, 1996; **93:** 2856–2861.
[22] Kim DR, McHenry CS. In vivo assembly of overproduced DNA polymerase III. Overproduction, purification, and characterization of the alpha, alpha-epsilon, and alpha-epsilon-theta subunits. *J Biol Chem*, 1996; **271:** 20681–20689.
[23] Taft-Benz SA, Schaaper RM. Mutational analysis of the 3′-5′ proofreading exonuclease of DNA polymerase III. *Nucleic Acids Res*, 1998; **26:** 4005–4011.
[24] Fijalkowska IJ, Schaaper RM. Antimutator mutations in the alpha subunit of *Escherichia coli* DNA polymerase III: identification of the responsible mutations and alignment with other DNA polymerases. *Genetics*, 1993; **134:** 1039–1344.
[25] Shin YC, Lee H, Lee H *et al.* Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *J Clin Microbiol*, 2000; **38:** 4535–4538.
[26] Iyer RR, Pluciennik A, Rosche WA *et al.* DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in *Escherichia coli*. *J Biol Chem*, 2000; **275:** 2174–2184.
[27] Mizrahi V, Andersen SJ. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol*, 1998; **29:** 1331–1339.
[28] Jiang D, Hatahet Z, Blaisdell JO *et al.* *Escherichia coli* endonuclease VIII: cloning, sequencing, and over-expression of the *nei* structural gene and characterization of *nei* and *nei nth* mutants. *J Bacteriol*, 1997; **179:** 3773–3782.
[29] Laval J, Jurado J, Saparbaev M, Sidorkina O. Antimutagenic role of base-excision repair enzymes upon free radical-induced DNA damage. *Mutat Res*, 1998; **402:** 93–102.
[30] Dinglay S, Trewick SC, Lindahl T, Sedgwick B. Defective processing of methylated single-stranded DNA by *E. coli AlkB* mutants. *Genes Dev*, 2000; **14:** 2097–2105.
[31] Boshoff HIM, Durbach SI, Mizrahi V. DNA metabolism in *Mycobacterium tuberculosis:* implications for drug resistance and strain variability. *Scand J Infect Dis*, 2001; **33:** 101–105.
[32] Durbach SI, Andersen SJ, Mizrahi V. SOS induction in mycobacteria: analysis of the DNA binding activity of a LexA-like repressor and its role in DNA damage induction of the *recA* gene from *Mycobacterium smegmatis*. *Mol Microbiol*, 1997; **26:** 643–653.
[33] Movahedzadeh F, Colston MJ, Davis EO. Determination of DNA sequences required for regulated *Mycobacterium tuberculosis* RecA expression in response to DNA-damaging agents suggests that two modes of regulation exist. *J Bacteriol*, 1997; **179:** 3509–3518.
[34] Brooks PC, Movahedzadeh F, Davis EO. Identification of some DNA damage-inducible genes of *Mycobacterium tuberculosis:* apparent lack of correlation with LexA binding. *J Bacteriol*, 2001; **183:** 4459–4467.
[35] Colston MJ, Davis EO, Papavinasasundaram KG. General recombination. In: Hatfull GF, Jacobs WR Jr (eds) *Molecular genetics of mycobacteria.* ASM Press, Washington D.C., 2000, pp. 85–91.
[36] Kooistra J, Haijema BJ, Venema G. The *Bacillus subtilis addAB* genes are fully functional in *Escherichia coli*. *Mol Microbiol*, 1993; **7:** 915–923.
[37] Aravind L, Walker DR, Koonin EV. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res*, 1999; **27:** 1223–1242.

Lepr Rev (2001) **72,** 415–428

Genomic evidence for the retention of the essential mycobacterial cell wall in the otherwise defective *Mycobacterium leprae*

PATRICK J. BRENNAN & VARALAKSHMI D. VISSA
Department of Microbiology, Colorado State University, Fort Collins, CO 80523, USA

Summary The obligate intracellularism of *Mycobacterium leprae* may be attributable to the effects of mutations in major metabolic areas due to a genome capable of encoding only about 1600 proteins. Yet cell wall biosynthesis capability remains relatively intact and comparisons with the genome of *Mycobacterium tuberculosis* provide insights into the genetic basis of a minimal mycobacterial cell wall.

Abbreviations

ACP:acyl carrier protein; AG: arabinogalactan; Ara*f*: arabinofuranose; C_{50}-P: decaprenyl 5-phosphate; CoA: coenzyme A; DAP: meso-diaminopimelic acid; DMAPP: dimethylallyl diphosphate; DPA: decaprenyl phosphoarabinose; DXP: deoxyxylulose 5-phosphate; FPP: farnesyl diphosphate; Gal*f*: galactofuranose; GAP: glyceraldehyde 3-phosphate; GDP: guanosine 5' diphosphate; GlcNAc: *N*-acetyl glucosamine; GPP: geranyl diphosphate; IPP: isopentenyl diphosphate; LAM: lipoarabinomannan; LM: lipomannan; mAGP: mycolyl-arabinogalactan-peptidoglycan; Mal: malonyl; Man: mannose; Me: methyl; MurNAc: *N*-acetylmuramic acid; MurNGly: *N*-glycolylmuramic acid; P: phosphate; PG: peptidoglycan; PDIM: phthiocerol dimycocerosate; PGL: phenolic glycolipids; PIMs: phosphoinositol mannosides; PAPP: 5' phospho arabinofuranosyl pyrophosphate; PRPP: 5' phospho ribosyl pyrophosphate; Rha: rhamnose; TDM: trehalose dimycolate; TDP: thymidine 5'-diphosphate; TMM: trehalose monomycolate; UDP: uridine 5′-diphosphate

Introduction

In the last decade of biological research dominated by the pursuit of genomic sequence of organisms, the genus *Mycobacterium* has not been left behind. In fact, several species of mycobacteria have been or are currently being sequenced to aid in the prevention and treatment of diseases such as leprosy and tuberculosis in humans and Johne's disease and tuberculosis in cattle.[1–3] The sequences of the virulent *M. tuberculosis* H37Rv, avirulent fast growing *M. smegmatis* and that of the slow-growing *M. leprae* are a valuable data set for comparative studies on physiology and virulence of mycobacteria. Unraveling the genetics of

Correspondence: P. J. Brennan (e-mail: Patrick.Brennan@ColoState.edu)

0305-7518/98/064053+07 $1.00

essential biosynthetic pathways is the preferred approach for identifying new dug targets and has already yielded several candidates.[4] The sequencing of the *M. leprae* bacterium is especially significant to medicine because it gives the genetic blue print of an organism yet to be cultivated in a laboratory. The sequence has revealed 'a decaying genome' with a dramatic loss of functional capacity by mutations that gave rise to 1116 pseudogenes in every aspect of central and intermediary metabolism resulting in the obligate *in vivo* and intracellular habitat of *M. leprae* [1,5](see also 'The decaying genome of *Mycobacterium leprae*', this issue). The very long doubling time of 11–12 days[6] and the preference for a cooler environment (30°C)[7] may also be due the presence of temperature sensitive mutations in some of the remaining 1600 or so genes. Nevertheless, *M. leprae* can sustain growth and cause disease. The survival of mycobacteria under unfavorable conditions has been attributed to the characteristic permeability barrier of the cell envelope and its role in infection of host macrophages and modulation of immune responses.[8] The properties of the cell envelope are particularly relevant for *M. leprae*, because the organism may depend on the host for several nutrients due to defects in the synthesis of methionine, cysteine, purine rings, and uptake of several amino acids, ions and sugars[1,5] (see also 'The microbial physiologist's guide to the leprosy genome', this issue). In this report we have therefore analyzed the impact of the genome down sizing and decay on the cell wall of *M. leprae*. We have compared the genomes of *M. leprae* and *M. tuberculosis* and assume that genes for the biosynthesis of similar molecules are orthologous (same gene in different genomes) or at least homologous (similar gene) and those genes that are present in *M. tuberculosis*, but absent or pseudogenes in *M. leprae* are probably not involved in these syntheses or are redundant (spare).

We refer the readers to the reviews by Brennan and Nikaido[8], McNeil *et al.*[9], Chatterjee and Khoo[10], Barry *et al.*[11] Baulard *et al.*[12], Belanger and Inamine[13] and Crick *et al.*[14] for details of the structure, function, biosynthesis and distribution among mycobacteria of these envelope components and only provide a brief outline in this review. Instead we have focused primarily on gene assignments for *M. leprae* in relation to those for *M. tuberculosis* based on the completed and annotated genome sequences.

Morphology of *M. leprae* and the ultrastructure of the cell wall

M. leprae is a strongly acid fast staining rod 1–8 μm long and 0.3 μm in diameter and thus does not differ remarkably from *M. tuberculosis*. Studies of the ultrastructure of *M. leprae*, in sections and as whole bacteria from man, mouse and armadillo, have been extensive[15], but have not shown any gross unique features compared to other mycobacteria. However, Draper[16] has described three ultrastructural features of the cell walls that may be characteristic of *M. leprae*: aberrant morphology, wall bands and paracrystalline bodies. He had observed departure from the classical cylindrical shape of a bacillus in suspensions prepared from armadillos, in that cells typically have a tapered or double-tapered shape with hemispherical ends, which he attributed to a defect in the normal process of cell wall construction. Wall bands first observed by Nishiura *et al.* are described as 'circumferential ridges on the outer surface of the cell', very numerous and positioned at random along the length of the cell.[17] These may be scars left when the cell wall separated during the division process and their randomness may reflect a defect in the cell wall construction. The paracrystalline, quasi-crystalline bodies seen in sectioned *M. leprae*[18] probably correspond to the capsular matrices and foamy structures responsible for binding hundreds of bacilli into

'clumps' or globi and into smaller clumps where the individual cells occur in parallel arrays, the noted 'bundles of cigars'.

Biochemical structure and composition

Current knowledge on the biosynthesis and genetics of several components of the mycobacterial cell wall that are described below has evolved primarily from studies with *M. smegmatis, M. tuberculosis*, *M. bovis BCG* and *M. avium* using a combination of methods including chemical and structural analysis, metabolic labeling, cell free assay systems, isolation and characterization of naturally occurring variants or mutants and, more recently, genetic manipulation such as mutagenesis of the genome and recombinant gene expression. However, such opportunities are limiting for *M. leprae* research due to the inability to cultivate the organism *in vitro* and the lack of proven genetic tools. Despite these limitations, sufficient information was gathered on the chemical and structural composition using small amounts of cells obtained from animal or human sources to conclude that the basic architecture of the cell wall is the covalently linked peptidoglycan-arabinogalactan-mycolic acids complex (mAGP) seen amongst all mycobacteria and the related corynebacteria and nocardia, except for few modifications.[19,20]

The peptidoglycan (PG) of *M. leprae* is characteristic of the chemotype IV group that includes mycobacteria, corynebacteria and nocardia because they contain *meso*-diaminopimelic acid (DAP), in the peptide chains.[8] The muramic acids of the sugar backbone are modified with *N*-glycolyl rather than *N*-acetyl groups in mycobacteria and nocardia.[8] A feature unique to *M. leprae* is the substitution of L-ala with glycine in the peptide of peptidoglycan.[21] The effect of this change on the physical properties of the peptidoglycan is not known.

Arabinogalactan (AG) is a polymer of furanose sugars of galactose and arabinose, not found in humans.[22] Typically, a homogalactan (~30 units in *M. tuberculosis*) composed of alternating 5 and 6 linked βD-Gal$_f$ residues is linked to the peptidoglycan via a disaccharide bridge (-L-Rha-D-GlcNAc-P-) called the linker unit (LU).[23] Three branches of 5-linked arabinan are attached near the reducing end of the galactan. The arabinan is composed of 5-linked Ara$_f$, which further branch (3- and 5-linked Ara$_f$ -α). The non reducing ends are composed of the hexaarabinofuranosyl motif [βD-Ara$_f$ -(1→2)-α-D-Ara$_f$]$_2$-3,5-α-D-Ara$_f$-(1→5)-α-D-Ara$_f$ (Ara$_6$).[24] All these major motifs are identical in *M. tuberculosis* and *M. leprae*, with the exception that *M. leprae* has 40–50% fewer galactan residues.[20] Two-thirds of the terminal arabinoses of the arabinan chains are esterified with mycolic acids in *M. tuberculosis*.[25] The extent in *M. leprae* has not been determined.

The α-alkyl branch of the α-alkyl, β-hydroxy fatty acids called mycolic acids which range from C_{14} to C_{26}, is C_{20} in *M. leprae*.[11] The β-hydroxy (meromycolate) chain is often modified with double bonds (cis and trans), cyclopropane, methyl, epoxy, keto, and methoxy groups that render flexibility (fluidity) to the wall. *M. leprae* does not have methoxymycolates[26] due to the lack of a functional *mmaA3* gene, as demonstrated in *Mycobacterium bovis* BCG (Pasteur).[27] It appears that ketomycolates have a more specific role for growth in macrophages in *M. tuberculosis* as their abundance increases 5-fold *in vivo*, and under low oxygen tensions *in vitro* while the methoxymycolates decrease 2-fold. The absence of ketomycolates reduces ability to survive in macrophage like cell lines.[28] Therefore, the lack of methoxymycolates in *M. leprae* may not impair viability in macrophages.

In addition to the mycolates esterified to AG, mycolic acids are also present in the extractable lipids as esters of trehalose: 6-*O*-mycolyl and 6, 6'-O-dimycolyl trehalose (TMM and TDM respectively). Small amounts of TMM but not TDM were identified in *M. leprae*.[29]

The wall of *M. leprae* is also endowed with an unusually abundant proportion of the extractable intercalated lipoglycans, phosphatidylinositol mannosides (PIMs), lipomannans (LM) and lipoarabinomannans (LAM) characterized in several mycobacteria which may be anchored in the plasma membrane via the acyl chains (tuberculostearic and palmitic acid) of phosphatidyl inositol (PI).[30] LM and LAM are made up of linear α-1→6 linked mannan chain originating from PIM_2 in which each mannose is further branched with α-1→2 D-mannose.[31] LAM is a heterogeneous macromolecule arising from LM that contains arabinan branches similar in composition and structure to that of AG that may be terminated with 'caps' of variable numbers of mannose or inositol residues.[10,32] The LAM of *M. leprae* has fewer Ara_6 termini, and a lesser degree of mannose capping than that of *M. tuberculosis*.

However, there are other solvent extractable components[8] such as glycopeptidolipids (GPLs) typical of *M. avium* spp that define serovar specificity and colony morphology; trehalose based lipids such as acylated trehaloses (containing straight chain, mycerocerosic acids, mycolipanolic and mycolipenic fatty acids) and sulpholipids (trehalose 2' sulphate acylated with phthioceranic and hydroxyphthioceranic acids) present in strains of *M. tuberculosis*; and lipooligosaccharides (LOSs) isolated from several species are absent in *M. leprae*. Instead, the dominant lipid is the phenolic glycolipid PGL-1 [33], a glycosylated derivative of the phenolphthiocerol dimycocerosate. The trisaccharide-of PGL-1 of *M. leprae* is β-D-3, 6, di-*O*-methyl Glu (1→4)-α-2,3-di-*O*-methyl-L-Rha-(1→2)-α-3-*O*-methyl-L-Rha and is highly antigenic. Synthetic glycoconjugates containing this trisaccharide are sensitive tools for serodiagnostics of leprosy.[34] Recently the trisaccharide was shown to be involved in the specific interaction of *M. leprae* with the laminin of Schwann cells.[35] This discovery is an important step towards the identification of a mechanism for entry of *M. leprae* into nerve cells that can initiate the subsequent nerve damage that is the hallmark of leprosy. *M. tuberculosis* H37Rv contains only phthiocerol dimycocerosates.[8]

Biosynthesis and genetics of cell wall (envelope) in *M. leprae*: insights from the genome sequence

The biosynthetic pathway for the individual components and their assembly to form the mAGP complex, as deciphered from the approaches mentioned before, is depicted in Figure 1. In the recent review by Crick *et al.*[14], only the 13 *M. tuberculosis* genes for AGP synthesis that have been functionally characterized have been highlighted. Eleven of these were identified by the 'cloning by homology' approach. In Table 1, we show the homologs for these genes in *M. leprae*. We predict that these genes will be functional in *M. leprae* without the need for their functional characterization as the homology is high, and genetic context is similar. Furthermore, we have included putative genes for reactions not yet characterized in mycobacteria, by finding homologs for known genes in other organisms using the BLAST algorithm.[36] The genes for the mycolic acid and phenolphthiocerol dimycocerosate synthesis are also included. In the following section, some of the biosynthetic pathways are described in brief. The common names of the genes involved are indicated in italic font in parenthesis.

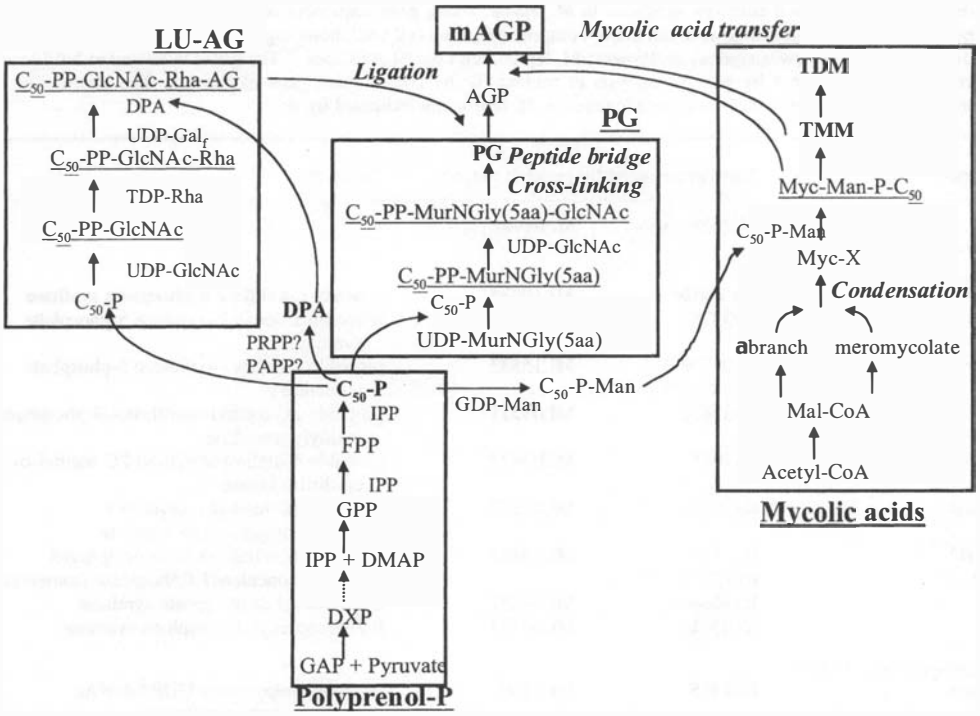

Figure 1. Biosynthesis of the mycolic acid-arabinogalactan-peptidoglycan complex (mAGP) of the cell wall of mycobacteria. The pathways for the synthesis of the individual components and the stages at which they are assembled to form the mAGP complex are shown schematically. The genes involved in these reactions and in the synthesis of the sugar donors are listed in Table 1. Polyprenyl-phosphate linked biosynthetic intermediates have been underlined. 5aa represents the pentapeptide linked to MurNGly in PG synthesis. The Myc-X represents mycolic acid esterified to an unknown carrier. Myc-Man-P-C_{50} is a polyprenyl-phosphate linked mycolate, shown to be involved in transport of mycolic acids through the membrane for deposition on AG (Besra *et al*, *Proc Natl Acad Sci USA*, 1994, **91:** 12735–12739).

PEPTIDOGLYCAN

The basic peptidoglycan biosynthetic machinery is similar to *E. coli* and appears to be remarkably well preserved in *M. tuberculosis*[13,14] and *M. leprae.* The main events are the synthesis of the unusual sugar-nucleotide UDP-MurNAc from UDP-GlcNAc (*murA, murB*), the sequential addition of five amino acids to the MurNAc (*murC, murD, murE* and *murF*), transfer of the sugar-pentapeptide to a polyprenyl-phosphate carrier (*murX*), followed by attachment of GlcNAC from UDP-GlcNAc (*murG*). Glycan chains are formed by transglycosylation reactions of the disaccharide—pentapeptide chains (releasing the lipid carrier from the incoming unit), followed by cross-linking of such chains via DAP-DAP and DAP-D-ala bridges in the final stage of peptidoglycan synthesis. The enzymes for transglycosylation and transpeptidation are members of the penicillin-binding protein (PBP) family. Two *M. leprae* genes (*ponA* and *pon1*) have been cloned and expressed but their role in peptidoglycan synthesis has not been reported.[37,38] The *M. leprae* genome also contains several pseudogenes with homology to PBPs. In mycobacteria, amidation of the carboxyl side chains of glutamate and DAP and the oxidation of the *N*-acetyl group of MurNAc to a glycolyl moiety occur at an undetermined stage in this pathway.

Table 1. Genetics of cell envelope synthesis in *M. leprae*. Using gene sequences of the *M. tuberculosis* or *E. coli* genes involved in the biosynthesis of the major components of the cell wall, homologs were identified in the genome of *M. leprae* (http://www. sanger.ac.uk/Projects/M_leprae) with the BLAST tool.[36] The genes indicated in **bold** have been characterized either by genetic analysis of mutants or by recombinant gene expression. The absence of a homolog is indicated by a dash (–). Pseudogenes in *M. leprae* are indicated by ψ

Name	Number assigned by genome project		Function
	M. tuberculosis	*M. leprae*	
Polyprenyl-P synthesis			
C_{50}-P synthesis			
dxs-I	**Rv 2682c**	ML10388	1-deoxy-D-xylulose 5-phosphate synthase
dxs-II	Rv3379c	–	probable 1-deoxy-D-xylulose 5-phosphate synthase
dxr	Rv2870c	ML15833	probable 1-deoxy-D-xylulose 5-phosphate reductase
ygbP	Rv3582c	ML03211	probable 2C-methyl-D-erithritol-4-phosphate cytidylyltransferase
ychB	Rv1011	ML02422	probable 4-diphosphocytidyl-2C-methyl-D-erythritol kinase
ygbB	Rv 3581c	ML03222	probable 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
lytB2	Rv1110	ML19388	gene function has not been deciphered
idi	Rv1745c	–	probable isopentenyl diphosphate isomerase
	Rv1086	ML24677	E,Z farnesyl disphosphate synthase
	Rv2361c	ML06344	E,Z decaprenyl diphosphate synthase
Peptidoglycan synthesis			
murA	**Rv1315**	ML1150	phosphoenolpyruvate:UDP-GlcNAc enolypyruvate transferase
murB	Rv0482	ML2447	UDP-*N*-acetylenolpyruvoylglucosamine reductase
murC	**Rv2152c**	**ML0915**	UDP-MurNAC: L-alanine ligase
murD	**Rv2155c**	ML0912	UDP-MurNAc-L-ala: D-glutamate ligase
murE	Rv2158c	ML0909	UDP-*N*-acetylmuramyl-tripeptide synthetase
murF	Rv2157c	ML0910	UDP-MurNAc-pentapeptide synthetase
murX	Rv2156c	ML0911	UDP-acetylmuramyl-tripeptide synthetase
murG	Rv2153	ML0915	pentapeptide pyrophosphoryl-decaprenol *N*-GlcNAc transferase
ponA	Rv0050	**ML2688**	PBP; transpeptidase or transglycosylase
ponA′	Rv3682	**ML2308**	PBP; transpeptidase or transglycosylase
Linker unit-arabinogalactan synthesis			
dTDP-rhamnose synthesis			
rmlA	**Rv0334**	ML2503	D-glucose 1-phosphate thymidylyl transferase
rmlB	**Rv3464**	ML1964	dTDP-D-glucose-4,6 dehydratase
rmlC	**Rv3465**	ML1965	dTDP-4-dehydrorhamnose 3,5-epimerase
rmlD	**Rv3266c**	ML0751	dTDP-4-dehydro-rhamnose reductase
UDP-galactofuranose synthesis			
galE	**Rv3634c**	ML0204	UDP-glucose-4-epimerase
glf	**Rv3809c**	ML0092	UDP-galactopyranose mutase
Lipid linked linker unit-arabinogalactan polymerization			
rfe	Rv1302	ML1137	probable UDP-GlcNAc: C_{50}-P GlcNAc transferase
wbbl	Rv3265c	ML0752	probable dTDP-rhamnose: C_{50}-PP-GlcNAc rhamnose transferase
glfT	**Rv3808c**	ML0093	UDP-galactofuranose transferase
embC	Rv3793	ML0106	arabinofuranose transferase? (arabinan synthesis)
embA	Rv3794	ML0105	
embB	Rv3795	ML0104	

Table 1. Continued

Name	Number assigned by genome project		Function
Mycolic acid synthesis and deposition			
α-branch synthesis			
fas	Rv2524c	ML1191	fatty acid synthase (FASI)
Meromycolic acid synthesis			
accD6	Rv2247	ML1657	acetyl-CoA carboxylase (malonyl-CoA synthase)
acpM	**Rv2244**	ML1654	acyl carrier protein
fabD	Rv2243	ML1653	malonyl-CoA-[ACP]-transacylase (malonyl-ACP) synthase
fabH	**Rv0533c**	–	β-ketoacyl-ACP synthase III
kasA	**Rv2245**	ML1655	β-ketoacyl-ACP synthase
kasB	**Rv2246**	ML1656	"
mabA	Rv1483	ML1807	3-ketoacyl-ACP reductase
inhA	**Rv1484**	ML1806	enoyl-ACP reductase
Meromycolic acid modification			
cmaA1	**Rv3392c**	ML0404	cyclopropane mycolic acid synthase (distal)
cmaA2	**Rv0503c**	ML2426	cyclopropane mycolic acid synthase (proximal)
mmaA1	**Rv0645c**	ML1900	trans cyclopropane mycolic acid synthase (oxygenated mycolates)
mmaA2	**Rv0644c**	ML1901ψ	cyclopropane mycolic acid synthase (oxygenated mycolates)
mmaA3	**Rv0643c**	ML1902ψ	methoxymycolate synthase
mmaA4	**Rv0642c**	ML1903	hydroxymycolate synthase
umaA1	Rv0469	ML2460ψ	probable mycolic acid methyltransferase
umaA2 (pcaA)	**Rv0470c**	ML2459	cyclopropane mycolic acid synthase (proximal, α-mycolates)
desA1	Rv0824c	ML2185	probable acyl-ACP desaturase
desA2	Rv1094	ML1952	"
desA3	Rv3229c	ML0789ψ	"
Deposition of mycolic acids			
fbpA	**Rv3804c**	ML0097	mycolyltransferase (TMM, TDM and mAGP synthesis?)
fbpB	**Rv1886c**	ML2028	"
fbpC2	**Rv0129c**	ML2655	"
PGL-1 synthesis			
Mycocerosoic acid synthesis			
mas	**Rv2940c**	ML0139	mycocerosoic acid synthase
fadD28	**Rv2941**	ML0138	probable acyl-CoA synthase
mmpL7	Rv2942	ML0137	mycobacterium membrane protein (transport of PDIM)
Phthiocerol synthesis			
fadD26	**Rv2930**	ML2358	probable acyl-CoA synthase
ppsA	**Rv2931**	ML2357	(phenol) phthiocerol synthase
ppsB	**Rv2932**	ML2356	"
ppsC	**Rv2933**	ML2355	"
ppsD	**Rv2934**	ML2354	"
ppsE	**Rv2935**	ML2353	"
drrA	**Rv2936**	ML2352	?
drrB	**Rv2937**	ML2351	?
drrC	**Rv2938**	ML2350	transport of PDIM
papA5	**Rv2939**	ML2349	polyketide associated protein (transport of PDIM)
Glycosylation of PDIM (trisaccharide synthesis)			
	Rv1524	ML2348	probable TDP-Rhamnose:phenol PDIM rhamnose transferase
	Rv1526c	ML2348?	
	Rv2962c	ML0125	probable UDP-glucose: phenol PDIM-rhamnose glucose transferase

Table 1. Continued

Name	Number assigned by genome project		Function
	Rv2958c	ML0128	"
	Rv2959c	ML0127	probable methyltransferase
	Rv2952	ML0130	probable methyltransferase
PIMs, LM and LAM synthesis			
	Rv0486	ML2443	probable GDP-mannose: polyprenyl-P mannosyl transferase
	Rv2051c	ML1440	"
pgsA	**Rv2612c**	ML0454	CDP-diacylglycerol: inositol phosphitidyl transferase
	Rv2611c	ML0452	probable phosphitidylinositol: GDP-mannose mannose transferase
	Rv2610c	ML0453	probable diacylphosphitidylinositol mannose: palmitoyl-CoA acyltransferase
pimB	**Rv0557**	ML2272ψ	triacylphosphitidylinositol mannose: GDP-mannose mannosyl transferase
	Rv2188c	ML0886	probable mannose transferase (mannan synthesis?)
	Rv3032	ML1715	"
	Rv0225	ML2583	"

Since the *M. leprae* peptidoglycan has glycine rather than L-alanine in the peptide cross-links, it was thought there might be a genetic basis for this substitution. Mahapatra *et al.*[39] analysed the genome but could not find a second ligase gene. In fact they demonstrated by *in vitro* studies with the recombinant MurC enzyme, that it can use L-alanine or glycine substrate with comparable affinities. Therefore the presence of glycine in *M. leprae* is perhaps an *in vivo* phenomenon driven by the ambient amino acid milieu.

ARABINOGALACTAN

The synthesis of AG requires the sugar donors UDP-GlcNAc,[40] TDP-rhamnose (*rmlA, rmlB, rmlC* and *rmlD*)[40,41], UDP-galactofuranose UDP-gal$_f$ (*galE* and *glf*)[42] and decaprenyl-phospho-arabinose DPA.[43] Successive addition of GlcNAc (*rfe*), rhamnose (*wbbl*), galacto-furanose and arabinofuranose on a prenyl-phosphate lipid carrier[44] occurs before the entire LU-AG is transferred (ligated) to approximately 1 in 10 MurNAc units of peptidoglycan.

The synthesis of DPA is interesting and is proposed to originate from the pentose phosphate pathway as phosphoribose pyrophosphate (PRPP).[45] Epimerization of the ribose to arabinose may occur before or after transfer to a decaprenyl-phosphate carrier. Regarding the galactosyltransferases, Mikusova *et al.*[44] showed that the gene Rv3808c (*glfT*) of *M. tuberculosis,* is a galactosyltransferase in AG synthesis. There is an ortholog in *M. leprae*. Furthermore, Kremer *et al.*[46] suggest that it encodes a bi-functional transferase for the alternating 5 and 6 linked galactose residues of the galactan by use of synthetic acceptors. It has not been demonstrated if GlfT can also catalyse the addition of the first galactose unit to the rhamnose sugar of the linker unit and the second galactose of the galactan. GlfT contains the sugar nucleotide binding motif hhhhDxDxh where 'h' represents an amino acid with hydrophobic nature.

With regard to arabinosyltransferases, the work of Belanger[47] suggests that the *embA* or *embB* genes of *M. avium* encode putative transferases for AG. A third gene, *embC,* also exists in all mycobacteria sequenced thus far and Escuyer *et al.*[48] have shown that knocking out the *embA or embB* genes of *M. smegmatis* causes changes in the arabinan content and structure of AG. These Emb proteins are very homologous to each other; they are large and hydrophobic, and span membranes. The genes are well conserved amongst many mycobacteria and are intact in *M. leprae* in a gene cluster very similar to that seen in *M. tuberculosis.* However, the proteins they encode have no significant homology to any other proteins in the database and no known domains or motifs. It is not clear if the Emb proteins are involved in the actual glycosyltransferase catalysis or in the assembly of the arabinan.

PRENYL-PHOSPHATE CARRIERS

The biosynthesis of many of the key cell wall polymers requires prenyl-phosphate carriers as sugar donors (DPM and DPA) and for carriers of the intermediates of cell wall polysaccharide synthesis (AG and PG).[14] The synthesis begins with the formation of deoxyxylulose phosphate from pyruvate and glyceraldehydes-3P (*dxs-I*), which is converted to the 5-carbon isoprene compounds isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) by the non-mevalonate pathway, also called the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway.[49] In *M. tuberculosis* there are two possible genes for this function (*dxs-I* and *dxs-II*). DXS-I has been shown to be functional by expression in *E. coli.*[50] Studies on DXS-II, are in progress. However, *dxs-II* may be redundant in *M. tuberculosis*, since the *M. leprae* genome has only one gene that is homologous to *dxs-I*. The complete pathway for the synthesis of IPP and DMAPP is not known. However, for all genes identified in *E. coli*[51,52], homologous genes have been found in *M. tuberculosis* and *M. leprae* for this essential pathway. A non-essential IPP isomerase (*idi*) for the interconversion of IPP and DMAPP is present in *E. coli*[53] and *M. tuberculosis* but not in *M. leprae*. The gene responsible for condensation of IPP with DMAP, to form geranyl diphosphate (GPP, C_{10}-PP) has not been identified in any organism. The addition of IPP to GPP results in farnesyl diphosphate (FPP, C_{15}-PP), which is subsequently elongated by seven cycles of polymerization to form decaprenyl diphosphate in a specific stereochemistry in *M. tuberculosis*[54] and also probably in *M. leprae* because homologous polymerase genes exist. In terms of synthesis of other isoprenoids, there are no homologs in *M. leprae* for four other prenyldiphosphate synthase genes found in the *M. tuberculosis* genome, except for the *grcC1* gene, which may be involved in the transfer of a prenyl moiety in the menaquinone pathway. There are also no homologs in *M. leprae* for the *M. tuberculosis* squalene synthase, monoxygenase and cyclase genes probably involved in steroid synthesis.

Biosynthesis of PI, PIMs, LM and LAM

The biosynthesis of phosphatidylinositol mannosides is initiated on the precursor PI using the gene product of *pgsA*[55], followed by mannosylation using GDP-mannose as the sugar donor.[56] Genes for a mannosyltransferase and an acyltransferase are linked to *pgsA* and may be required for the synthesis of PIM_1 and the acylation of mannose to form triacyl PIM_1, respectively. The gene product of Rv0557 (*pimB*) of *M. tuberculosis* has been identified as the second mannosyltransferase.[57] The rPimB was shown to convert labeled

tri-acylated PIM_1 to PIM_2 in the presence of GDP-mannose. PimB belongs to the family of glycosyltransferases that have a conserved C terminal motif EX_7E.[58] This gene is a pseudogene in *M. leprae* and we predict that one of the other genes sharing a homologous active site is used (Table 1). Subsequent mannosylation with undefined enzymes generates LM.

Of the many putative glycosyltransferases in the *M. tuberculosis* genome, several (Rv0539, Rv0696, Rv1781c, Rv1500, Rv1513, Rv1514c, Rv1516c, Rv1518, Rv1520 and Rv1525) do not have orthologs in *M. leprae*. We propose that these enzymes are not involved in the biosynthesis of mannan, arabinan and galactan, since there are no significant structural differences between these molecules in *M. tuberculosis* and *M. leprae*.

Mycolic acids, TMM and TDM and deposition on AG

The synthesis of mycolates occurs in several stages: synthesis of the α-alkyl chain and the primer for the meromycolic acid by the multifunctional fatty acyl synthase FASI enzyme (*fas*)[59]; linking by the β-keto acyl synthase III (*fabH*)[60] and elongation of the primer[11] by the disassociated fatty acyl synthase complex FASII (*accD6, fabD, kasA, kasB, acpM, mabA, inhA*);[61,62] modification of the meromycolic acid (introduction of double bonds, cyclopropane rings, keto, methyl and methoxy groups), probably in parallel with elongation;[63] and finally condensation of the α-alkyl chain and meromycolate. The condensation step and the carrier on which this reaction occurs are not known. The mycolates are then transferred to AG and trehalose to form mAGP and TMM/TDM respectively and may occur using the mycolyltransferases encoded by the members of the antigen 85 complex (*fbpA, fbpB* and *fbpC2*)[64,65]. All of these steps have been characterized in *M. tuberculosis* and homologs for the genes are present in *M. leprae*, with the exception of the gene encoding the linking enzyme FabH. The mechanism in *M. leprae* is not clear.

Phenolic glycolipids PGLs (glycosylphenolpthiocerol dimycocerosates)

The phthiocerol moiety is synthesized using a set of multifunctional enzymes (*ppsA, ppsB, ppsC, ppsD and ppsE*) that contain one or more of the acyltransferase, ketoacyl synthase, keto reductase, dehydratase, enoyl reductase and acyl carrier modules for the polymerization of malonyl-CoA and methylmalonyl-CoA units on a C_{22}-CoA fatty acid precursor.[66] Mycocerosic acid syntheisis occurs by the elongation of fatty acyl-CoA primers with methylmalonyl-CoA[67] (*mas),* followed by the transfer to the phthiocerol using a specific acyl-CoA synthase (*fadD28*). Two membrane associated proteins MmpL7 and DrrC have been shown to be responsible for the transport of the PDIM. In *M. tuberculosis*, genes for all these functions are clustered on the genome and mutations in these genes result in disruption of PDIM synthesis and loss of virulence.[68] In *M. leprae*, the *ppsA-E* genes are intact but have been separated from the *mas*, *fadD28* and *mmpL7* genes. For the addition of the first rhamnose in the trisaccharide of PGL-1 in *M. leprae*, we have analyzed the genome for genes homologous to the rhamnosyltransferases such as *rtfA* of *M. avium* (for addition of L-rhamnose to the 6-deoxy talose in GPL synthesis[69]), and the *wbbL* gene of *M. tuberculosis* (involved in linker unit synthesis). Based on homology searches with *rtfA*, we have

identified the gene ML2348 in *M. leprae* as a candidate. Co-incidentally, ML2348 is located where the phthiocerol gene cluster of *pps/drr/papA5* has separated from the *mas/fadD28/mmpL7* in *M. leprae* and may indicate its role in PGL-1. ML2348 is also homologous to genes used for the synthesis of glycosylated steroids in plants, and for antibiotics such as balhimycin and tylosin (tylN, 6-deoxyallosyltransferase) in microbes.[70] All these homologs use sterol/phenol like acceptors. Furthermore, *rtfA and tylN* encode glycosyltransferases for sugars other than glucose (particularly 6-deoxy hexoses). Combining these pieces of information, we propose that ML2348 is a good candidate for the enzyme that transfers the first rhamnose. We postulate that the genes ML0125 and ML0128 are glycosyltransferases and ML0127 and ML0130 are methylases for the synthesis of the second and third sugars of PGL-1. These genes are located close to the mycocerosoic acid gene cluster *mas/fadD28/mmpL7* genes. Genes that are highly homologous to ML2348 are also present in *M. tuberculosis* (RV1524 and 1526c). However, these are clustered with another *pks* system (pks5). There are also homologs in *M. tuberculosis* for the candidate glycosyltransferases (Rv2958c and Rv2962c) and methyltransferases (RV 2952 and Rv2959c) in a cluster similar to that in *M. leprae*, but may have no function since only phthiocerols and not the phenolpthiocerols are found in *M. tuberculosis*.[8]

Concluding remarks

In this review, we have identified putative genes of *M. leprae* for some biosynthetic pathways by homology searches with known genes of other organisms. We are comfortable with the premise that the genome of *M. leprae* approaches a minimal and perhaps also an 'essential' gene set for all basic structural and biological properties shared by virulent and avirulent mycobacteria, particularly for cell wall core synthesis. Genes for the synthesis of precursor molecules that cannot be obtained from the host environment such as UDP-Gal_f and TDP-Rha for AG, PI for PIMs, LM and LAM and DXP for polyprenyl phosphates have been retained in *M. leprae*.

In terms of the practical approaches to verifying and harnessing this genetic information, recombinant proteins can be purified and used in suitable assays. In the future, it may even be possible to set up *de novo* synthesis of complex macromolecules in crude extracts of *M. leprae*, since viable and high titre *M. leprae* are now available from nude mouse foot pads.[71] In addition, *M. leprae* from this source have been kept viable for up to 6 weeks in broth culture at 30°C. Since the number of functional ORFs is smaller than *M. tuberculosis*, use of microarrays and comparative proteomics, is a reasonable approach to identify genes that are preferentially regulated under defined and modified test conditions *in vitro*. We believe the *M. leprae* sequence information will be valuable towards efforts for elimination of leprosy.

Acknowledgements

Research conducted in the authors' laboratory was supported by NIH, NIAID, DMID Contract NO1 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. We thank Dean Crick for helpful discussions.

References

[1] Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.

[2] Cole ST, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998; **393:** 537–544.

[3] The *Mycobacterium leprae* Genome Project [http://www.sanger.ac.uk/Projects/M_leprae/]; The *Mycobacterium tuberculosis* Genome Project [http://www.sanger.ac.uk/Projects/M_tuberculosis/]; the *Mycobacterium bovis* Genome Project [http://www.sanger.ac.uk/Projects/M_bovis/]; The *Mycobacterium paratuberculosis* Genome Project [http://www.cbc.umn.edu/ResearchProjects/AGAC/Mptb/Mptbhome.html]; The *Corynebacterium diphtheriae* Genome Project [http://www.sanger.ac.uk/Projects/C_diphtheriae] The Institute for Genomic Research [http://www.tigr.org].

[4] Crick DC, Brennan PJ. Antituberculosis drug research. *Curr Opin Anti-Infect Invest Drugs* 2000; **2:** 154–163.

[5] Vissa VD, Brennan PJ. The genome of *Mycobacterium leprae:* a minimal mycobaacterial gene set. *Genome Biol*, 2001; **2:** 1023–1030.

[6] Levy L. Studies of the mouse footpad technique for cultivation of *Mycobaacterium leprae* 3. Doubling time during logarithmic multiplication. *Lepr Rev*, 1976; **47:** 103–106.

[7] Shepard CC. Stability of *Mycobacterium leprae* and the temperature optimum for growth. *Int J Lepr,* 1965; **33:** 541–550.

[8] Brennan PJ, Nikaido H. The envelope of mycobacteria. *Annu Rev Biochem*, 1995; **64:** 29–63.

[9] McNeil M, Besra GS, Brennan PJ. Chemistry of the mycobacaterial cell wall. In: Rom WN, Garay SM (eds) *Tuberculosis.* Little, Brown and Company, Boston, 1996, pp. 171–185.

[10] Chatterjee D, Khoo KH. Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects. *Glycobiology*, 1998; **8:** 113–120.

[11] Barry CE 3rd, Lee RE, Mdluli K *et al.* Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res*, 1998; **37:** 143–179.

[12] Baulard AR, Besra GS, Brennan PJ. The cell-wall core of *Mycobacterium:* structure, biogenesis and genetics. In: Ratledge C, Dale J (eds) *Mycobacteria, molecular biology and virulence.* Blackwell Science, Oxford, 1999, pp 240–259.

[13] Belanger AE, Inamine JM. Genetics of cell wall biosynthesis. In: Hatfull GF, Jacobs WR Jr (eds) *Molecular genetics of mycobacteria.* ASM Press, Washington DC, 2000, pp 191–202.

[14] Crick DC, Mahapatra S, Brennan PJ. Biosynthesis of the arabinogalactan-peptidoglycan complex of Mycobacterium tuberculosis. *Glycobiology*, 2001, **11:** 107R–118R.

[15] Hirata T. Electron microscopic observations of cell wall and cytoplasmic membrane in murine human leprosy bacilli. *Int J Lepr,* 1985; **53:** 433–440.

[16] Draper P. The bacteriology of *Mycobacterium leprae. Tubercle,* 1983; **64:** 43–56.

[17] Nishiura M, Okada S, Izumi S, Takizawa H. An electron microscope study of the band structure of the leprosy bacillus and other mycobacteria. *Int J Lepr*, 1969; **37:** 225–238.

[18] David HL, Clavel S, Clement F, Lesourd M. Paracrystalline inclusions in *Mycobacaterium leprae. Annal Microbiol* (Paris), 1981; **132A:** 41–50.

[19] Draper P, Kandler O, Darbre A. Peptidoglycan and arabinogalactan of *Mycobacterium leprae.J Gen Microbiol*, 1987; **133:** 1187–1194.

[20] 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.

[21] Draper P. Cell walls of *Mycobacterium leprae. Int J Lepr*, 1976; **44:** 95–98.

[22] McNeil M, Wallner SJ, Hunter SW, Brennan PJ. Demonstration that the galactosyl and arabinosyl residues in the cell-wall arabinogalactan of *Mycobacterium leprae* and *Mycobacterium tuberculosis* are furanoid. *Carbohydr Res*, 1987; **166:** 299–308.

[23] 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.

[24] Daffe M, Brennan PJ, McNeil M. Predominant structural features of the cell wall arabinogalactan of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by ^{1}H and ^{13}C NMR analyses. *J Biol Chem*, 1990; **265:** 6734–6743.

[25] 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.

[26] Minnikin DE, Dobson G, Goodfellow M *et al.* Quantitative comparison of the mycolic and fatty acid compositions of *Mycobacterium leprae* and *Mycobacterium gordonae. J Gen Microbiol*, 1985; **131:** 2013–2021.

[27] Behr MA, Schroeder BG, Brinkman JN *et al.* A point mutation in the *mma3* gene is responsible for impaired methoxymycolic acid production in *Mycobacterium bovis BCG* strains obtained after 1927. *J Bacteriol*, 2000; **182:** 3394–3399.

[28] Yuan Y, Zhu Y, Crane DD, Barry CE 3rd. The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis. Mol Microbiol*, 1998; **29:** 1449–1458.

[29] 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.
[30] Hunter SW, Brennan PJ. Evidence for the presence of a phosphatitidylinositol anchor on the lipoarabinomannan and lipomannan of *Mycobacterium tuberculosis. J Biol Chem.* 1990; **265:** 9272–9279.
[31] Khoo KH, Dell A, Morris HR *et al.* Structural definition of acylated phosphatidylinositol mannosides from *Mycobacterium tuberculosis:* definition of a common anchor for lipomannan and lipoarabinomannan. *Glycobiology*, 1995; **5:** 117–127.
[32] Khoo KH, Dell A, Morris HR *et al.* Inositol phosphate capping of the nonreducing termini of lipoarabinomannan from rapidly growing strains of *Mycobacterium. J Biol Chem*, 1995; **270:** 12380–12389.
[33] Hunter SW, Brennan PJ. A novel phenolic glycolipid from *Mycobacterium leprae* possibly involved in immunogenicity and pathogenicity. *J Bacteriol*, 1981; **147:** 728–735.
[34] 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.
[35] Rambukkana A. Molecular basis for the peripheral nerve predilection of *Mycobacterium leprae. Curr Opin Microbiol*, 2001; **4:** 21–27.
[36] Altschul SF, Madden TL, Schäffer AA *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 1997; **25:** 3389–3402.
[37] Lepage S, Dubois P, Ghosh TK *et al.* Dual multimodular class A penicillin-binding proteins in *Mycobacterium leprae. J Bacteriol*, 1997; **179:** 4627–4630.
[38] Basu J, Mahapatra S, Kundu M, Mukhopadhyay S *et al.* Identification and overexpression in Escherichia coli of a *Mycobacterium leprae* gene, *pon1*, encoding a high-molecular-mass class A penicillin-binding protein, PBP1. *J Bacteriol*, 1996; **178:** 1707–1711.
[39] Mahapatra S, Crick DC, Brennan PJ. Comparison of the UDP-N-acetylmuramate:L-alanine ligase enzymes from *Mycobacterium tuberculosis* and *Mycobacterium leprae. J Bacteriol*, 2000; **182:** 6827–6830.
[40] Mikusova K, Mikus M, Besra GS *et al.* Biosynthesis of the linkage region of the mycobacterial cell wall. *J Biol Chem*, 1996; **271:** 7820–7828.
[41] Ma Y, Stern RJ, Scherman MS *et al.* Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrob Agents Chemother*, 2001; **45:** 1407–1416.
[42] Weston A, Stern RJ, Lee RE *et al.* Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. *Tuberc Lung Dis*, 1997; **78:** 123–131.
[43] Wolucka BA, McNeil MR, de Hoffmann E *et al.* Recognition of the lipid intermediate for arabinogalactan/arabinomannan biosyshtesis and its relation to the mode of action of ethambutol on mycobacteria. *J Biol Chem*, 1994; **269:** 23328–23335.
[44] Mikusova K, Yagi T, Stern R *et al.* Biosynthesis of the galactan component of the mycobacterial cell wall. *J Biol Chem*, 2000; **275:** 33890–33897.
[45] Scherman MS, Kalbe-Bournonville L, Bush D *et al.* Polyprenylphosphate-pentoses in mycobacteria are synthesized from 5-phosphoribose pyrophosphate. *J Biol Chem*, 1996; **271:** 29652–29658.
[46] Kremer L, Dover LG, Morehouse C *et al.* Galactan biosynthesis in *Mycobacterium tuberculosis:* identification of a bifunctional UDP-galactofuranosyltransferase. *J Biol Chem*, 2001; **276:** 26430–26440.
[47] Belanger AE, Besra GS, Ford ME *et al.* The *embAB* genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc Natl Acad Sci USA*, 1996; **93:** 1919–1924.
[48] Escuyer VE, Lety ME, Torrelles JB *et al.* The role of the *embA, B* gene products in the biosynthesis of the terminal hexaarabinofuranosyl motif of *Mycobacterium smegmatis* arabinogalactan. *J Biol Chem*, 2001; Oct 24.
[49] Rohmer, M, Knani, M, Simonin, P *et al.* Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J*, 1993; **295:** 517–524.
[50] Bailey AM, Mahapatra S, Brennan PJ, Crick DC. Purification and characterization of 1-deoxyxylulose 5-phosphate synthase from *Mycobacterium tuberculosis. Glycobiology*, 2000; **10:** 46.
[51] Takahashi S, Kuzuyama T, Watanabe H, Seto H. A 1-deoxy-D-xylulose 5-phosphate reductoisomease catalyzing the formation of 2-*C*-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis. *Proc Natl Acad Sci USA*, 1998; **95:** 9879–9884.
[52] Campos N, Rodriguez-Concepcion M, Sauret-Gueto S *et al. Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *Biochem J*, 2001; **353:** 59–67.
[53] Rodriguez-Concepcion M, Campos N, Lois LM *et al.* Genetic evidence of branching in the isoprenoid pathway for the production of isopentenyl diphosphate and dimethylallyldiphosphate in *Escherichia coli. FEBS lett*, 2000; **473:** 328–332.
[54] Schulbach MC, Brennan PJ, Crick DC. Identification of a short (C_{15}) chain Z-isoprenyl diphosphate synthase and a homologous long (C_{50}) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis. J Biol Chem,* 2000; **275:** 22876–22881.

[55] Jackson M, Crick DC, Brennan PJ. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J Biol Chem*, 2000; **275:** 30092–30099.
[56] Besra GS, Morehouse CB, Rittner CM *et al.* Biosynthesis of mycobacterial lipoarabinomannan. *J Biol Chem*, 1997; **272:** 18460–18466.
[57] Schaeffer ML, Khoo KH, Besra GS *et al.* The *pimB* gene of *Mycobacterium tuberculosis* encodes a mannosyltransferase involved in lipoarabinomannan biosynthesis. *J Biol Chem*, 1999; **274:** 31625–31631.
[58] Geremia RA, Petroni EA, Ielpi L *et al.* Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic α-mannosyltransferases. *Biochem J*, 1996; **318:** 133–138.
[59] Fernandes ND, Kolattukudy PE. Cloning, sequencing and characterization of a fatty acid synthase-encoding gene from *Mycobacterium tuberculosis* var. bovis BCG. *Gene*, 1996; **170:** 95–102.
[60] Choi KH, Kremer L, Besra GS, Rock CO. Identification and substrate specificity of beta -ketoacyl (acyl carrier protein) synthase III (mtFabH) from *Mycobacterium tuberculosis. J Biol Chem*, 2000; **275:** 28201–28207.
[61] Mdluli K, Slayden RA, Zhu Y *et al.* Inhibition of a *Mycobacterium tuberculosis* beta-ketoacyl ACP synthase by isoniazid. *Science*, 1998; **280:** 1607–1610.
[62] Quemard A, Sacchettini JC, Dessen A *et al.* Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis. Biochemistry*, 1995; **34:** 8235–8241.
[63] Yuan Y, Mead D, Schroeder B G *et al.* The biosynthesis of mycolic acids in *Mycobacterium tuberculosis.* Enzymatic methyl(ene) transfer to acyl carrier protein bound meromycolic acid *in vitro. J Biol Chem*, 1998; **273:** 21282–21290.
[64] Belisle JT, Vissa VD, Sievert T *et al.* Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science*, 1997; **276:** 1420–1422.
[65] Puech V, Bayan N, Salim K *et al.* Characterization of the *in vivo* acceptors of the mycoloyl residues transferred by the corynebacterial PS1 and the related mycobacterial antigens 85. *Mol Microbiol*, 2000; **35:** 1026–1041.
[66] Azad AK, Sirakova TD, Fernandes ND, Kolattukudy PE. Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem*, 1997; **272:** 16741–16745.
[67] Mathur M, Kolattukudy PE. Molecular cloning and sequencing of the gene for mycocerosic acid synthase, a novel fatty acid elongating multifunctional enzyme, from *Mycobacterium tuberculosis* var. bovis Bacillus Clamette-Guerin. *J Biol Chem*, 1992; **267:** 19388–19395.
[68] Camacho LR, Constant P, Raynaud C *et al.* Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis:* evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem*, 2001; **276:** 19845–19854.
[69] Eckstein TM, Silbaq FS, Chatterjee D *et al.* Identification and recombinant expression of a *Mycobacterium avium* rhamnosyltransferase gene (*rtfA*) involved in glycopeptidolipid biosynthesis. *J Bacteriol*, 1998; **180:** 5567–5573.
[70] Wilson VT, Cundliffe E. Characterization and targeted disruption of a glycosyltransferase gene in the tylosin producer, *Streptomyces fradiae. Gene*, 1998; **214:** 95–100.
[71] Truman RW, JL Krahenbuhl. Viable *M.leprae* as a research reagent. *Int J Lepr*, 2001; **69:** 1–12.

Lepr Rev (2001) **72,** 429–440

Genomics and the chemotherapy of leprosy

JACQUES H. GROSSET* & STEWART T. COLE**
**Laboratoire de Bactériologie et Hygiène, Faculté de Médecine Pitié-Salpêtrière, 91 Boulevard de l'Hôpital, 75634 Paris, Cedex 13, France*
***Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, Cedex 15, France*

Summary The information deduced from the genome sequence of *Mycobacterium leprae* is of immense value for the chemotherapy of leprosy. Knowing the complete set of genes, enzymes and proteins allows us to understand why some drugs are without effect whereas others are fully active. It may also enable better use to be made of existing drugs, such as β-lactams, and opens new avenues for the development of novel compounds. *M. leprae* is relatively susceptible to a wide range of drugs, unlike the highly related tubercle bacillus, and several new multidrug regimens are in clinical trials. Genomics provides a number of possible explanations for this broader susceptibility as some of the genes encoding enzymes involved in antibiotic inactivation have decayed whereas the number of transporters available to contribute to drug efflux is considerably lower than in *Mycobacterium tuberculosis.* Several leads for new drug targets have been uncovered.

Introduction

The WHO-recommended multidrug therapy (MDT) for leprosy has been, without question, one of the major success stories in the field of public health.[1] There is, however, no room for complacency as the incidence of detected cases of leprosy has not fallen during the last decade[2] and the spectre of drug resistance is never far away. This has been well illustrated by the increased spread of multidrug resistant tuberculosis during the last decade.[3] Furthermore, regimens can always be improved by increasing efficacy or reducing duration and this is only likely to be achieved by employing drugs that are stronger or more effective than dapsone and clofazimine.

A number of compounds, such as minocycline, various fluoroquinolones and macrolides, have shown excellent activity in the mouse model of leprosy, and in limited clinical trials,[4–6] and ROM, a new regimen for the treatment of single lesion paucibacillary leprosy, comprising rifampicin, ofloxacin and minocycline,[7,8] has shown particular promise. In contrast to the situation in tuberculosis, where new chemotherapeutic agents are desperately

*Correspondence to: Prof. S.T. Cole, Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, Cedex 15, France. Phone: 33-1-45 68 84 46. Fax: 33-1-40 61 35 83 (e-mail: stcole@pasteur.fr)

0305-7518/98/072429+12 $1.00

needed, the future of leprosy treatment is well poised. Nevertheless, it is important not to neglect any opportunities for improvement and to remain aware of potential new leads for strengthening the chemotherapy of leprosy.

Drug resistance in bacteria can result from four different mechanisms. The commonest one, found in bacteria, is the enzymatic inactivation of the antibiotic by modification or hydrolysis. Since the genes encoding these enzymes are generally transferable, this resistance mechanism has often limited the usefulness of numerous β-lactam and aminoglycoside compounds. Transferable drug resistance has not been reported among the intracellular mycobacteria where the most frequent mechanism encountered is alteration of the drug target by mutation,[9,10] which is the case for rifampicin resistance in *Mycobacterium leprae*.[11,12] In some pathogens, innate drug resistance results from efficient permeability barriers preventing the entry of sufficient concentrations of drugs into the cell, and this appears to be operational in some mycobacteria.[13,14] Active efflux of antibiotics also occurs from various pathogens, including mycobacteria,[15] and is mediated by protein pumps that belong to different families such as the major facilitator superfamily (MFS), the resistance-nodulation-division (RND) superfamily, the ATP-binding cassette (ABC) superfamily, or the small multidrug resistance (SMR) superfamily.[16–18]

The genome sequence of *M. leprae*[19] provides clear explanations for the natural resistance of the leprosy bacillus to many antitubercular compounds and also indicates areas of potential susceptibility where existing drugs might find application. Our aim here is to discuss these topics in the light of the available clinical and microbiological knowledge.

Materials and methods

Bioinformatic analysis was performed using the procedures outlined previously[19,20] in conjunction with the relational databases, Leproma (http://genolist.pasteur.fr/Leproma/[21]) and TubercuList (http://genolist.pasteur.fr/TubercuList/).

Results and discussion

RIFAMPICIN, DAPSONE AND CLOFAZIMINE

WHO-recommended MDT relies on the association of three agents: rifampicin, dapsone and clofazimine. Since rifampicin is the backbone of leprosy treatment, the genome sequencing project began by studying the region encompassing the *rpoB* gene, encoding the drug target, the β-subunit of DNA-dependent RNA polymerase.[22] The sequence information, in conjunction with a panel of well-characterized resistant patient isolates,[23] enabled the molecular basis of rifampicin resistance to be elucidated.[11] Missense mutations were found in a restricted region of *rpoB* that probably reduces the affinity of the drug for RNA polymerase, and this information has been used to develop rapid diagnostic tests for resistance.[12,24,25] There is only one *rpoB* and no other genes associated with rifampicin resistance, like those mediating its ribosylative inactivation in certain actinomycetes,[26,27] could be found in the genome. This explains why other rifamycin derivatives that have the same drug target as rifampicin are inactive because of cross-resistance, and cannot be used as replacement drugs in the case of rifampicin resistance. However, some of them have pharmacokinetic

parameters more favourable than those of rifampicin for monthly administration. For example, rifapentine, a long lasting rifamycin derivative, has a serum half life 6 times longer than that of rifampicin and gives in humans an AUC after oral administration of 600 mg, 3 times larger than that given by 600 mg of rifampicin. Consequently, a monthly regimen containing rifapentine combined with a new fluoroquinolone and minocycline is under intensive investigation at the present time.[67]

Possibly as a result of its widespread, initial use in monotherapy, primary resistance to dapsone became common in the 1960s. Again, genomics was of great importance for understanding the basis of this resistance as the availability of the folate synthase gene sequence, *folP1*, enabled Kai and coworkers to demonstrate the presence of mutations in the gene from dapsone resistant strains.[28] Intriguingly, both *Mycobacterium tuberculosis* and *M. leprae* have a second gene, *folP2*, that resembles *folP1*, but this does not encode folate synthase nor mediate dapsone sensitivity.[29] Since folate synthesis is an essential activity, missense mutations occur that lead to amino acid substitutions at positions 53 and 55,[28,29] and from the crystal structure of the *M. tuberculosis* enzyme it is apparent that these are located in loop 2 near the active site.[30] The availability of the three-dimensional structure will allow lead compounds for antimycobacterial drug design to be designed in a rational manner, and clearly, if successful, this could benefit leprosy treatment.

The third component of MDT, clofazimine, is the least well understood in terms of its mode of action and resistance mechanism.[31] Its activity is primarily confined to mycobacteria where it is believed to bind to DNA. The genome sequence has yet to provide clues to possible targets.

ISONIAZID, ETHIONAMIDE, CELL WALL INHIBITORS

The unusual cell wall of mycobacteria has provided a rich and specific source of drug targets and, with the exception of rifampicin, three of the four agents used in the short course chemotherapy of tuberculosis block the synthesis of cell wall components. Somewhat surprisingly, in light of the conserved, if somewhat simplified nature of the cell wall of *M. leprae*,[32] none of these compounds is active, but explanations for their lack of efficacy are provided by the genome sequence. Isoniazid (INH) is exquisitely potent on the tubercle bacillus, where it is converted by catalase-peroxidase,[33] to an iso-nicotinoyl radical[34] which forms an adduct with NADH that blocks the action of the enoyl-ACP reductase, InhA.[35,36] This enzyme is involved in mycolic acid synthesis, as part of the fatty acid synthase II complex (FAS II), and is also a target for ethionamide.[35] In addition, INH interacts directly with another FAS II component, the beta-ketoacyl-ACP synthase, KasA.[37,38]

Although *M. leprae* has an intact FAS II system and functional *kasA* and *inhA* genes, it is not susceptible to clinically significant levels of INH as a result of multiple lesions in the catalase-peroxidase (*katG)* gene.[39,40] Unlike some other mycobacteria,[41] the genome of *M. leprae* does not contain a second catalase-peroxidase gene.[19] By contrast, as the FAS II complex is active, it is possible that *M. leprae* will be susceptible to triclosan, which inhibits InhA, and to an old compound, thiolactomycin, which affects KasA in *M. tuberculosis*, and other bacteria.[38] The leprosy bacillus is known to be inhibited by prothionamide (combined with isoniazid and dapsone in the fixed drug combination isoprodian), a hepatoxic drug[42] which, like ethionamide, should target InhA. It has recently been shown that ethionamide requires an activation step that is catalysed by a mono-oxygenase belonging to a 14-membered family (http://genolist.pasteur.fr/TubercuList/mast/

P14.2.html) in *M. tuberculosis*.[43,44] It is of some interest that expression of the monooxygenase gene, Rv3854c, (*ethA* or *etaA*) is regulated by a TetR repressor protein encoded by the neighbouring gene Rv3855, (*ethR*, *etaR*). Both the monoxygenase and the regulatory gene have been conserved in *M. leprae*, (ML0064, ML0065), implying that thioamide activation should proceed in a similar manner. Astonishingly, the monoxygenase gene is the sole survivor of the P14.2 family and, in the face of such extensive gene loss,[19] this implies that its physiological function must be important for *M. leprae* in particular, and mycobacteria in general.

Ethambutol inhibits the arabinosyltransferase(s), membrane-bound enzymes involved in the polymerization of arabinan which, in turn, impacts on arabinogalactan and lipoarabinomannan production thereby destabilizing the cell wall.[45] Using complementary approaches with *M. tuberculosis* and *Mycobacterium avium*,[46,47] these enzymes were shown to be encoded by linked genes, *emb(C)AB,* that have evolved by a gene duplication mechanism and are probably controlled by the regulatory gene, *embR*. While *M. leprae* has the *embCAB* operon it appears to have lost *embR*, which may mean that arabinosyltransferase production is no longer regulated.[19] Missense mutations located in a tetrapeptide at positions 303–306 of EmbB are responsible for acquired drug resistance in the majority of clinical isolates of *M. tuberculosis*[48] and in laboratory mutants of *Mycobacterium smegmatis*.[49] The *embB* gene of *M. leprae* harbours a 'mutation' at this position[46] and this undoubtedly accounts for its natural resistance to ethambutol.

Pyrazinamide is the third antitubercular agent that is believed to affect cell wall biogenesis possibly through indirect effects on fatty acid synthesis mediated by FAS I in the tubercle bacillus.[50] Drug activity requires its conversion to pyrazinoic acid in a reaction catalysed by the pyrazinamidase enzyme encoded by *pncA* in *M. tuberculosis*[51] or by a broad-spectrum amidase coded for by *pzaA* in *M. smegmatis*.[52,53] Resistance is associated with mutations that inactivate *pncA*, or alter the stability of pyrazinamidase, and in some mycobacteria, pyrazinoic acid efflux systems may also contribute.[54,55] The *pncA* gene of *M. leprae* has been inactivated and the genome contains no counterpart of *pzaA*. Pyrazinamide is therefore likely to be of no therapeutic value for leprosy.

TRANSLATION INHIBITORS

Antibiotics belonging to the tetracyline, aminoglycoside and macrolide families are potent inhibitors of protein synthesis. The aminoglycosides streptomycin and kanamycin show strong bactericidal activity in the mouse, and streptomycin has been used to treat leprosy in humans.[56] Resistance to streptomycin in mycobacteria arises as a result of missense mutations to the *rpsL* and *rrs* genes encoding the drug targets, the ribosomal protein S12 and the 16S rRNA, respectively.[57,58] Likewise, resistance to kanamycin is due to base changes around position 1400 in the 16S rRNA of *M. tuberculosis* that prevent the drug from binding.[59] As expected, both genes have wild type sequences in *M. leprae* thereby explaining its susceptibility to streptomycin and kanamycin.

Clarithromycin is a macrolide antibiotic that shows bactericidal activity against *M. leprae* in mice and humans.[5,6] Susceptibility can be attributed to the wild type sequence of the *rrl* gene, encoding the 23S rRNA. A to G transitions affecting positions 2058 and 2059 of this RNA have been described in clarithromycin resistant strains of *M. avium* and *Mycobacterium kansasii*, among others.[60,61] Minocycline, a second-generation tetracycline, is also active on *M. leprae*,[7,62] probably as a result of its ability to bind to a site on the ribosome comprising

proteins S7, S14, S19 and the 3′ domain of 16S rRNA. At present nothing is known about minocycline resistance in mycobacteria, although tetracycline resistance has been studied intensively in other Gram positive bacteria where it often involves efflux or ribosome protection systems[63]

Fusidic acid is another broad spectrum antibiotic that targets the ribosome and inhibits the growth of *M. leprae*.[64,65] It acts by preventing release of elongation factor EF-G from the ribosome and, in enteric bacteria, fusidic acid resistance is due to missense mutations in three highly conserved regions of the *efg* gene.[66] The availablity of the *efg* sequence of *M. leprae* allows genotypic tests for resistance to be developed should the need arise.[22]

QUINOLONES

Fluoroquinolones offer great potential to the future therapy of leprosy and have shown outstanding activity *in vitro* and *in vivo*.[4,62,67] Ofloxacin is a key component of the ROM regimen, discussed above, although some of the newer fluoroquinolones such as moxifloxacin appear to be even more bactericidal. A limited number of cases of fluoroquinolone resistance have been reported[68] and, as in *M. tuberculosis*,[69] these involve amino acid substitutions in the quinolone resistance determining region, QRDR, of the DNA gyrase A protein. In some bacteria, resistance also results from alterations of DNA topoisomerase IV, but genomics suggests that both *M. leprae* and *M. tuberculosis* lack this function.[19,70] The QRDR is situated very near the active site of GyrA and most unusually this region of *gyrA* has acquired an intein sequence coding for a putative homing endonuclease.[71] In consequence, production of active GyrA requires excision of the intein from the nascent polypeptide by protein splicing. There are three other inteins in the ML0593, *dnaB*, and *recA* genes of *M. leprae,* although these all differ in size and sequence.[72] Since both *gyrA* and *dnaB* encode essential functions they are valid drug targets whose functions would be lost if protein splicing were blocked.

ANTIBIOTIC INACTIVATION

Of the several reasons why *M. leprae* shows susceptibility to a broader range of drugs than *M. tuberculosis,* one is provided by the finding that its genome contains far fewer genes encoding enzymes that could inactivate or modify antibiotics. Only two of the 10 genes, annotated as being putatively involved in antibiotic modification in *M. tuberculosis,* are predicted to be functional in *M. leprae*. One of these, ML2551, encodes an aminoglycoside-2′-N-acetyltransferase that may be involved in peptidoglycan modification.[73]

While the tubercle bacillus has seven known or potential β-lactamase genes (Table 1), *M. leprae* has only two that appear to be functional (ML0270, ML1923). Consequently, β-lactam antibiotics may be more active in leprosy than in tuberculosis. It should be noted that there is no *blaC* ortholog encoding the class-A β-lactamase present in *M. tuberculosis* which contributes to its innate resistance to β-lactam antibiotics.[74,75] As expected of a class-A enzyme, inhibition was achieved by the β-lactamase inhibitors clavulanate or sulbactam, thereby raising the possibility of treating tuberculosis with penicillins and cephalosporins in conjunction with such inhibitors. There has been some recent interest in using these combinations to treat leprosy as well, inspired in part by Shephard's observations that, of the 12 β-lactams tested, two cephalosporins and one cephamycin were active in the mouse

Table 1. Predicted β-lactamase-like proteins in tubercle and leprosy bacilli

Gene	Predicted function in *M. tuberculosis*	*M. leprae**
blaC (Rv2068c)	Class A β-lactamase	del
lipD (Rv1923)	Similar to esterases, β-lactamase	del
lppW (Rv2905)	Lipoprotein with slight similarity to β-lactamase	ML1614, ps
lpqF (Rv3593)	Lipoprotein with slight similarity to class C β-lactamase	ML1923
Rv0406c	β-lactamase-like protein with Pfam match PF00753 lactamase_B, metallo-beta-lactamase superfamily	ML0270
Rv0907	Similar to PBP 4, class C β-lactamase	ML2116, ps
Rv1913	Similar to dehydrase, metallo-β-lactamase	ML2001, ps

*ps, denotes pseudogene; del, missing probably deleted.

footpad model.[76] In two studies involving β-lactam antibiotics with β-lactamase inhibitors, bactericidal activity was reported.[77,78].

DRUG EFFLUX

Another potential mechanism that could contribute to natural antimicrobial resistance is drug efflux and, in some pathogenic bacteria, this is known to be mediated by transmembrane proteins belonging to the ATP-binding cassette (ABC), and major facilitator superfamilies (MFS),[16] the small multidrug resistance family (SMR) and the resistance/nodulation/cell division family (RND).[17,18,79] *M. leprae* has proteins belonging to all four of these families but they are considerably less abundant than in the tubercle bacillus.

Careful analysis of the ABC transport proteins of *M. tuberculosis* has been undertaken,[80] and these can be divided into import and export systems on the basis of their structure and organization. Of the 11 potential drug export systems predicted (Table 2), only six remain in *M. leprae*. Pseudogenes for two may be found and the remainder appear to have been deleted. The *drrABC* system is very similar to those produced by various *Streptomyces spp.* and like them may also be involved in the export of daunorubicin-like molecules.[81] Investigations into the possible contribution of some of the ~30 MFS proteins of *M. tuberculosis* to drug efflux have been reported and these can be used to interpret the likely role of the few remaining orthologs in *M. leprae*. Two MFS proteins, Rv1258c and Rv1410c, have been shown to serve as proton motive force-dependent drug pumps that confer increased resistance to several aminoglycosides and tetracycline when expressed in *M. smegmatis*,[82,83] Both of these functions have been conserved in *M. leprae,* together with a third MFS protein, EfpA (Table 2) which is similar to the multidrug resistance pump, QacA.[84] Recently, expression of EfpA has been shown to be strongly induced during drug-mediated inhibition of cell wall synthesis in *M. tuberculosis*.[85] It is possible that these three conserved MFS proteins also act as drug pumps in *M. leprae* (Table 2).

M. leprae, like the tubercle bacillus, has only one member of the SMR family, the 108 residue ML1756 protein (Table 2, equivalent to Rv3065) and this has four transmembrane stretches like its relatives.[19,70] When expressed in *M. smegmatis* the Rv3065gene confers resistance to a variety of compounds, including acriflavine, erythromycin, ethidium bromide, safranin O, and pyronin Y 8586. *M. tuberculosis* is somewhat unusual as its genome contains 16 genes (Table 2) encoding members of the RND superfamily,[79] an exceptionally high

Table 2. Predicted drug efflux systems in tubercle and leprosy bacilli

ABC systems		
Gene	Predicted function in *M. tuberculosis*	*M. leprae**
drrABC	Daunorubicin resistance	*drrABC*
Rv1456-58c	Antibiotic resistance	ML0590, ML0589
Rv2686-88c	Antibiotic resistance	ML1033-35, ps
Rv1217-18c	Antibiotic resistance	ML1072-73, ps
Rv1272-73c	Multidrug resistance	ML1113-14
Rv1348-49	Multidrug resistance	del
Rv0194	Multidrug resistance	del
Rv1819c	Multidrug resistance	ML2084
Rv1473	Macrolide resistance	ML1816
Rv2477c	Macrolide resistance	ML1248
Rv1667-68c	Macrolide resistance	ML1239-40, ps
MFS systems		
Gene	Observed or predicted function in *M. tuberculosis*	*M. leprae*
tap (Rv1258c)	Aminoglycoside, tetracycline efflux	ML1104
efpA (Rv2846c)	Induced by drugs	ML1562
Rv1410c	P55, aminoglycoside, tetracycline efflux	ML0556
Rv1877, Rv2044, Rv2333c	Probable drug efflux proteins	All missing
Rv2459, Rv3728, Rv3239c	Probable drug efflux proteins	All missing
SMR systems		
Gene	Predicted function in *M. tuberculosis*	*M. leprae*
emrE (Rv3065)	Acriflavine, erythromycin, ethidium bromide, safranin O, pyronin Y resistance	ML1756
RND systems		
Gene	Known or predicted function in *M. tuberculosis*	*M. leprae*
mmpL1-6	Lipid transport	*mmpL3* (ML2620) *mmpL4* (ML2378)
mmpL7	PDIM transport	*mmpL7* (ML0137)
mmpL8-14	lipid transport	*mmpL10* (ML1231) *mmp11* (ML2617)

ps, denotes pseudogene; del, missing probably deleted.

number compared to other fully sequenced bacterial genomes rivalled only by *Pseudomonas aeruginosa*.[87] There are only five RND proteins predicted in *M. leprae* (Table 2).

In the Gram negative pathogens *P. aeruginosa* and *Escherichia coli,* RND proteins, such as MexAB or AcrAB (~1000 amino acids), act as proton motive force-dependent efflux systems and confer high levels of resistance to fluoroquinolones and other antimicrobial agents.[88,89] The genetic context of the *mmpL* genes, encoding the *M. tuberculosis* RND proteins, suggested an involvement in the export of lipids or glycolipids, and a body of experimental evidence to support this has since been amassed.[90,91] In particular, the MmpL7 protein is responsible for the export of the complex lipid phthiocerol-dimycocerosate (PDIM) and in *M. leprae,* modifies PDIM to produce phenolic glycolipid 1 (PGL1), MmpL7 may be

involved in the transport of PGL1. Given the similarities with other RND transporters it is possible that the MmpL proteins can also act in drug efflux, and as *M. leprae* produces far fewer than the tubercle bacillus does, it should be susceptible to more drugs as has been observed.

NEW DRUG REGIMENS AND NEW LEADS FOR DRUG DISCOVERY

Given the cost of developing new drugs, it seems certain that the pharmaceutical industry will not invest in the field of leprosy although tuberculosis may present a somewhat more lucrative market. Consequently, pharmacogenomics and high-throughput screening technologies will not be applied directly to *M. leprae* and we must look elsewhere for new leads for drug discovery. An exciting opening has emerged recently from studying the action of nitroimidazopyran derivatives on *M. tuberculosis* and a novel compound, PA824, shows great promise.[92] In order to be active, PA824 requires the F420-dependent glucose-6-phosphate dehydrogenase encoded by *fgd*, (Rv0407), and resistance mutations reside in this gene which is also conserved in *M. leprae* (ML0269). It is thus conceivable that PA824 could find use in leprosy treatment and a particularly attractive feature of this drug is its action under microaerophilic growth conditions similar to those which *M. leprae* is believed to favour.[19]

Another area where novel drug targets may be found is in signal transduction. In most bacteria, changes in gene expression in response to environmental cues are mediated by the His-Asp phosphorelay system effected by the two-component systems.[93] These are common in prokaryotes and comprise a membrane-bound sensor protein with histidine kinase activity which phosphorylates an aspartyl residue in a response regulator protein that in turn controls the target genes. In mycobacteria, a second, eukaryotic like phosphorelay system may be found in the form of the serine-threonine protein kinases and their cognate phosphoprotein phosphatases and these may also control cellular processes such as division and septation.[70,94] The pharmaceutical industry has batteries of lead compounds for both of these protein kinase families, since there has been intensive research into the histidine kinases as novel drug targets in recent years,[95,96] and numerous inhibitors of serine threonine protein kinases have been developed for use in cancer treatment. If these kinase inhibitors became available, they should certainly be tested on *M. leprae* as this pathogen has only four complete two-component systems and four serine-threonine protein kinases, thus increasing the chances of attaining complete inhibition.

Acknowledgements

We wish to acknowledge the financial support of the Institut Pasteur, the Association Française Raoul Follereau, and the European Community (QLRT-2000-02018).

References

[1] WHO Study Group. *Chemotherapy of leprosy for control programmes*. Technical Report Series No. 675. World Health Organization, Geneva, Switzerland, 1982.
[2] WHO Leprosy – Global Situation. *Weekly Epidemiological Record*, 2000; No. 28, 14 July.
[3] Snider DE Jr, Raviglione M, Kochi A. Global burden of tuberculosis. In: Bloom BR (ed) *Tuberculosis: pathogenesis, protection, and control*. American Society for Microbiology, Washington DC 20005, 1994.

[4] Consigny S, Bentoucha A, Bonnafous P *et al*. Bactericidal activities of HMR 3647, moxifloxacin, and rifapentine against *Mycobacterium leprae* in mice. *Antimicrob Agents Chemother*, 2000; **44:** 2919–2921.
[5] Ji B, Jamet P, Perani EG *et al*. Bactericidal activity of single dose of clarithromycin plus minocycline, with or without ofloxacin, against *Mycobacterium leprae* in patients. *Antimicrob Agents Chemother*, 1996; **40:** 2137–2141.
[6] Ji B, Perani EG, Petinom C, Grosset JH. Bactericidal activities of combinations of new drugs against *Mycobacterium leprae* in nude mice. *Antimicrob Agents Chemother*, 1996; **40:** 393–399.
[7] Ji B, Sow S, Perani E *et al*. Bactericidal activity of a single-dose combination of ofloxacin plus minocycline, with or without rifampin, against *Mycobacterium leprae* in mice and in lepromatous patients. *Antimicrob Agents Chemother*, 1998; **42:** 1115–1120.
[8] Efficacy of single dose multidrug therapy for the treatment of single-lesion paucibacillary leprosy. Single-lesion Multicentre Trial Group. *Ind J Lepr*, 1997; **69:** 121.
[9] Cole ST. *Mycobacterium tuberculosis:* drug resistance mechanisms. *Trends Microbiol*, 1994; **2:** 411–415.
[10] Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis:* 1998 update. *Tuberc Lung Dis*, 1998; **79:** 3–29.
[11] Honoré N, Cole ST. Molecular basis of rifampin-resistance in *Mycobacterium leprae. Antimicrob Agents Chemother*, 1993; **37:** 414–418.
[12] Williams DL, Waguespack C, Eisenach K *et al*. Characterization of rifampin resistance in pathogenic mycobacteria. *Antimicrob Agents Chemother*, 1994; **38:** 2380–2386.
[13] Brennan PJ, Nikaido H. The envelope of mycobacteria. *Annu Rev Biochem*, 1995; **64:** 29–63.
[14] Jarlier V, Nikaido H. Permeability barrier to hydrophilic solutes in *Mycobacterium chelonei. J Bacteriol*, 1990; **172:** 1418–1423.
[15] Liu J, Takiff HE, Nikaido H. Active efflux of fluoroquinolones in *Mycobacterium smegmatis* mediated by LfrA, a multidrug efflux pump. *J Bacteriol*, 1996; **178:** 3791–3795.
[16] Pao SS, Paulsen IT, Saier MJ. Major facilitator superfamily. *Microbiol Mol Biol Rev*, 1998; **62:** 1–34.
[17] Paulsen IT, Brown MH, Skurray RA. Proton-dependent multidrug efflux systems. *Microbiol. Rev*, 1996; **60:** 575–608.
[18] Paulsen IT, Sliwinski MK, Saier MJ. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *J Mol Biol*, 1998; **277:** 573–592.
[19] Cole ST, Eiglmeier K, Parkhill J *et al*. Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.
[20] Tekaia F, Gordon SV, Garnier T *et al*. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuberc Lung Dis*, 1999; **79:** 329–342.
[21] Jones LM, Cole ST, Moszer I. Leproma: a *Mycobacterium leprae* genome browser. *Lepr Rev*, 2001; **72:** 429–440.
[22] Honoré N, Bergh S, Chanteau S *et al*. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol Microbiol*, 1993; **7:** 207–214.
[23] Grosset JH, Guelpa-Lauras CC, Bobin P *et al*. Study of 39 documented relapses of multibacillary leprosy after treatment with rifampin. *Int J Lepr*, 1989; **57:** 607–614.
[24] Honoré N, Perani E, Telenti A *et al*. A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. *Int J Lepr*, 1994; **61:** 600–604.
[25] Honoré N, Roche P, Grosset J, Cole ST. A method for rapid detection of rifampicin-resistant isolates of *Mycobacterium leprae*. *Lepr Rev*, 2001; **72:** 441–448.
[26] Quan S, Venter H, Dabbs ER. Ribosylative inactivation of rifampin by *Mycobacterium smegmatis* is a principal contributor to its low susceptibility to this antibiotic. *Antimicrob Agents Chemother*, 1997; **41:** 2456–2460.
[27] Andersen SJ, Quan S, Gowan B, Dabbs ER. Monooxygenase-like sequence of a *Rhodococcus equi* gene conferring increased resistance to rifampin by inactivating this antibiotic. *Antimicrob Agents Chemother*, 1997; **41:** 218–221.
[28] Kai M, Matsuoka M, Nakata N *et al*. Diaminodiphenylsulfone resistance of *Mycobacterium leprae* due to mutations in the dihydropteroate synthase gene. *FEMS Microbiol Lett*, 1999; **177:** 231–235.
[29] Williams DL, Spring L, Harris E *et al*. Dihydropteroate synthase of *Mycobacterium leprae* and dapsone resistance. *Antimicrob Agents Chemother*, 2000; **44:** 1530–1537.
[30] Baca AM, Sirawaraporn R, Turley S *et al*. Crystal structure of *Mycobacterium tuberculosis* 7,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action. *J Mol Biol*, 2000; **302:** 1193–1212.
[31] Chopra I, Brennan P. Molecular action of anti-mycobacterial agents. Tuberc Lung Dis, 1997; **78:** 89–98.
[32] Brennan PJ, Vissa VD. Genomic evidence for the retention of the essential mycobacterial cell wall in the otherwise defective *Mycobacterium leprae*. *Lepr Rev*, 2001; **72:** 415–428.
[33] Zhang Y, Heym B, Allen B *et al*. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis. Nature*, 1992; **358:** 591–593.
[34] Johnsson K, Schultz PG. Mechanistic studies of the oxidation of isoniazid by the catalase-peroxidase from *Mycobacterium tuberculosis. J Am Chem Soc*, 1994; **116:** 7425–7426.
[35] Banerjee A, Dubnau E, Quémard A *et al*. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis. Science*, 1994; **263:** 227–230.

[36] Rozwarski DA, Grant GA, Barton DHR *et al.* Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis. Science*, 1998; **279:** 98–102.
[37] Slayden RA, Barry CE 3rd. The genetics and biochemistry of isoniazid resistance in mycobacterium tuberculosis. *Microbes Infect*, 2000; **2:** 659–669.
[38] Slayden RA, Lee RE, Barry CE 3rd. Isoniazid affects multiple components of the type II fatty acid synthase system of *Mycobacterium tuberculosis. Mol Microbiol*, 2000; **38:** 514–525.
[39] Eiglmeier K, Fsihi H, Heym B, Cole ST. On the catalase-peroxidase gene, katG, of *Mycobacterium leprae* and the implications for treatment of leprosy with isoniazid. *FEMS Microbiol Lett*, 1997; **149:** 273–278.
[40] Nakata N, Matsuoka M, Kashiwabara Y *et al.* Nucleotide sequence of the *Mycobacterium leprae* katG region. *J Bacteriol*, 1997; **179:** 3053–3057.
[41] Menendez MC, Ainsa JA, Martin C, Garcia MJ. katGI and katGII encode two different catalases-peroxidases in *Mycobacterium fortuitum. J Bacteriol*, 1997; **179:** 6880–6886.
[42] Cartel JL, Millan J, Guelpa-Lauras CC, Grosset JH. Hepatitis in leprosy patients treated by a daily combination of dapsone, rifampin, and a thioamide. *Int J Lepr Other Mycobact Dis*, 1983; **51:** 461–465.
[43] DeBarber AE, Mdluli K, Bosman M *et al.* Ethionamide activation and sensitivity in multidrug-resistant *Mycobacterium tuberculosis. Proc Natl Acad Sci USA*, 2000; **97:** 9677–9682.
[44] Baulard AR, Betts JC, Engohang-Ndong J *et al.* Activation of the pro-drug ethionamide is regulated in mycobacteria. *J Biol Chem*, 2000; **275:** 28326–28331.
[45] Khoo KH, Douglas E, Azadi P *et al.* Truncated structural variants of lipoarabinomannan in ethambutol drug-resistant strains of *Mycobacterium smegmatis.* Inhibition of arabinan biosynthesis by ethambutol. *J Biol Chem*, 1996; **271:** 28682–28690.
[46] Telenti A, Philipp WJ, Sreevatsan S *et al.* The emb operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nature Med*, 1997; **3:** 567–570.
[47] Belanger AE, Besra GS, Ford ME *et al.* The embAB genes of Mycobacterium avium encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc Natl Acad Sci USA*, 1996; **93:** 11919–11924.
[48] Sreevatsan S, Stockbauer KE, Pan X *et al.* Ethambutol resistance in *Mycobacterium tuberculosis:* critical role of embB mutations. *Antimicrob Agents Chemother*, 1997; **41:** 1677–1681.
[49] Lety MA, Nair S, Berche P, Escuyer V. A single point mutation in the embB gene is responsible for resistance to ethambutol in *Mycobacterium smegmatis. Antimicrob Agents Chemother*, 1997; **41:** 2629–2633.
[50] Zimhony O, Cox JS, Welch JT *et al.* Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of *Mycobacterium tuberculosis. Nature Med*, 2000; **6:** 1043–1047.
[51] Scorpio A, Zhang Y. Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nature Med*, 1996; **2:** 662–667.
[52] Boshoff HI, Mizrahi V. Purification, gene cloning, targeted knockout, overexpression, and biochemical characterization of the major pyrazinamidase from *Mycobacterium smegmatis.* J Bacteriol, 1998; **180:** 5809–5814.
[53] Boshoff HI, Mizrahi V. Expression of *Mycobacterium smegmatis* pyrazinamidase in *Mycobacterium tuberculosis* confers hypersensitivity to pyrazinamide and related amides. *J Bacteriol*, 2000; **182:** 5479–5485.
[54] Zhang Y, Scorpio A, Nikaido H, Sun Z. Role of acid pH and deficient efflux of pyrazinoic acid in unique susceptibility of *Mycobacterium tuberculosis* to pyrazinamide. *J Bacteriol*, 1999; **181:** 2044–2049.
[55] Pym AS, Cole ST. Tuberculosis chemotherapy – from conception to genomics. In: Wax R, Lewis K, Salyers A, Taber H (eds) *Bacterial resistance to antimicrobials: mechanisms, genetics, medical practice and public health.* Marcel Dekker, New York, 2001, pp. 355–403.
[56] Gelber RH. Further studies of the killing of M. leprae by aminoglycosides: reduced dosage and frequency of administration. *Int J Lepr Other Mycobact Dis*, 1987; **55:** 78–82.
[57] Honoré N, Cole ST. Streptomycin and mycobacteria. *Antimicrob Agents Chemother*, 1994; **38:** 238–242.
[58] Meier A, Kirschner P, Bange F-C *et al.* Genetic alterations in streptomycin-resistant *Mycobacterium tuberculosis:* Mapping of mutations conferring resistance. *Antimicrob Agents Chemother*, 1994; **38:** 228–233.
[59] Suzuki Y, Katsukawa C, Tamaru A *et al.* Detection of kanamycin-resistant *Mycobacterium tuberculosis* by identifying mutations in the 16S rRNA gene. *J Clin Microbiol*, 1998; **36:** 1220–1225.
[60] Burman WJ, Stone BL, Brown BA *et al.* AIDS-related *Mycobacterium kansasii* infection with initial resistance to clarithromycin. *Diagn Microbiol Infect Dis*, 1998; **31:** 369–371.
[61] Meier A, Heifets L, Wallace RJ Jr *et al.* Molecular mechanisms of clarithromycin resistance in *Mycobacterium avium:* observation of multiple 23S rDNA mutations in a clonal population. *J Infect Dis*, 1996; **174:** 354–360.
[62] Ji B, Perani EG, Petinom C, Grosset JH. Bactericidal activities of combinations of new drugs against *Mycobacterium leprae* in nude mice. *Antimicrob Agents Chemother*, 1996; **40:** 2137–2141.
[63] Speer BS, Shoemaker NB, Salyers AA. Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. *Clin Microbiol Rev*, 1992; **5:** 387–399.
[64] Franzblau SG, Chan GP, Garcia-Ignacio BG *et al.* Clinical trial of fusidic acid for lepromatous leprosy. *Antimicrob Agents Chemother*, 1994; **38:** 1651–1654.

[65] Franzblau SG, Biswas AN, Harris EB. Fusidic acid is highly active against extracellular and intracellular *Mycobacterium leprae*. *Antimicrob Agents Chemother*, 1992; **36:** 92–94.

[66] Johanson U, Hughes D. Fusidic acid-resistant mutants define three regions in elongation factor G of *Salmonella typhimurium*. *Gene*, 1994; **143:** 55–59.

[67] Ji B, Grosset J. Combination of rifapentine-moxifloxacin-minocycline (PMM) for the treatment of leprosy. *Lepr Rev*, 2000; **71:** S81–87.

[68] Cambau E, Perani E, Guillemin I *et al*. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet*, 1997; **349:** 103–104.

[69] Takiff HE, Salazar L, Guerrero C *et al*. Cloning and nucleotide sequence of the *Mycobacterium tuberculosis* gyrA and gyrB genes, and characterization of quinolone resistance mutations. *Antimicrob Agents Chemother*, 1994; **38:** 773–780.

[70] Cole ST, Brosch R, Parkhill J *et al*. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998; **393:** 537–544.

[71] Fsihi H, Vincent V, Cole ST. Homing events in the gyrA gene of some mycobacteria. *Proc Natl Acad Sci USA*, 1996; **93:** 3410–3415.

[72] Davis EO, Thangaraj HS, Brooks PC, Colston JM. Evidence of selection for protein introns in the RecAs of pathogenic mycobacteria. *EMBO J*, 1994; **13:** 699–703.

[73] Ainsa JA, Perez E, Pelicic V *et al*. Aminoglycoside 2'-N-acetyltransferase genes are universally present in mycobacteria: characterization of the aac(2′)-Ic gene from *Mycobacterium tuberculosis* and the aac(2')-Id gene from *Mycobacterium smegmatis*. *Mol Microbiol*, 1997; **24:** 431–441.

[74] Chambers HF, Moreau D, Yajko D *et al*. Can penicillins and other beta-lactam antibiotics be used to treat tuberculosis? *Antimicrob Agents Chemother*, 1995; **39:** 2620–2624.

[75] Voladri RK, Lakey DL, Hennigan SH *et al*.. Recombinant expression and characterization of the major beta-lactamase of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 1998; **42:** 1375–1381.

[76] Shepard CC, Van Landingham RM, Walker LL, Good RC. Activity of selected beta-lactam antibiotics against *Mycobacterium leprae*. *Int J Lepr Other Mycobact Dis*, 1987; **55:** 322–327.

[77] Randhawa B, Harris EB, Prabhakaran K. Bactericidal action of oral ampicillin/sulbactam against *Mycobacterium leprae*. *J Antimicrob Chemother*, 1999; **44:** 279–281.

[78] Gelber RH. The activity of amoxicillin plus clavulanic acid against *Mycobacterium leprae* in mice. *J Infect Dis*, 1991; **163:** 1374–1377.

[79] Tseng T-T, Gratwick KS, Kollmann J *et al*. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and develoment proteins. *J Mol Microbiol Biotechnol*, 1999; **1:** 107–125.

[80] Braibant M, Gilot P, Content J. The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev*, 2000; **24:** 449–467.

[81] Guilfoile PG, Hutchinson CR. A bacterial analog of the mdr gene of mammalian tumor cells is present in *Streptomyces peucetius*, the producer of daunorubicin and doxyrubicin. *Proc Natl Acad Sci USA*, 1991; **88:** 8553–8557.

[82] Ainsa JA, Blokpoel MCJ, Otal I *et al*. Molecular cloning and characterization of Tap, a putative multidrug efflux pump present in *Mycobacterium fortuitum* and *Mycobacterium tuberculosis*. *J Bacteriol*, 1998; **180:** 5836–5843.

[83] Silva PE, Bigi F, de la Paz Santangelo M *et al*. Characterization of P55, a multidrug efflux pump in *Mycobacterium bovis* and *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 2001; **45:** 800–804.

[84] Doran JL, Pang Y, Mdluli KE *et al*. *Mycobacterium tuberculosis* efpA encodes an efflux protein of the QacA transporter family. *Clin Diag Lab Immunol*, 1997; **4:** 23–32.

[85] Wilson M, de Risi J, Kristensen H-K *et al*. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA*, 1999; **96:** 6854–6859.

[86] de Rossi E, Branzoni M, Cantoni R *et al*. mmr, a *Mycobacterium tuberculosis* gene conferring resistance to small cationic dyes and inhibitors. *J Bacteriol*, 1998; **180:** 6068–6071.

[87] Stover CK, Pham XQ, Erwin AL *et al*. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature*, 2000; **406:** 959–964.

[88] Li XZ, Nikaido H, Poole K. Role of mexA-mexB-oprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrob Agents Chemother*, 1995; **39:** 1948–1953.

[89] Ma D, Cook DN, Alberti M *et al*. Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli*. *Mol Microbiol*, 1995; **16:** 45–55.

[90] Camacho LR, Ensergueix D, Perez E *et al*. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol*, 1999; **34:** 257–267.

[91] Cox JS, Chen B, McNeil M, Jacobs WR Jr. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. *Nature*, 1999; **402:** 79–83.

[92] Stover CK, Warrener P, VanDevanter DR *et al*. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. *Nature*, 2000; **405:** 962–966.

[93] Grebe TW, Stock JB. The histidine protein kinase superfamily. *Adv Microb Physiol*, 1999; **41:** 139–227.

[94] Fsihi H, De Rossi E, Salazar L *et al*. Gene arrangement and organisation in a ~76 kilobase fragment encompassing the oriC region of the chromosome of *Mycobacterium leprae*. *Microbiology*, 1996; **142:** 3147–3161.

[95] Barrett JF, Goldschmidt RM, Lawrence LE *et al*. Antibacterial agents that inhibit two-component signal transduction systems. *Proc Natl Acad Sci USA*, 1998; **95:** 5317–5322.

[96] Hlasta DJ, Demers JP, Foleno BD *et al*. Novel inhibitors of bacterial two-component systems with gram positive antibacterial activity: pharmacophore identification based on the screening hit closantel. *Bioorg Med Chem Lett*, 1998; **8:** 1923–1928.

Lepr Rev (2201) **72,** 441–448

A method for rapid detection of rifampicin-resistant isolates of *Mycobacterium leprae*

NADINE HONORÉ*, PAUL W. ROCHE**,
JACQUES H. GROSSET*** & STEWART T. COLE*
**Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, Cedex 15, France*
***Mycobacterial Research Laboratory, The Leprosy Mission, Anandaban Leprosy Hospital, PO Box 151, Kathmandu, Nepal*
****SAMU Social de Paris, 35 Avenue Courteline, 75012 Paris, France*

Summary A genotypic method for predicting rifampicin resistance in *Mycobacterium leprae* has been developed and rigorously tested on mouse footpad-derived and clinical specimens. A series of immobilized oligonucleotide capture probes can discriminate between wild type and mutant *rpoB* alleles, and positive controls are available for the most frequent mutation affecting Ser425. Two different non-radioactive detection formats have been tested with comparable success in both an industrialized and a developing country. The standardized procedure could now be used in a prospective study of potential rifampicin resistance among multibacillary patients.

Introduction

Rifampicin is the backbone of the multidrug therapy currently used to treat leprosy[1,2] and it, or a related rifamycin, will almost certainly be included in any new regimens that may be developed.[3] Fortunately, resistance to rifampicin is rare and has only been documented in a limited number of lepromatous patients, mostly receiving monotherapy or other inadequate treatment.[4] On characterization, all rifampicin resistant isolates of *Mycobacterium leprae* were found to harbour mis-sense mutations in a short segment of the *rpoB* gene, encoding the β-subunit of DNA-dependent RNA polymerase.[5–8] The same region, termed RIF-RDR has been shown to be associated with rifampicin resistance in almost all resistant isolates, of *Mycobacterium tuberculosis*,[7,9–12] and in a range of other bacteria.[13,14]

Testing *M. leprae* isolates for rifampicin susceptibility is generally performed after inoculation of mice[15] but this is expensive and requires specialized resources that are gradually disappearing from leprosy control programmes. Radio-respirometry is a useful, more rapid alternative[16] but this also needs a sophisticated laboratory infrastructure and

* Correspondence to: Prof. S. T. Cole, Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France. Phone: 33-1-45 68 84 46. Fax: 33-1-40 61 35 83. E-mail: stcole@pasteur.fr

0305-7518/01/072441+08 $1.00

relatively large numbers of metabolically active organisms that can only be obtained by passage and growth in experimental animals. In recent years, several groups have established predictive molecular methods for detecting rifampicin resistance that are based on the *rpoB* genotype. The first such approach involved single strand conformation polymorphism (SSCP)-PCR,[6] yielding excellent results within 48 h and giving perfect agreement with those obtained from the mouse. Similar results were subsequently obtained using PCR-facilitated heteroduplex analysis (HDA) of *rpoB*[7] and this method has also been applied to detecting dapsone resistance.[17] However, both PCR-SSCP and PCR-HDA require gel electrophoresis and sophisticated laboratory equipment and may not be suitable for resource-poor settings. Here we describe the development of a simpler method for predicting rifampicin resistance, involving solid-phase hybridization to oligonucleotide capture probes, and its application in the field.

Materials and methods

MEMBRANE PREPARATION

Oligonucleotides corresponding to both strands of the RIF-RDR region were designed (Figure 1, Table 1) then tested empirically for use as capture probes. The best results were obtained with probes corresponding to the complementary strand so these were then used exclusively. Capture probes were synthesized with a 5′-amino group to facilitate covalent linkage to Biodyne C membranes and diluted to the optimized concentrations in 350 μl of 500 mM $NaHCO_3$, pH 8.4. Membranes (9·5 × 7 cm) were activated by incubation in 20 ml freshly prepared 16% (w/v) 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDAC), with shaking at room temperature for 10 min, rinsed with water and placed in a clean miniblotter system. Diluted oligonucleotides solutions (350 μl) were added to the slots and left for 5 min before removal. The membrane was removed from the miniblotter, inactivated in 100 mM NaOH for 10 min, then washed in 2 × SSPE/0·1% SDS for 5 min at 60°C. After washing in 20 mM EDTA pH 8, for 15 min, membranes were stored at 4°C in Saranwrap, then cut into appropriate strips before use.

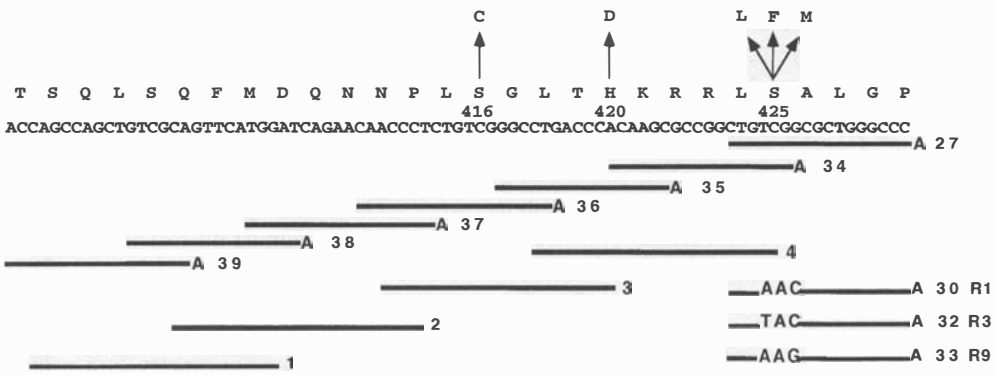

Figure 1. Organisation and sequence of the RIF-RDR of *rpoB* from *M. leprae* and location of capture probes. The upper part shows the DNA sequence and predicted protein sequence together with mutations known or suspected to confer rifampicin resistance. The lower part shows the position of the probes discussed in the text; their sequences may be found in Table 1.

Table 1. Oligonucleotides used in this study*

For PCR and sequencing	
Brpo22	bCAGGACGTCGAGGCGATCAC
rpo32	TCCTCGTCAGCGGTCAAGTA
rpo46	TCGATCGGGCACATCCGGC
Capture probes	
S1	aCCATGAACTGCGACAGCTGGCTG
S2	aGGGTTGTTCTGATCCATGAACTG
S3	aTGGGTCAGGCCCGACAGAGGGT
S4	aGACAGCCGGCGCTTGTGGGTCAG
A27	aAAAGGGCCCAGCGCCGACAG
A30R1	aAAAGGGCCCAGCGCCAACAG
A32R3	aAAAGGGCCCAGCGCCATCAG
A33R9	aAAAGGGCCCAGCGCGAACAG
A34	aAAACGACAGCCGGCGCTTGT
A35	aAAAGCTTGTGGGTCAGGCCC
A36	aAAAAGGCCCGACAGAGGGTT
A37	aAAAAGGGTTGTTCTGATCCA
A38	aAAAGATCCATGAACTGCGAC
A39	aAAATGCGACAGCTGGCTGGT

a, denotes the position of the 5′-amino link group; b, denotes biotin.

PREPARATION OF BIOTINYLATED PROBES AND DNA SEQUENCING

PCR was performed using the primers B-rpo22 and rpo32 (Table 1), where B-rpo22 is biotinylated, with samples of *M. leprae* DNA prepared by the freeze-boiling method,[18] as described previously.[5] For control purposes, plasmids carrying cloned RIF-RDR from wild type and three rifampicin resistant isolates of *M. leprae* were used in PCR reactions. In some experiments, DNA sequencing was performed to confirm the presence of mutations predicted by the kit. Primers B-rpo22 and rpo32 were used to generate a PCR fragment spanning the RIF-RDR, which was then sequenced using primer rpoB46 (Table 1) and fluorescent big dye terminators. Sequences were obtained using an ABI373 automated sequencer.

HYBRIDIZATION AND DETECTION

This requires the following prewarmed buffers (20°C, unless specified otherwise): denaturation buffer (DB) 2 M NaOH/2 mM EDTA; hybridization buffer (HB) 6×SSPE/0·1% SDS, 50°C; washing buffer (WB1) 4×SSPE/0·1% SDS, 55°C,and 42°C; washing buffer (WB2) 2×SSPE/0·1% SDS; washing buffer (WB3) 2×SSPE; staining buffer (SB) 0·1 M Tris-buffer pH 9·5/0·05 M $MgCl_2$, 0·1 M NaCl. PCR fragments were denatured in an equal volume of DB for 5 min then added to prewarmed HB (1 ml, 55°C), and incubated with the strips, in a shaking water-bath at 50°C, for 30–60 min. After discarding the hybridization solution and washing the strips with 2 ml of prewarmed WB1 for 15 min, the appropriate streptavidin-conjugate in WB1 (2 ml, 42°C) was added and incubated for 30–60 min. Streptavidin-peroxidase conjugate was diluted 1/4000 for chemiluminescent detection and streptavidin-alkaline phosphatase conjugate was diluted 1/1000 for colorimetric detection. After discarding the conjugate solution, strips were washed with 2 ml of WB2 for 10 min

before adding 2 ml of WB3 and washing for 10 min. For chemiluminescent detection, strips were incubated for 2 min in 2 ml ECL detection liquid (Amersham), washed, then covered with Saran-wrap and exposes to X-ray film for 1 min. For colorimetric detection, 2 ml of diluted NBT/BCIP stock solution in SB (200 μl/10 ml), was added and the strips incubated for 15–60 min.

Results

STANDARDIZING THE METHOD

The aim of this work was to construct and standardize a kit that allows mutations to be detected in the *rpoB* gene of *M. leprae*, and hence to predict whether the corresponding isolates would be sensitive or resistant to rifampicin. The approach involved the synthesis of a series of oligonucleotide capture probes spanning the lower strand of the RIF-RDR, their immobilization on a nylon membrane and use in reverse hybridization experiments. Many oligonucleotide probes of different lengths spanning RIF-RDR were synthesized, bound to the membrane then tested with biotinylated probes prepared from wild type and mutant *rpoB* alleles by PCR. The optimal combination of oligonucleotides, their sequence and location is shown in Figure 1. We found that the best results were obtained by combining four 23-mer probes with six (or seven) 17-mer probes for the bottom strand. Hybridization results were uneven in initial experiments with the shorter probes, probably due to steric hindrance from the membrane resulting in less of the capture probe being available to form hybrids. Consequently, these probes were resynthesized with three additional deoxyadenosine residues at the 5′-end thereby allowing better access of the biotinylated PCR fragment to the 17-mer capture sequence.

To date, only a limited number of mutations responsible for rifampicin resistance have been reported in *M. leprae*.[5,7] The amino acid substitutions involved are His420Asp (position 525 in *Escherichia coli*), Ser425Leu, Ser425Met and Ser425Phe. For control purposes, 17-mer capture probes (with 5′-adenosine) corresponding to mutant *rpoB* alleles of Ser425 were also synthesized and included on the membranes. This proved particularly useful as the DNA sequence encompassing Ser425 is very GC-rich and readily adopts stable secondary structures that often preclude hybridization.

Typical results obtained with this method are shown in Figure 2. Hybridization to capture probes 1–4 is seen in all cases except when a mutation affecting the codon for His420 is present as probe 4 then displayed no signal (column 5). Mutations affecting this codon can also be detected with probe A35. Probe 4 also covers part of the mutational hot-spot corresponding to codon 425 but does not detect mutations affecting this position probably because the site is too close to the 5′-end of the capture probe for heteroduplex formation to be sufficiently destabilized. However, the shorter probe A27, in combination with the mutant probes A30R1, A32R3 and A33R9, allows all known mutations affecting codon 425 to be detected efficiently (Figure 2, columns 2, 3 and 4). The mutant probes displayed excellent specificity as no signals due to crosshybridization with the wild type *rpoB* gene were obtained. Probes A34 and A37 sometimes displayed weaker signals than the others but the corresponding region are well covered by probes A27 and 4, or 2 and 3, respectively (Figure 1). Mutations were correctly revealed under standard laboratory conditions using either colorimetric or chemiluminescent detection methods.

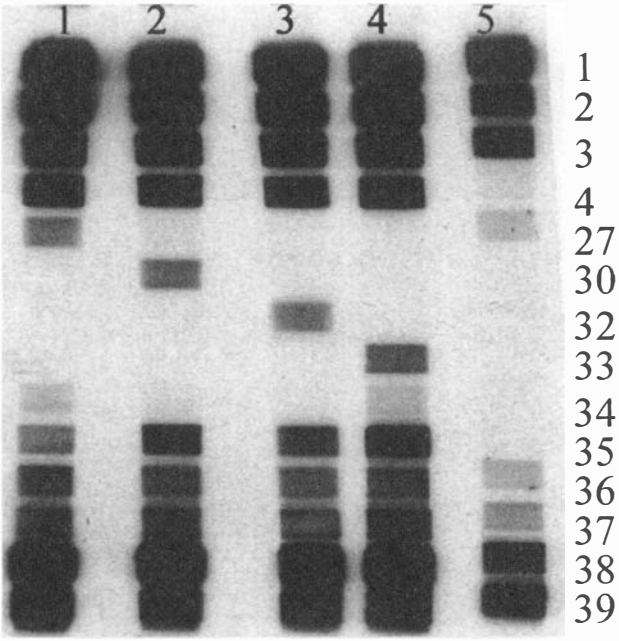

Figure 2. Detection of rifampicin resistance by reverse hybridization. The results were obtained using the protocol described in Materials and methods with biotinylated probes generated by PCR using *M. leprae* DNA from the following sources: column 1, wild type *M. leprae*; column 2, positive control *rpoB*Ser425Leu; column 3, positive control *rpoB*Ser425Met; column 4, positive control *rpoB*Ser425Phe; column 5, strain 92041 with *rpoB*His420Asp mutation. The capture probes are indicated on the right, for further details see Table 1 or Figure 1.

BLIND TESTING

Having established that the method reproducibly detected the common *rpoB* mutations at our disposal, it was then evaluated blindly on a set of six uncharacterized *M. leprae* isolates. Five of these proved to be wild type for *rpoB* (represented in column 2, Figure 3), but one of them appeared to harbour an *rpoB* mutation as no signal was obtained with capture probes 3 and A36 (column 1, Figure 3). These results were confirmed by DNA sequencing of PCR fragments spanning the RIF-RDR as the missense mutation Ser416Cys (TCG->TGT) was uncovered. Mutation of this residue to Leu and Ser has been reported in rifampicin resistant isolates of *M. tuberculosis* and *E. coli*, respectively.[9,11,13]

FIELD TESTING

Kits for detecting rifampicin resistance were established in Paris then tested in Kathmandu at the Anandaban Leprosy Hospital. The results obtained there were of comparable quality to those generated in France as were those obtained with membranes produced locally. A potential case of rifampicin resistance was successfully detected in a mouse footpad-derived specimen. The corresponding strain was predicted to harbour a Ser425Phe mutation by solid-phase hybridization and this was confirmed by DNA sequencing. An *M. leprae* isolate harboring the corresponding allele has previously been shown to be drug resistant in the

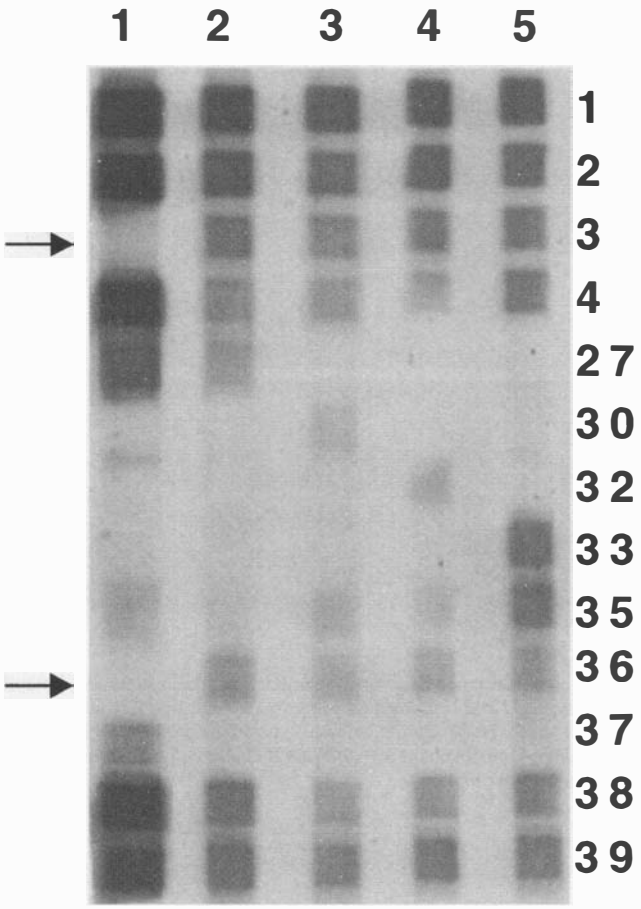

Figure 3. Detection of rifampicin resistance by reverse hybridization. As stated in legend to Figure 2 except: column 1, strain 25200 with *rpoB*Ser416Cys mutation; column 2, wild type *M. leprae*; column 3, positive control *rpoB*Ser425Leu; column 4, positive control *rpoB*Ser425Met; column 5, positive control *rpoB*Ser425Phe.

mouse model.[4,5] Having validated the system, attempts were then made to apply it to clinical specimens obtained from leprosy patients.

Skin biopsies were collected from ~60 lepromatous leprosy patients in Nepal and India, stored in 70% ethanol, then transported to the Anandaban Leprosy Hospital for further analysis. Most of these patients had either relapsed or not responded to treatment. After rehydration and mincing, the biopsies were prepared for PCR analysis by freeze-boiling. Only five samples reproducibly yielded PCR fragments and these were used successfully with the chemiluminescent detection system. Two samples from patients who had not responded to treatment were found to be resistant by both methods and harboured the Ser425Phe mutation. The remaining three samples were found to be wild type, or rifampicin-susceptible, by both methods.

Discussion

This study has convincingly demonstrated that detection of rifampicin resistance by molecular methods is a feasible and practical alternative to the mouse footpad assay. Equivalent results were obtained using two different non-radioactive detection systems in laboratories in France and Nepal. There is thus no technical reason why a kit such as that described here should not find more extensive application in developing countries. A similar method has been developed for detecting rifampicin resistant *M. tuberculosis* and is now commercially available.[19] It has yielded satisfactory results in several settings.[10] The availability of the *M. leprae* genome sequence also provides us with means to develop genotypic methods for detecting resistance to other drugs. This could now be done for dapsone and fluoroquinolones and, in the near future for other therapeutic agents as information about the molecular basis of resistance becomes available.

We believe that the time is now right to extend the study and to learn from the present work as this approach has much to offer leprosy control programmes. In the next phase, a much larger number of specimens should be tested. Although every available mutation has been detected by the reverse hybridization system, the number of sites within *rpoB* that can be monitored by positive controls is limited by the restricted set of base changes described.[5–7] As a result, the faint possibility that new mutations might be missed by the system cannot yet be excluded and more extensive evaluation is required.

This could be done in the framework of a prospective study of rifampicin resistance in which skin biopsies are examined from ex-multibacillary leprosy patients who have not responded properly to the treatment of MDT. Combined with PCR-based, DNA sequence analysis of the *rpoB* gene from *M. leprae* extracted from the biopsies rigorous appraisal of the reverse hybridization kit would be obtained and valuable information about the level of rifampicin resistance generated. Such a study would also enable us to evaluate, in a systematic manner, the long-term efficacy of multidrug therapy for multibacillary leprosy.

Acknowledgements

We wish to acknowledge the financial support of the Institut Pasteur, the Association Française Raoul Follereau, and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

References

[1] Grosset JH, Ji B. Recent advances in the chemotherapy of leprosy. *Lepr Rev*, 1990; **61:** 313–329.
[2] WHO Study Group. *Chemotherapy of leprosy for control programmes.* World Health Organisation, Geneva, Tech. Rep. Ser. 675, 1982.
[3] Ji B, Grosset J. Combination of rifapentine-moxifloxacin-minocycline (PMM) for the treatment of leprosy. *Lepr Rev*, 2000; **71:** S81–7.
[4] Grosset JH, Guelpa-Lauras CC, Bobin P *et al.* Study of 39 documented relapses of multibacillary leprosy after treatment with rifampin. *Int J Lepr*, 1989; **57:** 607–614.
[5] Honoré N, Cole ST. Molecular basis of rifampin resistance in *Mycobacterium leprae*. *Antimicrob Agents Chemother*, 1993; **37:** 414–418.
[6] Honoré N, Perani E, Telenti A *et al.* A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. *Int J Lepr*, 1994; **61:** 600–604.
[7] Williams DL, Waguespack C, Eisenach K *et al.* Characterization of rifampin resistance in pathogenic mycobacteria. *Antimicrob Agents Chemother*, 1994; **38:** 2380–2386.

[8] Cambau E, Perani E, Guillemin I *et al*. Multidrug resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet*, 1997; **349:** 103–104.
[9] Telenti A, Imboden P, Marchesi F *et al*. Detection of rifampicin resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 1993; **341:** 647–650.
[10] Telenti A, Honoré N, Bernasconi C *et al*. Genotypic assessment of isoniazid and rifampin resistance in *Mycobacterium tuberculosis:* a blind study at reference laboratory level. *J Clin Microbiol*, 1997; **35:** 719–723.
[11] Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis:* 1998 update. *Tuber Lung Dis*, 1998; **79:** 3–29.
[12] Miller LP, Crawford JT, Shinnick TM. The rpoB gene of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 1994; **38:** 805–811.
[13] Jin DJ, Gross CA. Mapping and sequencing of mutations in the *Escherichia coli* rpoB gene that lead to rifampicin resistance. *J Mol Biol*, 1988; **202:** 45–58.
[14] Carter PE, Abadi FJ, Yakubu DE, Pennington TH. Molecular characterization of rifampin resistant *Neisseria meningitidis*. *Antimicrob Agents Chemother*, 1994; **38:** 1256–1261.
[15] Shephard CC, Chang YT. Effect of several anti-leprosy drugs on multiplication of human leprosy bacilli in footpads of mice. *Proc Soc Exp Biol Med*, 1962; **109:** 636–638.
[16] Franzblau S. Drug susceptibility testing of *Mycobacterium leprae* in the BACTEC 460 system. *Antimicrob Agents Chemother*, 1989; **33:** 2115–2117.
[17] Williams DL, Pittman TL, Gillis TP *et al*. Simultaneous detection of *Mycobacterium leprae* and its susceptibility to dapsone using DNA heteroduplex analysis. *J Clin Microbiol*, 2001; **39:** 2083–2088.
[18] Woods SA, Cole ST. A rapid method for the detection of potentially viable *Mycobacterium leprae* in human biopsies: a novel application of PCR. *FEMS Microbiol Lett*, 1989; **65:** 305–310.
[19] Beenhouwer De H, Lhiang Z, Jannes G *et al*. Rapid detection of rifampicin resistance in sputum and biopsy specimens from tuberculosis patients by PCR and line probe assay. *Tuberc Lung Dis*, 1995; **76:** 425–430.

Lepr Rev (2001) **72,** 449–461

Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity

STEWART T. COLE*, PHILIPPE SUPPLY** & NADINE HONORÉ*

**Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris, Cedex 15, France*

***Institut National de la Sante et de la Recherche Medicale, U447, Institut Pasteur de Lille, 1 Rue du Professor Calmette, F-59019 Lille Cedex, France*

Summary About 2% of the genome of *Mycobacterium leprae* is composed of repetitive DNA. There are more than 26 extinct IS elements together with four families of dispersed repeats, present in five copies or more, RLEP (37 copies), REPLEP (15 copies), LEPREP (eight copies), and LEPRPT (five copies). Although there is no sequence similarity to known transposable elements, RLEP occurs predominantly at the 3′-end of genes and, in several cases, within pseudogenes, suggesting that it was capable of dissemination. Strikingly, on comparison of the genome sequences of *M. leprae* and the closely related tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, many of these repetitive sequences were found at sites of discontinuity in gene order. Evidence is presented that loss of synteny, inversion and genome downsizing may have resulted from recombination between dispersed copies of these repetitive elements.

Introduction

Repetitive sequences are common constituents of the genomes of all living organisms although they are far more prominent in higher eukaryotes where they can account for a substantial percentage of the chromosomal DNA. There are two principal forms of repetitive DNA in bacterial genomes: dispersed and tandem repeats. Dispersed repetitive sequences can correspond to duplicated genes, or to mobile genetic elements present in several copies like insertion sequences (IS). IS are often an important component of bacterial genomes and as a result of their ability to transpose have mutational potential based on their ability to locate within coding or regulatory regions. Hundreds of individual IS have been described and grouped into 17 families on the basis of their genetic organization, sequence similarities in their recombinases/transposases, the similarity of their ends (direct or terminal inverted repeats) and their target sites which are often duplicated during transposition.[1] The genome sequence of *Mycobacterium tuberculosis* H37Rv contains more than 56 IS, belonging to eight

* Correspondence to: Prof. S. T. Cole, Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France. Phone: 33-1-45 68 84 46. Fax: 33-1-40 61 35 83. E-mail: stcole@pasteur.fr

0305-7518/01/072449+13 $1.00

families and these are an important source of plasticity and genetic variability.[2–6] A novel repeated sequence, the REP13E12 family, is present in seven copies on the chromosome and contains a probable phage attachment site.[3] A large portion of the genome has also evolved from gene duplication events, followed by sequence divergence, leading to functional redundancy and expansion of the biological potential of the tubercle bacillus.[7]

Tandem repeats can be relatively simple, such as multiple repetitions of di- or tri-nucleotide sequences, or more complex such as the tandem duplication of large chromosomal segments,[8] like those described in *Mycobacterium bovis* BCG Pasteur.[9] Genetic variation is commonly associated with di- or tri-nucleotide repeats which are prone to amplification and contraction. These are often referred to as micro- or mini-satellites and are useful for typing purposes. One such mini-satellite that has been described in *M. tuberculosis* is the mycobacterial interspersed repetitive unit (MIRU), and this is also found in *M. leprae*.[10] A very promising epidemiological tool for tubercle bacilli has been developed that is based on variable number tandem repeats (VNTRs) of MIRU and this is capable of efficiently discriminating between outbreak strains.[11,12]

Here we describe the complete repertoire of repetitive DNA sequences identified in the genome of *M. leprae*[13] and discuss their potential impact on the evolution of the organism. In addition, attempts are being made to exploit some of these sequences for the development of a test that can distinguish between isolates of the leprosy bacillus.

Materials and methods

To identify repetitive DNA, the BLASTN program[14,15] was used to compare the genome sequence with itself. Areas showing >99% identity were then inspected visually and annotated using Artemis.[16] Potential IS elements were uncovered by database searches using BLASTX and tandem repeats identified using the program tandem repeats finder.[17] MIRUs were localized by BLASTN searches of the genome sequence using the consensus sequences of MIRU1-3 as strings, and all hits with scores >70 were investigated using a combination of Artemis the relational database, Lepromа.[18]

To investigate MIRU-based polymorphism, PCR primers were designed using the Oligo 5.0 software (National Biosciences, Plymouth, MN, USA), and the sequences are summarized in Table 1. *M. leprae* DNA was prepared by the freeze-boiling method.[19] For PCR reactions, 5 μl of DNA solutions was added to a final volume of 25 μl containing 10% DMSO, 0·5 mmol/l of each dATP, dCTP, dGTP and dTTP, 0·2 μmol/l of primers, 2·5 μl of PCR buffer [170 mmol/l $(NH4)_2SO_4$, 600 mmol/l Tris-HCL (pH 8·8), 20 mmol/l $MgCl_2$, 100 mmol/l β-mercaptoethanol] and 1·25 IU of *Taq* polymerase (Gibco-BRL). The PCR was performed using a PTC-100 (MJ Research, Inc.) for 35 cycles of 1 min at 94°C, 2 min at 59°C, 2 min at 72°C. The reactions were terminated by incubating for 10 min at 72°C and analysed by agarose gel electrophoresis using the appropriate controls.

Results and discussion

DIRECT REPEATS

On examination of the *M. leprae* genome sequence a series of perfect direct repeats was found ranging in size from 2 to 52 bp. All repeat sequences of >20 bp that were present in

Table 1. PCR primers used to study MIRU diversity

PRIMER F	SEQUENCE	PRIMER R	SEQUENCE	GENES
B937-MIRU1F	GTGCTGACCCGCTATCCTGA	B937-MIRU1R	CCCGCGACCCAGATTCTATC	*ML0534...carA*
B1308-MIRU1F	CGTTCTTGTGTGCGGGTGAGT	B1308-MIRU1R	TTACGACGCTGTTATGGAAACTGC	*ML0719...aldB*
B2235-MIRU1F	GCTGCGCCCGCGGTAGTCAC	B2235-MIRU1R	GAGGGGATGCCGACCATTTGG	*miaA...dapF*
B1764-MIRU1F	GGGCTTTCCATCGTCAACAG	B1764-MIRU1R	GCGTTAGGCACCCCAACA	*dut...ML1029*
B1764-MIRU2F	TTCACGCGAGTCCAGGTCAGAC	B1764-MIRU2R	CGTACGCAGGGAGGAGCAAAAC	*ML1042...ML1041*
L471-MIRU1F	CCAGGAGCCCACCAGGAC	L471-MIRU1R	GACGGCTGACATTGTCGGTCTAG	*ML1135...ML1136*
B1549-MIRU2F	TACCAGGAGCGGGATCGTAT	B1549-MIRU2R	CGGACGTGCTGACCATC	*cysM...ML1171*
B1549-MIRU1F	GTTCAGCGATACCAGCGTCA	B1549-MIRU1R	TCAGGGGACTGGTGAGGG	*rphA...ML1175*
B1133-MIRU1F	TGACGCTGGGTTTTTGGT	B1133-MIRU1R	GTTCGCGTGGAGTTCTTGTC	*argD...argF*
B2266-MIRU2F	GGAGGGAACTGGCAAGTCGT	B2266-MIRU2R	TGACCAGCCCAACAGACCTG	*ML2199...ML2200*
B2266-MIRU1F	TGAGCGGTCCACTAGCACAG	B2266-MIRU1R	CCGTCCAACGCGACTATCAC	*pabCML2203*
B2168-MIRU2F	CGCGGGTGGCTCGTAGAAGA	B2168-MIRU2R	TGACCGGCAAGCGACTTTGG	*ML2412...ML2413*
B2168-MIRU3F	TGGGCTCAAAACCTCCTTGC	B2168-MIRU3R	GGGCTGGCCATCGTCAAAC	*ML2439...ML2440*
B2168-MIRU1F	GGATGGCGTTGGTCTTGAG	B2168-MIRU1R	GCACTTTGGTGTTCGGACAT	*ML2442...ML2443*

two or more identical copies were annotated together with all perfect repeats present in three or more copies. There are far fewer tandem repeats in the genome of *M. leprae* compared to that of *M. tuberculosis*, mainly as a result of the much smaller number of PE-PGRS genes that are composed of such motifs. No tetranucleotide repeats are found in *M. leprae* nor in *M. tuberculosis* and dinucleotide repeats were only observed in the leprosy bacillus. A trinucleotide repeat (TTC) displaying copy number differences has been described in some isolates of the leprosy bacillus,[20] and this has 21 repetitions in the TN strain used for genome sequencing, while a hexanucleotide repeat in the *sigA* (*rpoT*) gene has been found recently to be present in three copies in most strains of *M. leprae*, including TN, but in four copies in others.[21] Promising discriminatory tests have been devised that target these polymorphisms,[20,21] and this encourages us to examine di- and trinucleotide repeats for variability in different isolates of *M. leprae*.

MYCOBACTERIAL INTERSPERSED REPETITIVE UNITS (MIRUS)

Prominent among the tandem repeats found in *M. tuberculosis*, are the MIRUs, as these can occur in from two to four tandem copies ranging in size from 46 to 101 bp, and are present at 41 loci.[10,12] MIRUs generally occur in intergenic regions and have the potential to encode small peptides as they contain short open reading frames whose start overlaps the stop codon of the upstream gene whereas the stop codon overlaps the iniation codon of the following gene. No MIRUs were detected by tandem repeat finder, but 20 single copies were found in the *M. leprae* genome by BLAST (Table 2). Eleven of these MIRUs had no counterparts in the *M. tuberculosis* genome, whereas four of the conserved loci contained multiple MIRUs in *M. tuberculosis* but only one in *M. leprae*.

To determine whether any of the sites harbouring MIRUs were occupied by multiple copies in other *M. leprae* strains, PCR primers were designed for 14 loci and used to screen a panel of 14 different isolates from Mali, Martinique, New Caledonia and the Philippines for diversity. In all cases, the size of the PCR fragments was consistent with the presence of a single MIRU, and when the DNA sequence was determined this was found to be identical to that of the TN strain. These results indicate that MIRUs are unlikely to represent a source of polymorphism in the leprosy bacillus, in contrast to the situation in *M. tuberculosis*.[11]

INSERTION SEQUENCES AND DUPLICATED GENES

Unlike *M. tuberculosis* H37Rv, which contains 56 IS elements, most of which are predicted to be functional,[3] *M. leprae* has only vestigial IS elements, as >26 transposase gene fragments were identified. These could not be classified reliably owing to the extensive levels of mutation and truncation incurred. All of these sequences appear in single copies.

Two identical copies of 16 dispersed repeats of >700 bp were detected and examined (Table 3). Three of these (1329, 1261 and 1179 bp) probably correspond to extinct IS, although it is interesting to note that despite loss of function their sequences are perfectly conserved, whereas two others (1054 and 753 bp) appear to be counterparts of the REP13E12 repeats described in the tubercle bacilli.[3] The remaining duplicated sequences correspond to genes or more rarely to pseudogenes (Table 3), and again it is unusual to find perfect conservation of the sequence in bacteria. This suggests that these duplication events may have occurred very recently or that sequence divergence occurs at an exceptionally slow rate in *M. leprae*.

Table 2. Features of MIRU in *M. leprae* and comparison with *M. tuberculosis*

Class	Position	Bases	Genes*	*M. tuberculosis*	Class (No.)°	Comments
MIRU2	153610...153666	57	*rfbe*...ML0112	*rfbE*...3781	–	
MIRU2	534710...534766	57	*scoA*...*scoB*	*scoA*...*scoB*	–	In-frame stop
MIRU1	648342...648425	84	ML0534...*carA*	Rv1382...*carA*	3 (2)	
MIRU1	862062...862114	>53	ML0719...*aldB*	Rv3292...*aldB*	–	Degenerate
MIRU3	877171...877223	53	*purK*...*purE*	*purK*...*purE*	3	Embedded in purK
MIRU2	1164265...1164320	57	*miaA*...*dapF*	*miA*...*dapF*	–	No ATG
MIRU1	1194835...1194886	52	*dut*...ML1029	*dut*...Rv2696c	1	Out-of-frame
MIRU2	1207116...1207172	57	ML1042...ML1041	Rv2680...*echA15*	2 (3)	–
MIRU2	1327764...1327823	60	ML1135...ML1136	Rv1300...Rv1301	2 (3)	–
MIRU1	1368453...1368519	67	*cysM*...ML1171	*cysM*...Rv1337	–	In-frame with ML1171
MIRU1	1371749...1371830	83	*rphA*...ML1175	*rphA*...Rv1341	2	Out-of-frame
MIRU2	1629692...1629746	55	ML1368...ML1369	Rv1709...Rv1710	2	Out-of-frame
MIRU2	1777816...1777872	57	ML1476...ML1475	Rv2454c...Rv2455c	–	–
MIRU1	1692100...1692180	81	*argD*...*argF*	*argD*...*argF*	–	
MIRU2	1915421...1915483	66	*pyrH*...*frr*	*pyrH*...*frr*	2 (2)	No ATG
MIRU2	2044283...2044339	57	*ilvb*...*ilvN*	*ilvB*...*ilvN*	–	No ATG
MIRU1	2613356...2613460	105	ML2199...ML2200	Rv0813c...Rv0814c	–	
MIRU1	2617793...2617873	81	*pabC*...ML2203	*pabC*...Rv0811c	–	–
MIRU2	2884042...2884095	54	ML2412...ML2413	Rv0525...Rv0526	2	No ATG
MIRU1	2917923...2918003	81	ML2442...ML2443	Rv0486...Rv0487	–	–

* Underlining indicates pseudogenes.
° –, denotes MIRU absent.

Table 3. Identical duplicated genes and sequences of >700 bp

Sequence	Genes*	Description
1329 bp	ML0040	Possible transposase remnant
1261 bp	ML1749	Possible transposase remnant
1179 bp	ML0444	Pseudogene similar to group II intron maturase
1054 bp	ML1290/ML1850	Pseudogenes orthologous to REP13E12 proteins
753 bp	ML1118/ML2236	Pseudogenes orthologous to REP13E12 proteins
1551 bp	*ansP1/ansP2*	L-asparagine transport proteins
1391 bp	ML2356/ML2357	Part of polyketide synthase
1219 bp	*mmpL2/mmpL4*	RND family transport proteins
1186 bp	ML0396/ML2692	Myo-inositol-1-phosphate synthase
1063 bp	ML1053/ML1183	PE proteins
1063 bp	ML1054/ML1182	PPE proteins
1006 bp	ML0125/ML0128	Putative glycosyl transferase
879 bp	ML1055/ML1180	QILSS family
879 bp	ML1056/ML1181	ESAT-6 family
871 bp	ML1047/ML1943	Pseudogene orthologous to Rv3714c
847 bp	*fadD5/fadD5*	Acyl-CoA synthase pseudogenes
740 bp	ML0447/ML2159	Similar to region of cytochrome P450s
704 bp	*umaA2/umaA1*	Mycolic acid synthase and pseudogene

* Underlining indicates pseudogenes.

Of particular interest are two regions of 1063 and 879 bp as these encode proteins of the PE, PPE, and ESAT-6 families.[2,7,22] In *M. tuberculosis* there are 11 regions containing ESAT-6 genes and these show two configurations comprising blocks of four or 10 conserved genes. There are three ESAT-6 regions of the larger type in *M. leprae* and two blocks of four genes. The latter consist of two identical repeat sequences of 1063 and 879 bp (Table 2; Figure 1). Interestingly, in one of these ESAT-6 regions, an additional unique sequence of 619 bp is present within the ML1182 gene, encoding a PPE protein, but one cannot tell whether this has been acquired by ML1182 or lost from ML1054. This is further evidence indicating that the ESAT-6 regions are dynamic[7,23] and that PPE proteins can undergo variation.[2,24]

RLEP

The RLEP element was intitially detected as a repetitive sequence in *M. leprae* by means of Southern blotting[25,26] and subsequently characterized at the molecular level by Woods

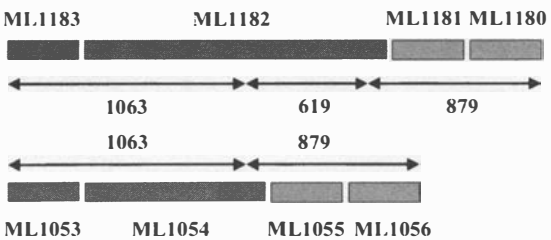

Figure 1. Organization of repeated loci encoding PE, PPE, and ESAT-6 proteins. Gene names are given and repeat sizes indicated in bp.

et al.[27] These authors estimated that there were at least 28 copies of RLEP in the genome and demonstrated that there was a central portion, common to all copies of RLEP, flanked by additional sequences whose presence was variable. With the complete genome sequence at our disposal we were able to perform the definitive bioinformatic analysis and this revealed that the TN strain of *M. leprae* contains 37 copies of RLEP, one of which, RLEP_29, lacks part of the central domain and will not be discussed further here. There is a conserved segment of 488 bp found in all intact copies of RLEP and this is flanked by additional sequences present in two or more independent RLEP elements. Consequently, the total length can vary from 601 to 1075 bp (Figure 2) and, as described previously, no open reading frames capable of coding for transposases, resolvases or other IS-associated functions could be found.[27] Further comparisons uncovered six polymorphic sites in the 488 bp conserved segment, three of which occurred only once while the remainder were found in numerous copies of RLEP. All of these polymorphisms can be accounted for by C-T transitions. On construction of a tree of RLEP sequences (Figure 2) by phylogenetic analysis using parsimony routines (PAUP), three large branches were established together with several outliers. However, there was little clear association between the length of the RLEP element or the presence of particular polymorphic nucleotides and its position in the tree (Figure 2). This is consistent with the complex organization of these sequences.

Roughly 1% of the chromosome is composed of RLEP DNA and these elements are distributed fairly randomly.[13] It is clear, however, that RLEP has contributed extensively to the remodelling of the *M. leprae* genome as copies are often found at breaks in synteny with *M. tuberculosis*. This will be discussed further below. There is a marked over-representation of RLEP elements, in either orientation, at the 3′-ends of genes, since ~30 of the copies are within 80 bp of the stop codon of the nearest gene. In several instances, RLEP is situated within the coding sequence at the 3′-end or overlapping the stop codon. Examples of this may be found in the *truA*, *truB* and *polA* genes. In two cases, important genes such as *glnA* and *polA* are flanked by inverted pairs of RLEPs in a configuration resembling that of a transposon. Data have been published that show that this composite *polA* structure is polymorphic between isolates of *M. leprae*[28] and it is conceivable that similar variability may also be associated with *glnA*. Transcription of *polA* may also have been impaired by RLEP,[29] as RLEP_22 is situated 6 bp upstream of the *polA* initiation codon. Some copies of RLEP are found within the sequences of pseudogenes that have intact functional orthologues in *M. tuberculosis*, notable examples are RLEP_8 in the spermidine biosynthetic gene, *speE*, RLEP_29 in the phosphoglucomutase gene, *pgmA*, and RLEP_28 in ML1722, a pseudogene orthologous to Rv3037c, a conserved hypothetical gene of *M. tuberculosis*. These observations suggest that RLEP may have been capable of transposition at one time although it is quite unclear how this was mediated. Furthermore, attempts to detect restriction fragment length polymorphisms linked to RLEP have revealed no diversity suggesting that RLEP is no longer capable of movement.[30]

REPLEP

There are 13 essentially intact sequences belonging to the REPLEP family and two large fragments (Figure 3). The largest elements are 881 bp long, with extensive complementarity between bases 1–95 and 783–880 (68% identity). REPLEP is bounded in most cases by an 8 bp inverted repeat (5′-GTTGTGGG) and contains no open reading frames. In several cases the inverted repeats continue past this octamer with certain REPLEP elements

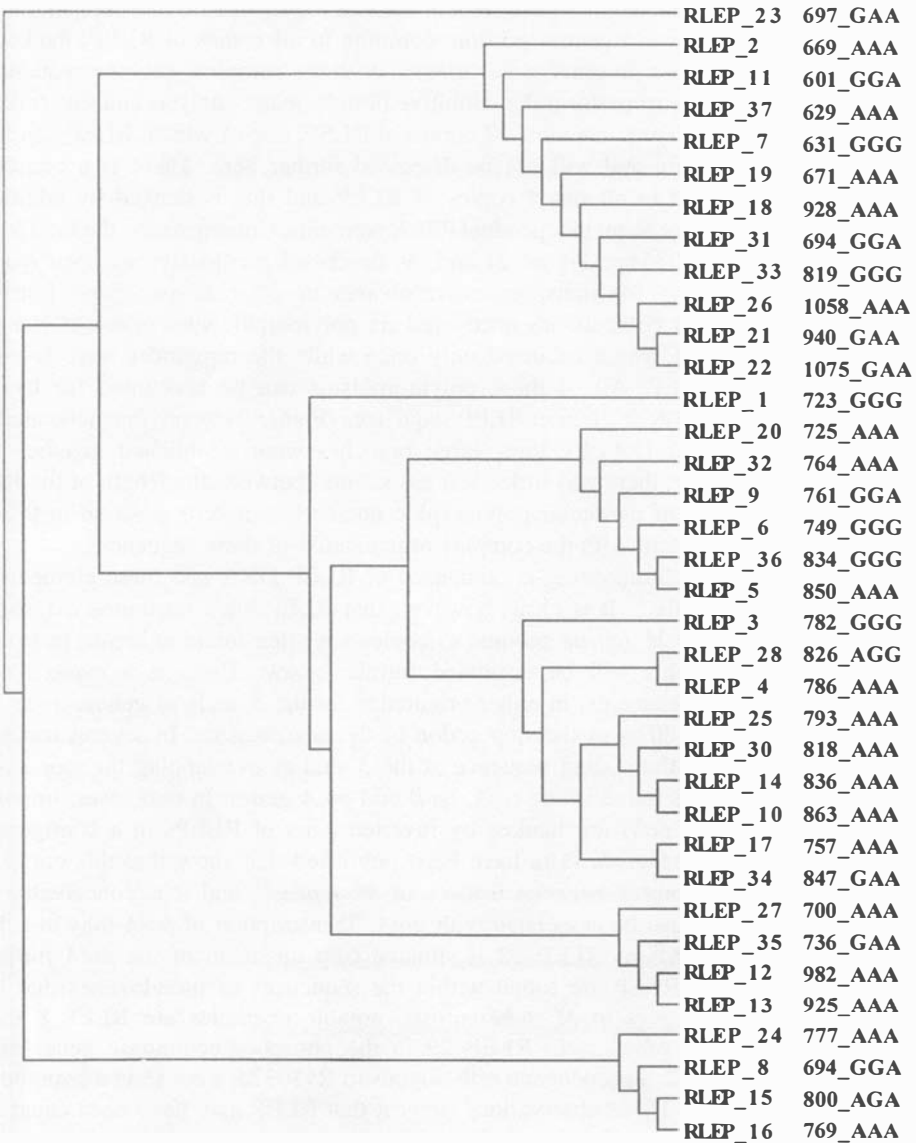

Figure 2. Phylogenetic tree of RLEP elements established using the phylogenetic analysis using parsimony routine of the GCG package. RLEP identifiers, sequence lengths and the sequence present at the concatenated polymorphic sites are indicated on the right. The sequence gaa signifies the presence of G, A and A at positions 501, 583 and 592 in the multiple alignment.

displaying a 45 bp sequence, or subsequence thereof, at the 3′-end that is also wholly or partially present at the 5′-end of some copies (Figure 3). With the exception of a single site, where six copies of REPLEP have a GGG tract whereas the remaining nine have GG, the nucleotide sequences are identical. No information as to a possible function is available.

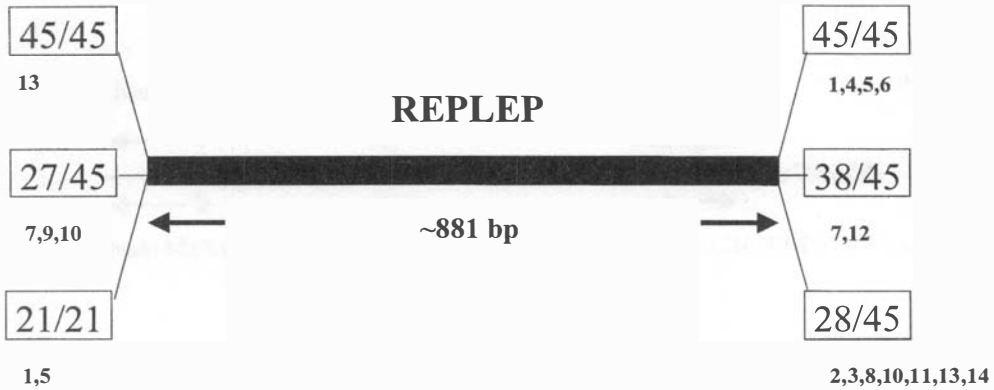

Figure 3. Schematic organization of REPLEP elements. The variable segments associated with the ends are shown as boxes with the numbers corresponding to the length in bp; 27/45 denote that 27 bp of the 45 bp segment are present. Identifiers for each REPLEP are shown below the boxes.

LEPREP

There are five intact sequences belonging to the LEPREP family and three fragments (Figure 4). These display near identical sequences with only three base differences being detected, two C–T transitions in REPLEP5 and one in REPLEP3. Unlike RLEP, REPLEP and LEPRPT, LEPREP displays a number of features commonly associated with IS elements and most probably corresponds to a degenerate version. The complete LEPREP sequence is 2383 bp long, contains a 54 bp palindromic inverted repeat and has a 6 bp inverted repeat (5′-CTAGTG) at its ends. Although there are no open reading frames that could code for

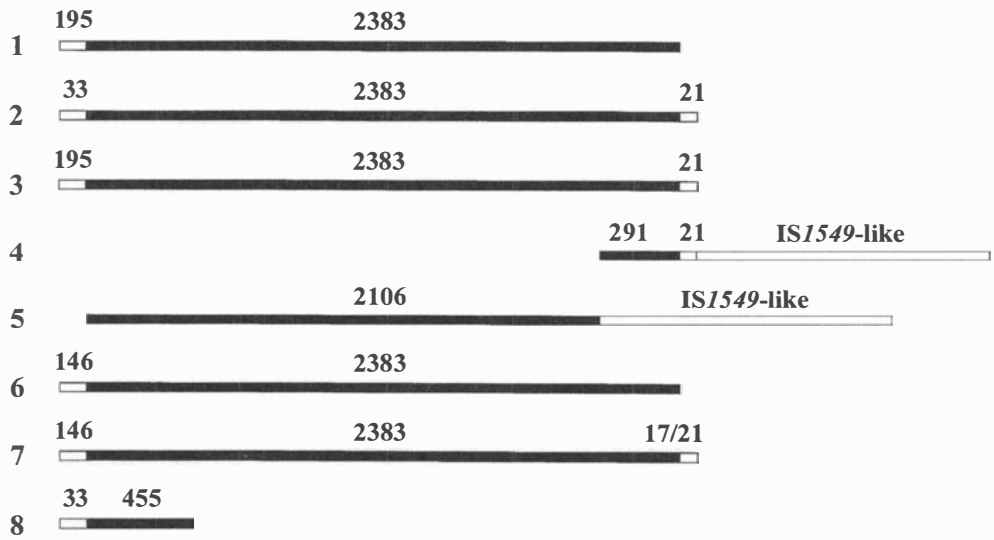

Figure 4. Schematic organization of LEPREP elements. The length in bp of each element is indicated. The variable segments associated with the ends are shown as stippled boxes with the numbers corresponding to the length in bp; 17/21 denotes that 17 bp of the 21 bp segment are present. Identifiers for each REPLEP are shown to the left.

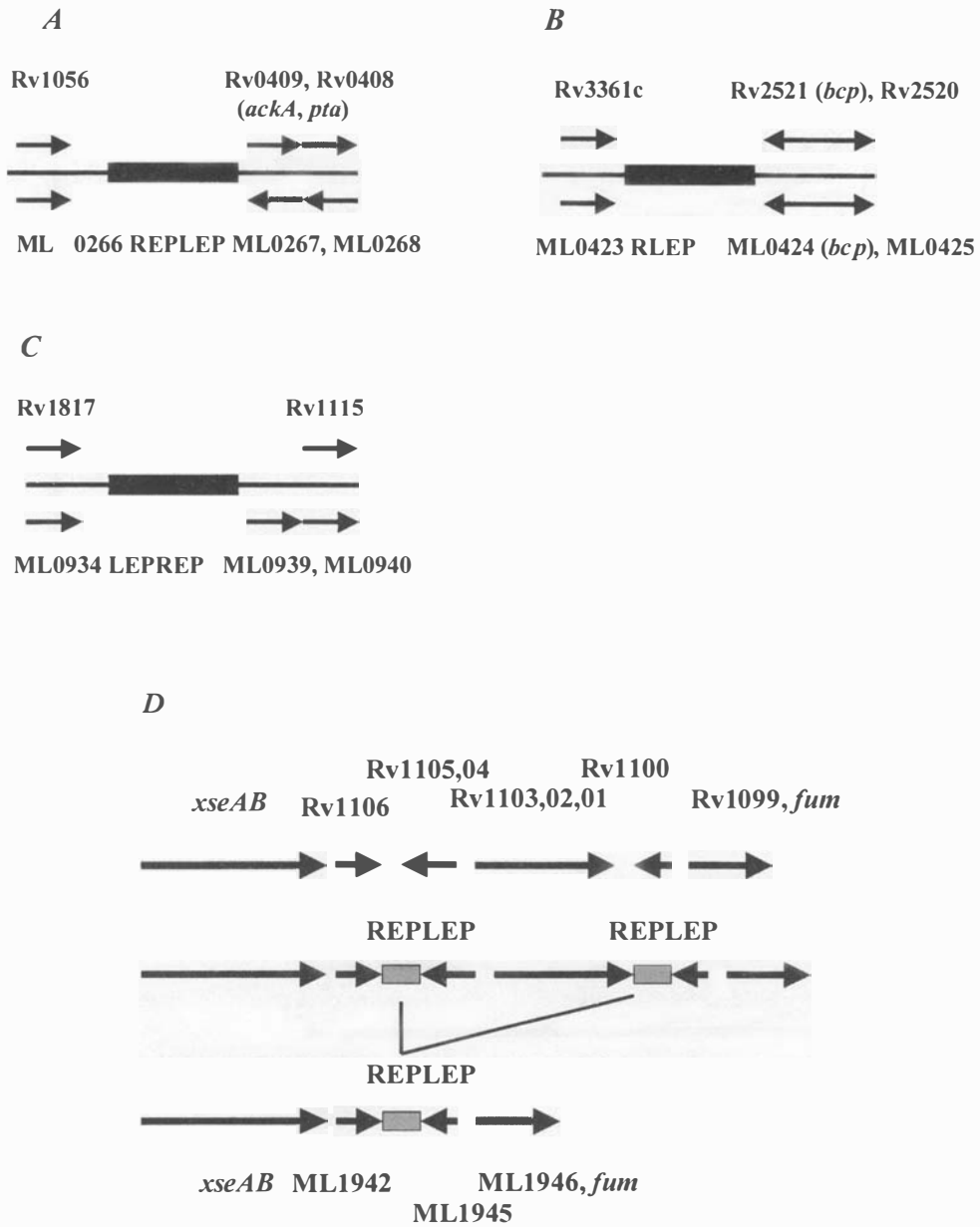

Figure 5. Repetitive elements and genome discontinuities. The three main repetitive elements in the *M. leprae* genome are shown together with examples of flanking genes and their counterparts in *M. tuberculosis*. The *M. tuberculosis* genes are designated with Rv prefixes, and the *M. leprae* genes with ML prefixes. Note the breaks in continuity of number of *M. tuberculosis* genes that indicate translocation event. **A.** REPLEP, 12 complete copies of ~875 bp, plus 2 fragments; **B.** RLEP, 36 complete copies of 545–700 bp plus 1 fragment; **C.** LEPREP, 6 complete copies of 2,400 bp plus 3 fragments. **D.** Example of gene loss by deletion following transposition of a REPLEP element and homologous recombination between the two copies. The gene organization in *M. tuberculosis* H37Rv is shown at the top followed by the corresponding region in an ancestor of *M. leprae*. The present situation in the TN strain of *M. leprae* is shown at the bottom.

functional proteins, BLASTX searches revealed extensive sequence similarity to parts of transposases from *Pseudomonas putida* (EMBL:AJ245436) and *Agrobacterium tumefaciens* (EMBL:Z18270), and to putative group II intron maturase-related proteins such as that of the fungus, *Cryphonectria parasitica* (EMBL:AF218567). Copies 4, 5 and 8 of LEPREP have been truncated and in two cases this appears to have resulted from the insertion of another IS element of 1261 bp that is now degenerate but still shows extensive similarity to IS*1549* from *Mycobacterium smegmatis*. Copies 4 and 5 are truncated at their 3′- and 5′-ends, respectively, share a 14 bp residual sequence and are followed by the IS*1549*-like element (Figure 4). They may once have comprised part of the same LEPREP element. Four copies of LEPREP are followed by the same 21 (or 17) bp sequence whereas two copies are preceded by common sequences of 195, 146 and 33 bp, respectively. It is conceivable that these sequences represent preferential sites of insertion for LEPREP.

LEPRPT

There are five sequences belonging to the LEPRPT family. Copies 1, 2 and 4 are 1252–1254 bp in length whereas copies 3 and 5 appear to have been shortened, as they only comprise 707 and 533 bp. The sequences of the LEPRPT elements are identical and contain no significant open reading frames. Copies 4 and 5 are preceded by the same 51 bp segment while copies 2 and 3 both have identical 7 bp sequences at their 5′-ends. Although the size of LEPRPT is consistent with that of an IS element, there is no other evidence to this effect.

REMODELLING THE GENOME

When whole genome comparisons of the tubercle and leprosy bacilli were performed it became apparent that there were ~65 conserved chromosomal segments common to both bacteria with loss of gene synteny occurring at sites occupied by repetitive elements in most cases.[13] This is illustrated in Figure 5, where one can see that gene order changes abruptly at sites harbouring RLEP, REPLEP and LEPREP. It is probable that this resulted from recombination events between dispersed repetitive sequences of the same family. If the elements were arranged in inverted orientation this would result in displacement and inversion of segments of the chromosome whereas recombination events between directly oriented repeats would result in deletion of the intervening segment. A potential example of this is shown in Figure 5D.

To conclude, it is likely that chromosomal rearrangements, gene deletions and duplications have had a profound effect on the biology of *M. leprae* and in turn on leprosy itself. One of the major forces that shaped this process was undoubtedly the dispersion of repetitive DNA, which may have been catalysed by enzymes encoded by the elements, followed by homologous recombination between these dispersed repeats effected by recombinases such as RecA. Characterization of the residual repetitive sequences has helped us to understand the past of the leprosy bacillus and may provide us with new tools to track its dissemination in the future.

Acknowledgements

We thank Edouard Yeramian for help with repeat analysis. STC wishes to acknowledge the financial support of the Institut Pasteur, the Association Française Raoul Follereau, the

Programme de Recherche Fondamentale en Microbiologie et Maladies Infectieuses, and the National Institutes of Health, National Institute of Allergy and Infectious Diseases (grant no. 1 RO1 AI47197-01A1) for parts of this work.

References

[1] Mahillon J, Chandler M. Insertion sequences. *Microbiol Mol Biol Rev*, 1998; **62:** 725–774.

[2] Cole ST, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998; **393:** 537–544.

[3] Gordon SV, Heym B, Parkhill J *et al.* New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, 1999; **145:** 881–892.

[4] Brosch R, Gordon SV, Eiglmeier K *et al.* Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. In: Hatfull GF, Jacobs WR Jr (eds) *Molecular genetics of mycobacteria.* ASM Press, Washington D.C., 2000, pp. 19–36.

[5] Brosch R, Philipp W, Stavropolous E *et al.* Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra. *Infect Immun*, 1999; **67:** 5768–5774.

[6] Fang Z, Doig C, Kenna DT *et al.* IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J Bacteriol*, 1999; **181:** 1014–1020.

[7] Tekaia F, Gordon SV, Garnier T *et al.* Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle Lung Dis*, 1999; **79:** 329–342.

[8] Lupski JR, Roth JR, Weinstock GM. Chromosomal duplications in bacteria, fruit flies, and humans. *Am J Hum Genet*, 1996; **58:** 21–27.

[9] Brosch R, Gordon SV, Buchrieser C *et al.* Comparative genomics uncovers tandem chromosomal duplications in some strains of *Mycobacterium bovis* BCG: implications for vaccination. *Comp Funct Genom (Yeast)*, 2000; **17:** 111–123.

[10] Supply P, Magdalena J, Himpens S, Locht C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Mol Microbiol*, 1997; **26:** 991–1003.

[11] Mazars E, Lesjean S, Banuls AL *et al.* High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proc Natl Acad Sci USA*, 2001; **13:** 1901–1906.

[12] Supply P, Mazars E, Lesjean S *et al.* Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol*, 2000; **36:** 762–771.

[13] Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.

[14] Altschul SF, Boguski MS, Gish W, Wooton JC. Issues in searching molecular sequence databases. *Nature Genet*, 1994; **6:** 119–129.

[15] Altschul S, Gish W, Miller W *et al.* A basic local alignment search tool. *J Mol Biol*, 1990; **215:** 403–410.

[16] Rutherford K, Parkhill J, Crook J *et al.* Artemis: sequence visualization and annotation. *Bioinformatics*, 2000; in press.

[17] Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*, 1999; **27:** 573–580.

[18] Jones LM, Cole ST, Moszer I. Leproma: a *Mycobacterium leprae* genome browser. *Lepr Rev*, 2001; **72:** 470–477.

[19] Woods SA, Cole ST. A rapid method for the detection of potentially viable *Mycobacterium leprae* in human biopsies: a novel application of PCR. *FEMS Microbiol Lett*, 1989; **65:** 305–310.

[20] Shin YC, Lee H, Walsh GP *et al.* Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *J Clin Microbiol*, 2000; **38:** 4535–4538.

[21] Matsuoka M, Maeda S, Kai M *et al. Mycobacterium leprae* typing by genomic diversity and global distribution of genotypes. *Int J Lepr Other Mycobact Dis*, 2000; **68:** 121–128.

[22] Berthet FX, Rasmussen PB, Rosenkrandt I *et al.* A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology*, 1998; **144:** 3195–3203.

[23] Gordon SV, Brosch R, Billault A *et al.* Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol*, 1999; **32:** 643–656.

[24] Gordon SV, Eiglmeier K, Garnier T *et al.* Genomics of *Mycobacterium bovis. Tubercle Lung Dis*, 2001; **6:** in press.

[25] Clark-Curtiss JE, Docherty MA. A species-specific repetitive sequence in *Mycobacterium leprae* DNA. *J Infect Dis*, 1989; **159:** 7–15.

[26] Grosskinsky CM, Jacobs Jr WR, Clark-Curtiss JE, Bloom BR. Genetic relationships among *Mycobacterium leprae*, *Mycobacterium tuberculosis*, and candidate leprosy vaccine strains determined by DNA hybridization: identification of an *M. leprae* specific repetitive sequence. *Infect Immun*, 1989; **57:** 1535–1541.

[27] Woods SA, Cole ST. A family of dispersed repeats in *Mycobacterium leprae*. *Mol Microbiol*, 1990; **4:** 1745–1751.

[28] Fsihi H, Cole ST. The *Mycobacterium leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel polA locus associated with genomic variability. *Mol Microbiol*, 1995; **16:** 909–919.

[29] Gordhan BG, Mizrahi V. The RLEP-flanked polA gene from *Mycobacterium leprae* is not transcribed in Mycobacterium smegmatis. *Gene*, 1997; **187:** 63–66.

[30] Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol*, 1990; **4:** 1653–1659.

Lepr Rev (2001) **72,** 462–469

The integrated genome map of *Mycobacterium leprae*

KARIN EIGLMEIER*, SYLVIE SIMON, THIERRY GARNIER & STEWART T. COLE
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28, rue du Dr Roux, 75724 Paris, Cedex 15, France

Summary The integrated map of the *Mycobacterium leprae* genome unveiled for the first time the genomic organization of this obligate intracellular parasite. Selected cosmid clones, isolated from a genomic library created in the cosmid vector Lorist6, were identified as representing nearly the complete genome and were subsequently used in the *M. leprae* genome sequencing project. Now a new version of the integrated map of *M. leprae* can be presented, combining the mapping results from the Lorist6 cosmids with data obtained from a second genomic library constructed in an *Escherichia coli*-mycobacterium shuttle cosmid, pYUB18. More than 98% of the *M. leprae* genome is now covered by overlapping large insert genomic clones representing a renewable source of well defined DNA segments and a powerful tool for functional genomics.

Introduction

When Gerhard Henrik Armauer Hansen established in 1873 the contagious nature of leprosy, and thereby associated for the first time bacteria with a disease in man, he probably did not envisage that over 100 years later leprosy, caused by *Mycobacterium leprae*, would still be an important health problem in several countries. Only in the last 2 decades of the 20th century, due to the leprosy elimination strategy recommended by the WHO[1,2] and the widespread implementation of multidrug therapy (MDT), there was a substantial reduction of leprosy prevalence achieved, resulting in an estimated 1–2 million registered cases. Effective MDT kills the bacteria; however, it does not reverse or cure the nerve damage in the afflicted patients. For this reason, some of the current major leprosy research priorities are the development of new diagnostic tools for the detection of early leprosy cases, new methods for detecting drug resistance or persistence, and the design of new drugs with increased effectiveness, capable of shortening the duration of antileprosy chemotherapy and which are operationally less demanding. To prevent the onset of infection by developing a leprosy vaccine could also contribute towards leprosy elimination efforts. The wide-spread administration of BCG, *Mycobacterium bovis* Bacille Calmette Guérin, is probably already

* Corresponding author: Karin Eiglmeier. Mailing address: Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28, rue du Dr Roux, 75724 Paris, Cedex 15, France. Phone: 33-1-45 68 84 49, Fax: 33-1-40 61 35 83, E-mail: kei@pasteur.fr

 0305-7518/01/072462+08 $1.00

contributing to the reduction of leprosy prevalence, offering a significant degree of protection from clinical leprosy.[3,4] The characterization of protective antigens from *M. leprae* to be included in a prophylactic vaccine and the understanding of its unusual life style, particularly its ability to enter and live inside Schwann cells, are research areas where many questions remain unanswered.

Research on *M. leprae*—the challenge

M. leprae, is an obligate intracellular parasite which has, so far, evaded cultivation *in vitro* outside of living animals. The development of the mouse model for leprosy research[5–8] was the first step towards the study of *M. leprae* in the laboratory, and is still in use for testing drug susceptibility and monitoring of treatment. The introduction of the armadillo (*Dasypus noveminctus* Linnaeus) as an animal model[9] made it possible for the first time to prepare a sufficient quantity of bacteria to envisage biochemical or genetic studies of *M. leprae*. Together with the rapidly developing recombinant DNA technologies, this represented a watershed for leprosy research. Preparation of chromosomal *M. leprae* DNA and first estimations of its G + C content and size became possible with the availability of armadillo-derived bacteria.[10] The values obtained (G + C 56%, genome size of about 3·38 Mb) were ultimately shown to be very close to the data generated by the *M. leprae* sequencing project (57·8%, 3,268 Mb) several years later.[11]

The λgt11 period in *M. leprae* research and the generation of other genomic libraries

The advent of recombinant DNA technologies opened new avenues for leprosy research by analysis of gene products expressed in a cultivable bacterium, *Escherichia coli*. In the bacteriophage λgt11 system, a relatively small (2–8 kb) foreign insert DNA is expressed under the control of the *E. coli* β-galactosidase promoter, therefore ensuring that the cloned sequence will be efficiently transcribed and translated in *E. coli*, generating a β-galactosidase fusion protein. Armadillo-derived genomic DNA was utilized in λgt11 expression libraries and many investigators successfully identified recombinant clones producing antigens recognized by a panel of monoclonal antibodies or patient's sera[12–19]; see references [20] and [21] for a compilation. These antigenic determinants were subsequently characterized; however, it became evident that despite using different murine and human sera to screen the libraries, a limited number of immunodominant protein antigens was identified,[22] and the same antigen was sometimes independently re-isolated in different laboratories,[20] suggesting an element of bias in the system used. In addition, the size limitation in the λgt11 cloning system, while ideal for the detection of small antigenic epitopes, was not suitable for the analysis of full length genes.

To date, all attempts to obtain high molecular weight chromosomal DNA isolated from armadillo-derived *M. leprae* have been unsuccessful. The strong physical barrier of the cell envelope is one explanation, necessitating a relatively rough treatment to obtain permeabilization and preventing *in situ* preparations of *M. leprae* DNA suitable for pulsed-field gel electrophoresis (PFGE). The second reason is that *M. leprae* isolated from an animal host is a heterogenous population of bacilli with many cells probably harbouring already degraded DNA. This limits the maximum insert length to 40–60 kb. Chromosomal DNA fragments

of this size can be prepared in sufficient quantities from armadillo-derived bacteria. The technique used to manipulate DNA of this size is the cosmid cloning strategy. However, because of the relatively small size of the starting material, the conventional partial digest methods employed for library construction are not always appropriate due to the risk of obtaining chimeric inserts. Nevertheless, investigators subjected *M. leprae* chromosomal DNA to partial restriction enzyme digestion and subsequently utilized size fractionated DNA fragments for the construction of different types of large insert libraries.[10]

The integrated map of *M. leprae*

With the aim of centralizing and unifying the *M. leprae* data and creating a new starting point for leprosy research, an integrated genome map was initiated, with the first step consisting of the construction of an ordered cosmid library. No *M. leprae* strain has ever been cloned, strictly speaking, so working with a clinical isolate seemed to be an acceptable solution. Thus *M. leprae*, originally derived from a patient from Tamil Nadu, was purified from an experimentally infected armadillo and the chromosomal DNA was prepared as described by Clark-Curtiss *et al.*[10] In order to circumvent the problem of chimeric clones, the DNA was gel-fractionated, end-repaired and blunt-end fragments in the 35–50 kb size range were cloned directly into the cosmid vector Lorist6.[23] About 1000 independent clones harbouring *M. leprae* DNA were obtained, corresponding to about 10 genome equivalents, and the cosmids were subjected to a modified fingerprint analysis.[24] Briefly, clones are 'fingerprinted' by treatment with restriction enzymes, measurement of the sizes of all the resulting fragments, followed by employing a computer-assisted matching technique[25] to detect regions common to several clones. The fingerprint data permit overlapping cosmids to be identified, enabling a contiguous physical map of the *M. leprae* chromosome to be built up of overlapping clones.[26]

This contig map was refined and completed by hybridization, either using complete cosmids or suitable end-fragments of cosmids as probes. The understanding of the basic organization of the *M. leprae* chromosome was further improved by the successful positioning of all the available cloned *M. leprae* genes and loci, along with well-conserved, heterologous genes from other bacteria (such as housekeeping genes or stable RNAs). The ultimate comparison of fingerprinting data and hybridization results with genetic markers finally resulted in the integrated map of the *M. leprae* chromosome.

Based on this integrated map, the chromosome was estimated to be ~2·8 Mb in size. Seven years later the *M. leprae* sequencing project revealed a genome of ≈3·2 Mb for the leprosy bacillus,[11] indicating that the data from the fingerprint analysis had led to the underestimation of the contig sizes, due to paucity of bands in the fingerprint or different restriction fragments having similar lengths and hence appearing to be the same fragment. Nevertheless, the overall clone and gene distribution on the integrated map was left nearly unchanged and the hypothesis of a circular chromosome was confirmed.

First genome comparisons of *M. leprae* with *M. tuberculosis*[27] based on the genetic maps were within reach, revealing existing regions of synteny. However, it rapidly became evident that the order of these limited genome sections was different, observations which were later substantiated by the corresponding genome sequencing projects of *M. leprae* and *M. tuberculosis*.

Shuttle vector libraries

A prerequisite for studying proteins produced by mycobacterial pathways or those whose expression is catalysed by the products of other mycobacterial genes is the introduction of their genes in an appropriate surrogate host such as *Mycobacterium smegmatis* or *M. bovis* BCG where faithful gene expression can be obtained. The replicative shuttle cosmid pYUB18,[28] which contains an origin of replication functional in *E. coli* and a second, mycobacterial origin of replication for stable maintenance in mycobacteria, proved to be very useful for this purpose. As a selective marker pYUB18 carries a kanamycin resistance gene, functional in both *E. coli* and mycobacteria; however, due to the size of the vector, slightly smaller chromosomal DNA inserts can be cloned (between 30 and 35 kb). In order to pave the way for new functional genomic experiments, we wanted to determine a minimum tiling path (i.e. the combination of cosmids which gave the minimum overlap) of *M. leprae* shuttle cosmid clones representing the entire genome. Having the complete *M. leprae* genome sequence at our disposal, the positioning of *M. leprae* shuttle clones by cosmid end-sequencing became feasible as a direct mapping approach. About 800 shuttle cosmid clones from a pYUB18-based library (obtained from W. Jacobs Jr), generated from a different source of chromosomal *M. leprae* DNA, were characterized by end-sequencing and positioned on the *M. leprae* genome ; however, only incomplete coverage with shuttle cosmids was obtained. A high number of chimeric clones were identified and eliminated, due to an apparent discrepancy with the genome sequence. Additional proof for the chimeric nature of these shuttle cosmids was generated by comparison of their fingerprint data with the complete cosmid fingerprint dataset.

These comparisons also reinforced the initial assumption, that *M. leprae* strains from different origins exhibit no obvious, important genome diversity,[10,29–31] a hypothesis that was also strengthened by sequence comparisons of the generated end-sequences with the genome sequence. However, the previously reported polymorphisms, the RLEP distribution in the genome[32] and the variations in the TTC repeats,[33] have not yet been exploited in the pYUB18 based library.

The integration of the Lorist6-based cosmids and the pYUB18 shuttle cosmid clones yielded a new integrated map, which, as in the first version, still exhibits the four gaps in the library coverage creating contig breaks. Comparisons with the annotated *M. leprae* genome reveal gap sizes of approximately 2·9, 7·5, 5·1 and 42·3 kb, corresponding to about 1·7% of the genome. Table 1 summarizes the open reading frames which are absent from the two cosmid libraries; in total 34 possible ORFs are concerned, comprising only 19 genes and 10 conserved hypotheticals. Why these regions are not present in the two different libraries is unclear. There could be, as in other similar genome projects,[34–37] *M. leprae* sequences that cannot be cloned in *E. coli* using these vectors. Another explanation could be the quality of the DNA preparation used as starter material for the library construction in that these regions might be underrepresented in the cloned, size fractionated region and more clones should be specifically screened for the presence of these missing sequences.

Conclusions and future studies

These ordered libraries are valuable tools and will facilitate genetic research considerably as they represent renewable sources of well-defined segments of the *M. leprae* genome, easy to distribute and convenient to handle in a non-pathogenic, cultivable host bacterium. They

Table 1. Regions of the *M. leprae* chromosome absent from the Lorist6 and also from the pYUB18 genomic libraries. The first column indicates the limits, corresponding to the sequenced *M. leprae* genome, of the regions absent in the two mapped libraries. In column two the putative genes present in these regions are given. Columns three and four indicate the number of conserved hypotheticals and hypothetical genes present in these uncloned regions and the fifth column contains the number of identified pseudogenes

Coordinates (nts)	Genes	Conserved hypotheticals	Hypotheticals	Pseudogenes
174 526–177 499	ML0131 - oxidoreductase *fadD29* - acyl-CoA synthetase	1	–	–
254 015–261 597	ML0192 - membrane protein	2	1	4
655 781–660 912	PPE - PPE-family protein *mihF* - integration host factor *gmk* - guanylate kinase *dfp* - flavoprotein *metK* - S-adenosylmethionine synthase	1	–	–
911 512–953 873	*sahH* - S-adenosyl-L-homocysteine hydrolase *tmk* - thymidylate kinase *mtrA* - two-component response regulator *mtrB* - two-component system sensor kinase *lpqB* - lipoprotein *secA* - preprotein translocase subunit ML0782 - transmembrane transport protein *aroA* - 3-phosphoshikimate 1-carboxyvinyl transferase ML0793 - bacteriophage protein ML0803 - two-component system sensor kinase *whiB1* - putative transcriptional regulator	6	4	17
Total	19	10	5	21

remove the tedious step of preparation of mycobacterial chromosomal DNA and, in addition, the insert size permits investigations of longer operons or genes with related functions, often linked on the chromosome.

Nowadays, complete bacterial genome sequences are relatively easy to produce without prior mapping of clones with large inserts or the establishment of a complete integrated map. However, sequencing projects concerning species necessitating laborious culture conditions, biohazard facilities or with complex genomes which make the sequencing approach by a whole genome shotgun difficult due to their special features (for example, extensive, numerous repetitive regions or DNA segments which are difficult to clone) benefit from clones carrying defined DNA fragments and a detailed cosmid map, serving as a framework for the assembly. Selected clones represented ideal starting material for the systematic genome sequencing, giving early access to thoroughly analysed and annotated regions of the chromosome, long before the genome was completely sequenced, as for example, the first completely sequenced cosmid, B1790.[38] Sequence analysis of this cosmid revealed the presence of the genes for four potential drug targets: *rpoB* (rifampicin), *tuf* (tetracyclin derivatives), *rpsL* (streptomycin) and *efg* (fusidic acid). These results permitted the early development of molecular diagnostic tests for determining rifampicin resistance in *M. leprae*[39] which, refined further, culminated in a relatively simple assay for resistance detection[40] (Honoré *et al.*, this issue). In addition, benefiting from the sequence data obtained

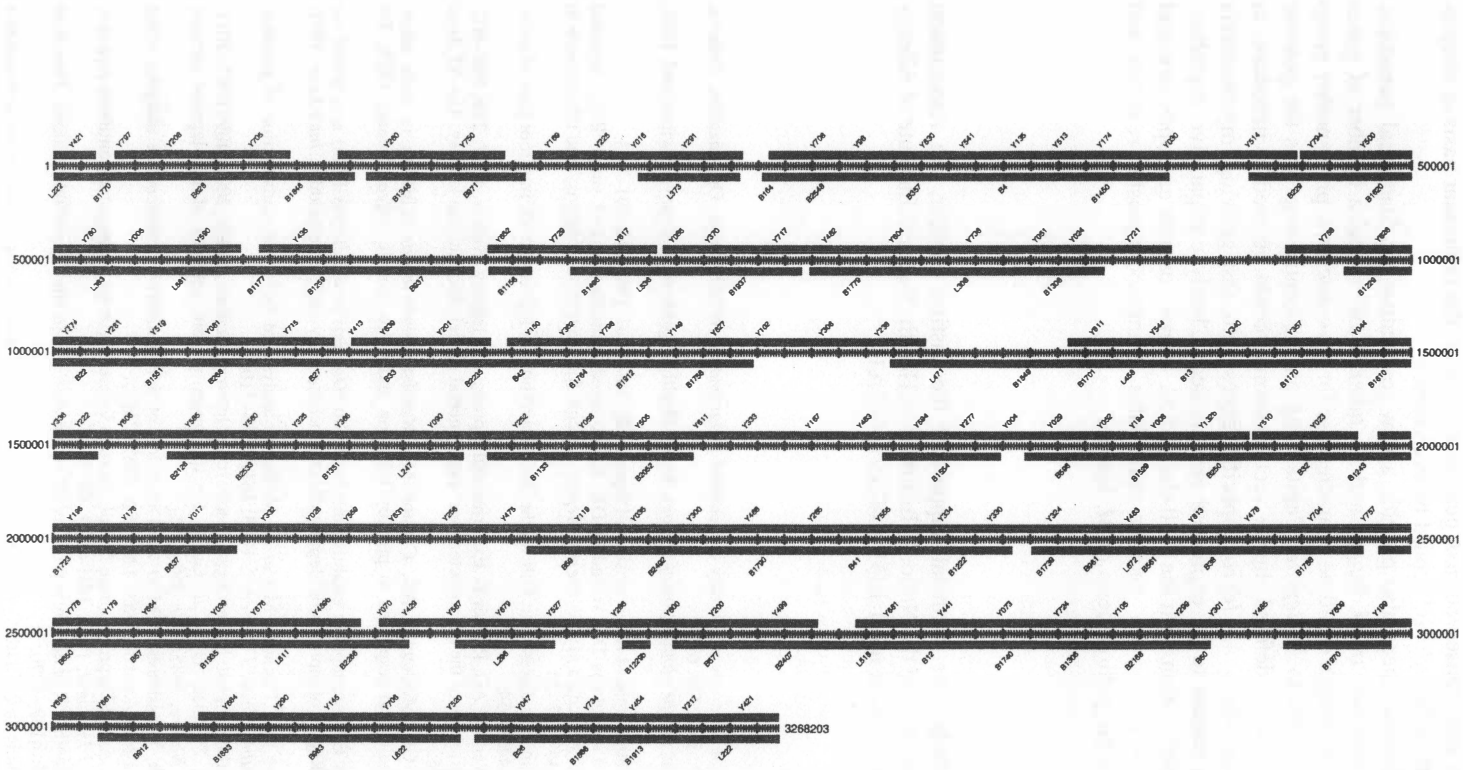

Figure 1. Latest version of the integrated map of *M. leprae*. A minimal set of Lorist6-based/pYUB18 shuttle vector clones is shown, covering in 4 contigs more than 98% of the *M. leprae* genome. The central black line corresponds to the genome sequence, with the positions indicated in nucleotides. Below the sequence line, the lower bars represent the sections of the chromosome present in completely sequenced clones of the Lorist6-based library. The upper bar indicates the genome sections represented in cosmids of the pYUB18-based library. The exact coordinates of the selected clones can be obtained at the following web site: http://genolist.pasteur.fr/Leproma.

from the evolutionary well-conserved *rpsL* gene of *M. leprae,* the molecular basis of streptomycin resistance in *M. tuberculosis*[41] could be determined.

The resources described here also provide a new foundation for functional genomics. Comparisons of the genomes of *M. leprae* and *M. tuberculosis* reveal a number of genes common to both species which could be characteristic for mycobacteria, plus another group of genes which are specific to *M. leprae*. Capitalizing on the contig map plus the genome sequence and its annotation, defined shuttle vector cosmids could be used to introduce, in a straightforward way, these *M. leprae* specific genes into fast-growing mycobacteria allowing *M. leprae* proteins to be expressed and studied. Clearly comparative mycobacterial genomics in combination with the well-defined *M. leprae* cosmids can open new and systematic ways of searching for vaccine candidates, immunodiagnostic reagents and molecules involved in the pathogenesis of *M. leprae*.

Acknowledgements

We wish to acknowledge the financial support of the Institut Pasteur, the Association Française Raoul Follereau, and the National Institutes of Health, National Institute of Allergy and Infectious Diseases (grant no. 1 RO1 AI47197—01A1).

References

[1] WHO Study group. *Chemotherapy of leprosy for control programmes.* World Health Organization, Geneva, Switzerland, 1982, Tech. Rep. Ser. 675.
[2] WHO. *Action programme for the elimination of leprosy.* World Health Organization, Geneva, Switzerland, 1998, Status Rep. 98.2.
[3] Fine PEM. BCG vaccination against tuberculosis and leprosy. *Br Med Bull*, 1988; **44:** 691–703.
[4] Fine PEM, Clayton D, Ponnighaus JM, Warndorf DK. Randomized controlled trial of single BCG, repeated BCG, or combined BCG and killed *Mycobacterium leprae* vaccine for prevention of leprosy and tuberculosis in Malawi. *Lancet*, 1996; **348:** 17–24.
[5] Shepard CC. The experimental disease that follows the injection of human leprosy bacilli into foot pads of mice. *J Exp Med*, 1960; **112:** 445–454.
[6] Rees RJ, Waters MF, Weddell AG, Palmer E. Experimental lepromatous leprosy. *Nature*, 1967; **215:** 599–602.
[7] Rees RJ, Weddell AG. Biology of the mycobacterioses: experimental models for studying leprosy. *Ann NY Acad Sci*, 1968; **154:** 214–236.
[8] Lancaster RD, Hilson GR, McDougall AC, Colston MJ. *Mycobacterium leprae* infection in nude mice: bacteriological and histological responses to primary infection and large inocule. *Infect Immun*, 1983; **39:** 865–872.
[9] Kirchheimer WF, Storrs EE. Attempts to establish the armadillo (*Dasypus novemcinctus* Linn) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int J Lepr*, 1971; **39:** 693–702.
[10] Clark-Curtiss JE, Jacobs WR, Docherty MA *et al.* Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. *J Bacteriol*, 1985; **161:** 1093–1102.
[11] Cole S, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.
[12] Young RA, Mehra V, Sweetser D *et al.* Genes for the major protein antigens of the leprosy parasite *Mycobacterium leprae*. *Nature*, 1985; **316:** 450–452.
[13] Young RA, Bloom BR, Grosskinsky CM *et al.* Dissection of *Mycobacterium tuberculosis* antigens using recombinant DNA. *Proc Natl Acad Sci USA*, 1985; **82:** 2583–2587.
[14] Cherayil BJ, Young RA. A 28-kDa protein from *Mycobacterium leprae* is a target of the human antibody response in lepromatous leprosy. *J Immunol*, 1988; **141:** 4370–4375.
[15] Young D, Lathigra R, Hendrix R *et al.* Stress proteins are immune targets in leprosy and tuberculosis. *Proc Natl Acad Sci USA*, 1988; **85:** 4267–4270.
[16] Sathish M, Esser RE, Thole JER, Clark-Curtiss JE. Identification and characterization of antigenic determinants

of *Mycobacterium leprae* that react with antibodies in sera from leprosy patients. *Infect Immun*, 1990; **58:** 1327–1336.

[17] Thole JE, Schoningh R, Janson AA *et al*. Molecular and immunological analysis of a fibronectin-binding protein antigen secreted by *Mycobacterium leprae*. *Mol Microbiol*, 1992; **6:** 153–163.

[18] Mehra V, Sweetser D, Young RA. Efficient mapping of protein antigenic determinants. *Proc Natl Acad Sci USA*, 1986; **83:** 7013–7017.

[19] Laal S, Sharma YD, Prasad HK *et al*. Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. *Proc Natl Acad Sci USA*, 1991; **88:** 1054–1058.

[20] Young DB, Kaufmann SHE, Hermans PWM, Thole JER. Mycobacterial protein antigens: a compilation. *Mol Microbiol*, 1992; **6:** 133–145.

[21] Thole JE, Wieles B, Clark-Curtiss JE *et al*. Immunological and functional characterization of *Mycobacterium leprae* protein antigens: an overview. *Mol Microbiol*, 1995; **18:** 791–800.

[22] Khanolkar-Young S, Kolk AHJ, Andersen AB *et al*. Results of the Third Immunology of Leprosy/Immunology of Tuberculosis Antimycobacterial Monoclonal Antibody Workshop. *Infect Immun*, 1992; **60:** 3925–3927.

[23] Gibson TJ, Rosenthal A, Waterson RW. Lorist 6, a cosmid vector with *BamHI, NotI, ScaI* and *HindIII* cloning sites and altered neomycin phosphotransferase gene expression. *Gene*, 1987; **53:** 283–286.

[24] Coulson A, Sulston J, Brenner S, Karn J. Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci USA*, 1986; **83:** 7821–7825.

[25] Sulston J, Mallett F, Staden R *et al*. Software for genome mapping by fingerprinting techniques. *CABIOS*, 1988; **4:** 125–132.

[26] Eiglmeier K, Honoré N, Woods SA *et al*. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol Microbiol*, 1993; **7:** 197–206.

[27] Philipp W, Poulet S, Eiglmeier K *et al*. An integrated map of the genome of the tubercle bacillus *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc Natl Acad Sci USA*, 1996; **93:** 3132–3137.

[28] Jacobs Jr WR, Kalpana GV, Cirillo JD *et al*. Genetic systems for mycobacteria. *Methods Enzymol*, 1991; **204:** 537–555.

[29] Clark-Curtiss JE, Walsh GP. Conservation of genomic sequences among isolates of *Mycobacterium leprae*. *J Bacteriol*, 1989; **171:** 4844–4851.

[30] Sela S, Clark-Curtiss JE, Bercovier H. Characterization and taxonomic implications of the rRNA genes of *Mycobacterium leprae*. *J Bacteriol*, 1989; **171:** 70–73.

[31] Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol*, 1990; **4:** 1653–1659.

[32] Fsihi H, Cole ST. The *Mycobacterium leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel *pol*A locus associated with genomic variability. *Mol Microbiol*, 1995; **16:** 909–919.

[33] Shin Y-C, Lee H, Lee H *et al*. Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *J Clin Microbiol*, 2000; **38:** 4535–4538.

[34] Kohara Y, Akiyama K, Isono K. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, 1987; **50:** 495–508.

[35] Knott V, Blake DJ, Brownlee G. Completion of the detailed restriction map of the *E. coli* genome by the isolation of overlapping cosmid clones. *Nucl Acids Res*, 1989; **17:** 5901–5912.

[36] Birkenbihl RP, Vielmetter W. Cosmid derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110. *Nucl Acids Res*, 1989; **17:** 5057–5069.

[37] Charlebois RL. Detailed physical map and set of overlapping clones covering the genome of the archebacterium *Halofax volcanii* DS2. *J Mol Biol*, 1991; **222:** 509–524.

[38] Honoré N, Bergh S, Chanteau S *et al*. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. *Mol Microbiol*, 1993; **7:** 207–214.

[39] Honoré N, Perrani E, Telenti A *et al*. A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. *Int J Lepr Other Mycobact Dis*, 1993; **61:** 600–604.

[40] Roche PW, Shrestha N, Thomas A *et al*. Rapid detection of resistance to rifampicin in *Mycobacterium leprae*. *Lepr Rev*, 2000; **71:** S96–97.

[41] Honoré N, Cole ST. Streptomycin resistance in mycobacteria. *Antimicrob Agents Chemother*, 1994; **38:** 238–242.

Lepr Rev (2001) **72,** 470–477

Leproma: A *Mycobacterium leprae* genome browser

L. JONES*, I. MOSZER** & S. T. COLE***
**Service Informatique Scientifique*
***Unite de Genetique des Genomes Bacteriens*
****Unite de Genetique Bacteriens, Institut Pasteur, Paris, France*

Introduction

Leproma is a powerful Web based tool for extracting information about annotations from a *Mycobacterium leprae* genome database. The URL for the Leproma web site is http://genolist.pasteur.fr/Leproma.

With *Leproma*, the user may search the *M. leprae* genome[1] database using several search criteria. One can search by gene name or synonym, by region in the genome, by gene function or classification, by DNA or protein patterns, by a BLAST[2,3] or FASTA[4] search in the DNA sequence or the protein sequences, or by free text.

Search results can be in the form of a list where the columns in the list are set by the user or in the form of a drawing if the search results in a region in the genome. The user can also download or view the DNA sequence or the protein sequence from a single gene or from the list of a search result.

Leproma opening Web page

The *Leproma* web server will show a page (Figure 1) with free frames. The left frame shows the form for entering search criteria. The top right frame is used for showing the search results, either in the form of a list or a drawing (if the search request results in a region). It is also used for entering additional search criteria for a BLAST/FASTA search, a pattern search, or an extended annotation search. The bottom right frame is used for showing detailed information for a gene. Other pages may be shown for the help pages, a view of a DNA or protein sequence, the pre-calculated BLAST results for a gene, or links to other external databases.

The opening Web page (Figure 1; see also accompanying colour poster) shows in the upper right frame, the *Leproma* data version and date of the data release and an image of the *M. leprae* genome showing CDS or ORF positions, gene classification, and GC%. This image is clickable and results in a list (or drawing) of all the genes ±10 kb in the region

Correspondence to: S. T. Cole (e-mail: stcole@pasteur.fr)

0305-7518/01/072470+07 $1.00

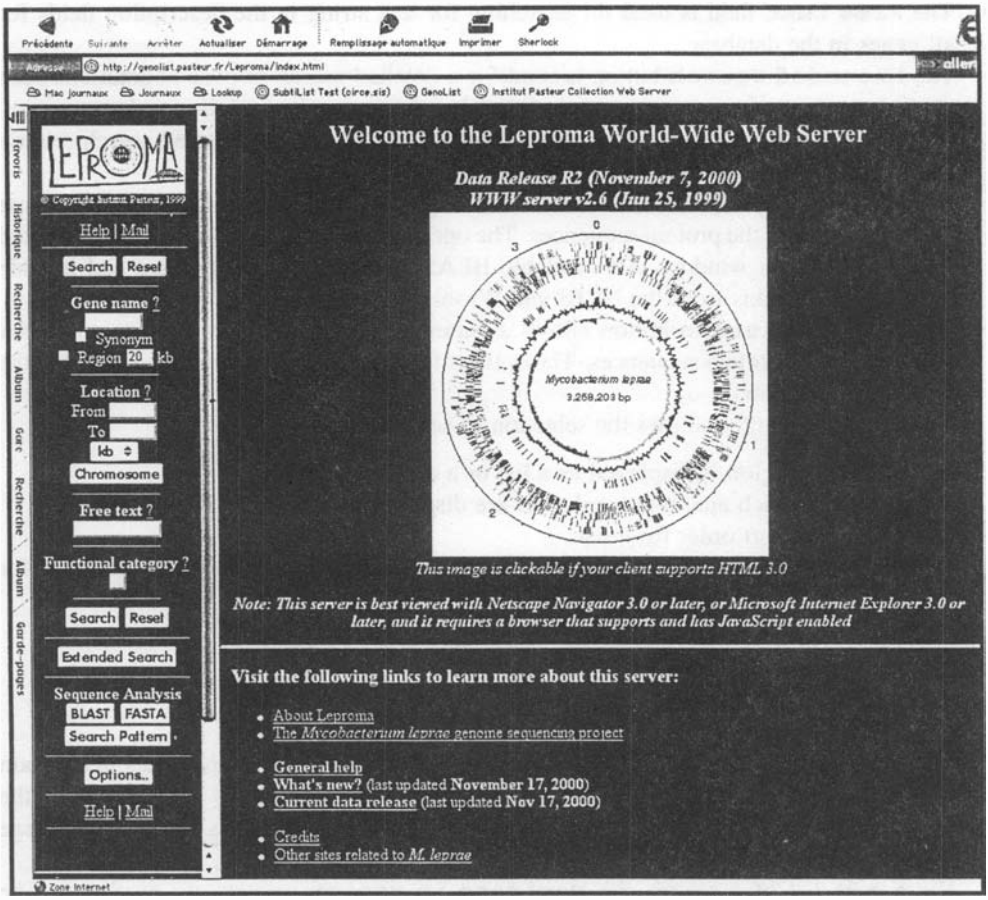

Figure 1. *Leproma* opening Web page (for colour version of this figure see also accompanying colour poster.

around the click point. The bottom right frame gives links to information about the *M. leprae* genome project.

The left side frame contains the form for entering basic search criteria.

- The **`Gene name`** field is used for specifying a gene name. One or more wild-card characters may be used for searching several genes. For example, 'dnaA' will search for the gene '`dnaA`' while '`dna*`' will search for all genes whose name starts with 'dna'. The **Synonym** option allows the use of synonyms or old gene names. The **Region** option gives a list of genes ±20 kb around the specified gene name. This option is ignored if a wild-card character is used in the **`Gene name`** field.
- The **Location** fields **From** and **To** will give a list (or drawing) of all genes within the given genome coordinates. The coordinates are in kilo-bases. The **`Chromosome`** button will display the image of the *M. leprae* genome from the opening page (Figure 1).
- The **`Functional category`** field is used to search for all genes of the given classification code.

- The **Free text** field is used for searching for text string in the description fields for all genes in the database.
- The **Extended Search** button is used for a detailed search of gene annotations such as Accession number, protein length, DNA length, molecular weight, isoelectric point, and a free text search limited to certain annotation fields. The extended search criteria form is displayed in the upper right frame.
- The **BLAST** and **FASTA** buttons will allow a BLAST or FASTA search of the genome DNA sequence or the protein sequences. The options for BLAST or FASTA are presented in the upper right window. All available BLAST programs are presented and include the BLAST versions from the NCBI and Washington University.
- The **Search Pattern** button allows a pattern or motif search of the genome DNA sequence or the protein sequences. The options for the pattern search are presented in the upper right window.
- The **Options** button allows the selection of several default options:
  - Whether a region is displayed as a list or a drawing.
  - For a list, which annotation columns are displayed.
  - The default sort order for a list.
  - How many items to be displayed at once in a list. A **More**... button is displayed for scrolling through the list.

Leproma list display of a region

The list display (Figure 2; see also accompanying colour poster) shows the genes in a certain region or the genes based on the user specified search criteria. A gene name in the **Gene Name** column may be clicked to display the detailed annotations for the given gene in the lower right frame.

For a gene list of a region, the **Navigate in region** options are displayed. The user can shift the region to the left (lower genome coordinates) by clicking the yellow left arrow, expand the region, contract the region, or shift the region to the right. The region coordinates may be changed directly with the **From** and **To** fields and clicking the **Update List** button. Also, the list display can be changed to a drawing (Figure 3; see also accompanying colour poster) by clicking the **Draw Region** button.

The list columns may be changed by selecting the appropriate gene list columns and clicking the **Update List** button.

The user may export information from the list including the list of genes and selected columns, the protein sequence for each CDS in the list, the DNA sequence for all entries in the list, and for a region list, the DNA sequence for the region in direct or reverse complement form. The data may be displayed on the user's screen or it may be downloaded to a local file on the user's system.

Leproma drawing of a region

The region drawing (Figure 3; see also accompanying colour poster) displays a region in the Leproma genome. Each genomic element is represented by a graphic symbol on the drawing.

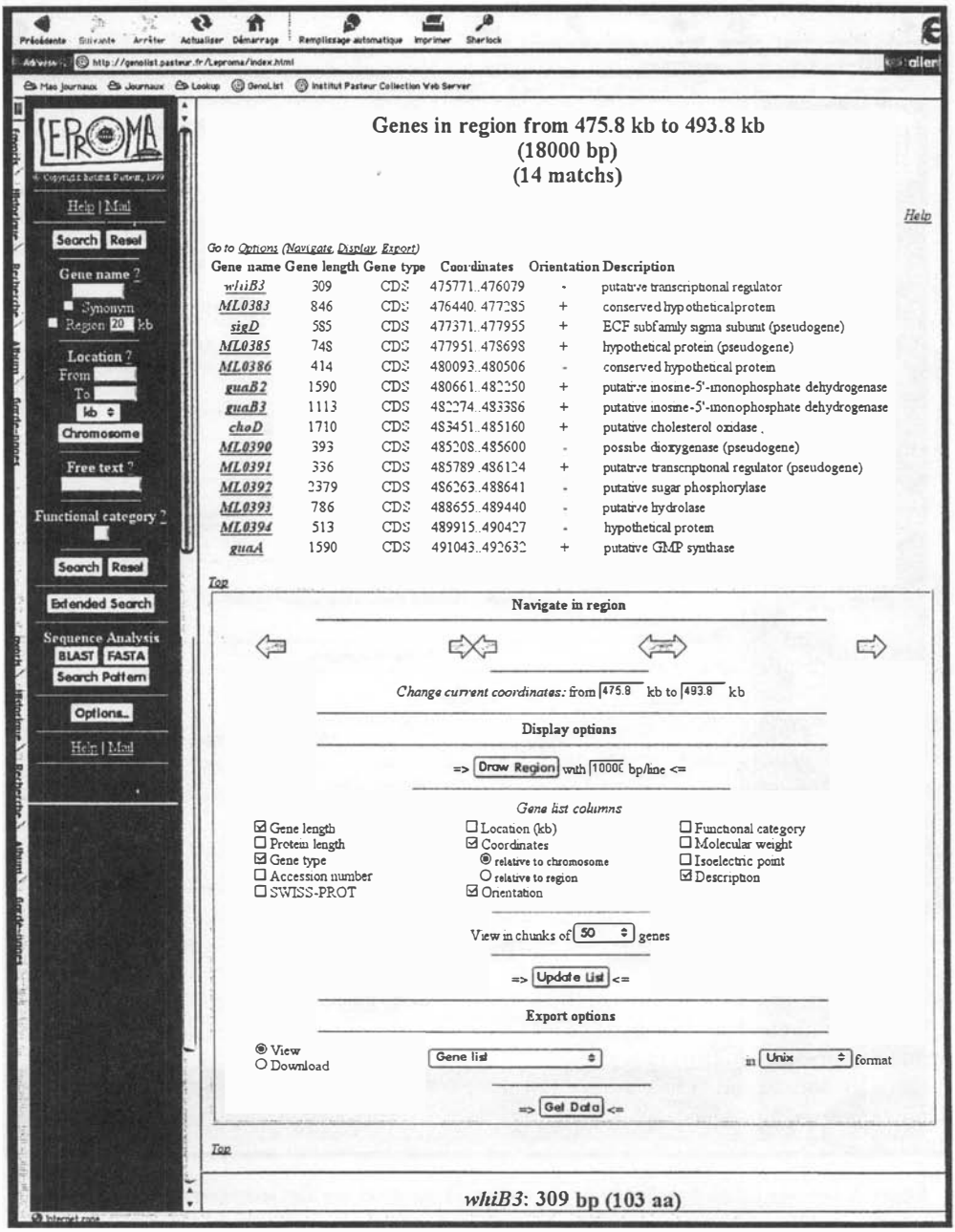

Figure 2. *Leproma* region list (for colour version of this figure see also accompanying colour poster).

A gene or CDS is represented by a coloured, horizontal line with the start-end points representing the beginning and end of the CDS and the direction of transcription. The colour represents the functional classification of the CDS that has been used for both the *M. leprae* and the *M. tuberculosis* genome projects.[5] There is one exception, as pink is used for the

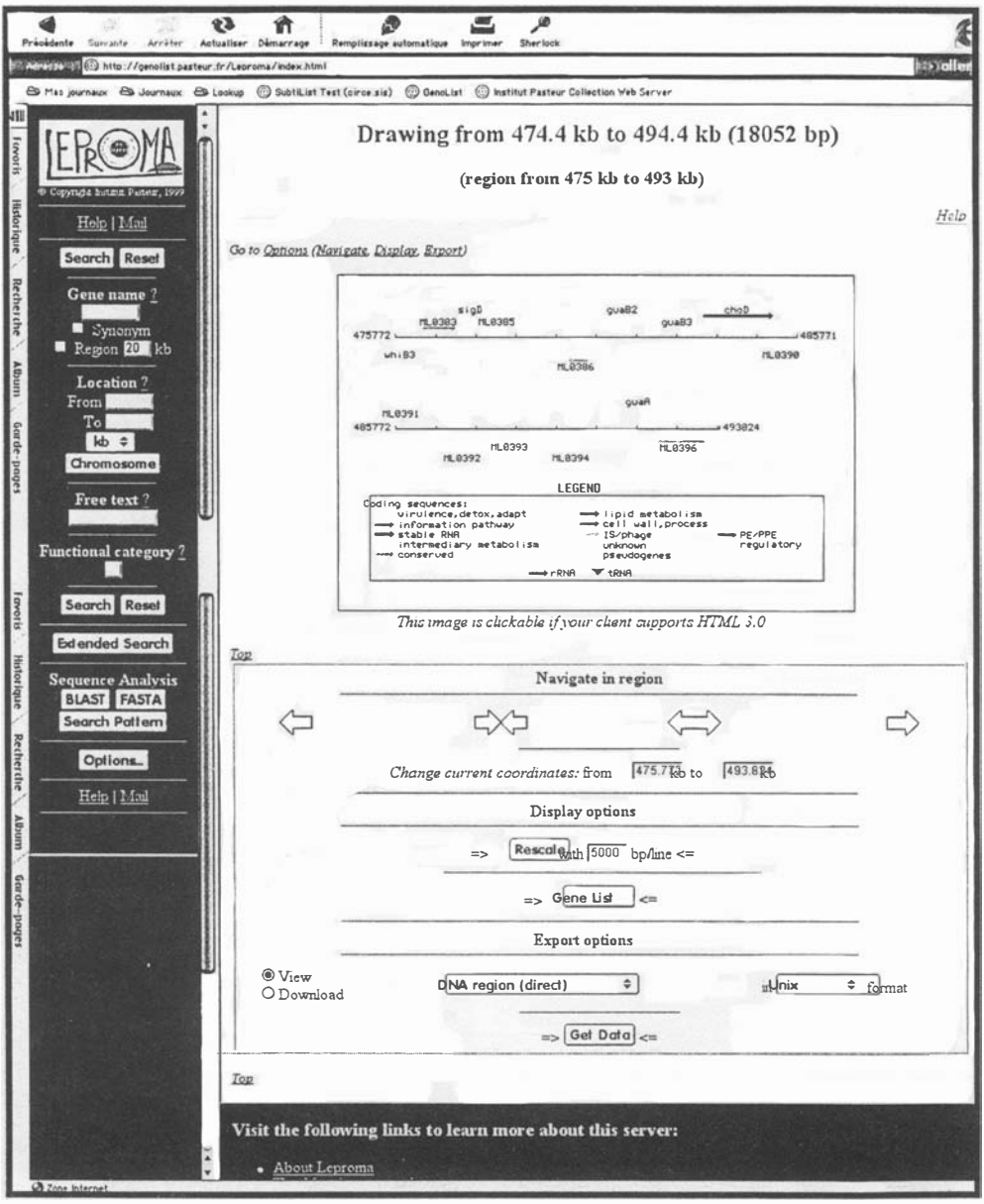

Figure 3. *Leproma* region drawing (for colour version of this figure see also accompanying colour poster).

many pseudogenes in *M. leprae*. The caption at the bottom of the drawing gives the colour code for each category of classification. RNA is represented by a black vertical arrow. Each graphic symbol and its name is clickable and displays detailed annotation in the lower right frame of the window.

As with the region gene list (Figure 2), the user may shift the region to the left, expand the region, contract the region, or shift the region to the right. The user may change the base-pair

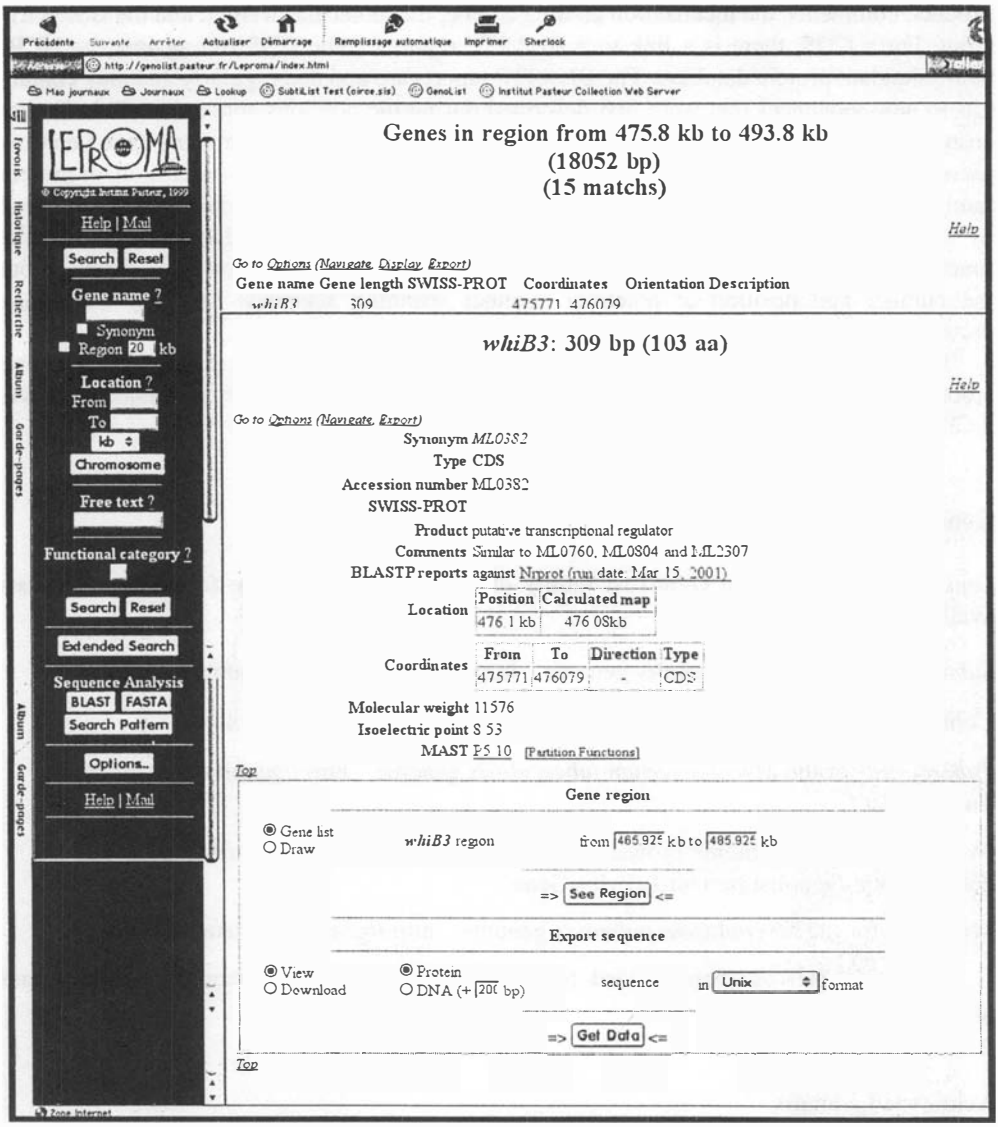

Figure 4. *Leproma* detailed gene annotation (for colour version of this figure see also accompanying colour poster).

density per line and change to a list display of the region. The user may export the genes or the DNA sequence in the region as for the gene list of a region.

Leproma detailed gene annotation

The gene detail display (Figure 4; see also accompanying colour poster) shows all annotations available for the selected CDS or RNA. These annotations include the size in base-pairs and in amino acids, for a CDS, the synonyms (if any), the accession number, the

product, comments, the localization on the genome, the molecular weight, and the isoelectric point. For a CDS, there is a link to a BLASTP report for the CDS run against the NCBI non-redundant protein database. The BLASTP reports are updated regularly and may contain hits to new sequences that were first described during the previous annotation. This facility ensures that new clues to function are made available and we recommend that Leproma users check them regularly. Information about membership of a CDS to a partition, or protein family, can be found by clicking the MAST link as this provides graphic output generated by the MEME[6] and MAST[7] programs. A table of partition functions summarizes the functional information about all the families and the TMHMM[8] link provides details about the number and position of potential member spanning segments in likely membrane proteins.

The user has the option to display in a list (Figure 2) or a drawing (Figure 3) the region around the gene or export the gene sequence as a DNA sequence or a protein sequence if it is a CDS.

GenoList genome browsers

Leproma is a part of the **GenoList** family of genome browsers. The following table lists available **GenoList** browsers:

SubtiList for the *Bacillus subtilis* genome http://genolist.pasteur.fr.subtilList

Colibri for the *Escherichia coli* genome http://genolist.pasteur.fr/Colibri

TubercuList for the *Mycobacterium tuberculosis* genome http://genolist.pasteur.fr/TubercuList

Pylorigene, a multi-genome browser for the two *Helicobacter pylori strains* J99 and 26695 http://genolist.pasteur.fr/PyloriGene

MypuList, for the *Mycoplasma pulmonis* genome http://genolist.pasteur.fr/MypuList

Other genomes are being added regularly. An updated status can be seen at http://genolist.pasteur.fr

Acknowledgements

We are grateful to Fredj Tekaia for his help with data analysis. This work was supported by the Association Française Raoul Follereau, the Wellcome Trust, and the Institut Pasteur.

References

[1] Cole ST, Eiglemeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409:** 1007–1011.
[2] Altschul S, Gish W, Miller W *et al.* A basic local alignment search tool. *J Mol Biol*, 1990; **215:** 403–410.
[3] Gish W, States D. Identification of protein coding regions by database similarity search. *Nature Genet*, 1993; **3:** 266–272.
[4] Pearson W, Lipman D. Improved tools for biological sequence comparisons. *Proc Natl Acad Sci USA*, 1988; **85:** 2444–2448.

[5] Cole ST, Brosch R, Parkhill J *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature (Lond)*, 1998; **393:** 537–544.
[6] Bailey T, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: *Second International Conference on Intelligent Systems for Molecular Biology*, 1994, 28–36.
[7] Bailey T, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, 1998; **14:** 45–48.
[8] Sonnhammer ELL, von Heijne C, Kmogh A. A hidden markov model for predicting transmembrane helices in protein sequences. *Sixth International Conference on Intelligent Systems.*

Lepr Rev (2001) **72,** 478

Erratum

In Editor's Choice in the September issues of Leprosy Review, the Editorial 'Needing to know? Ethical dilemmas in leprosy treatment and control' was erroneously attributed to John Porter and Anthony Zwei, rather than John Porter and Anthony Kessel. We apologise for this error, and for any confusion that it may have caused.

Lepr Rev (2001) **72,** 479–484

Index: volume 72

Instructions to Authors

Article submission. Articles submitted for publication in Leprosy Review should be sent to the Editor at the following address: Diana Lockwood, LEPRA, Fairfax House, Causton Road, Colchester CO1 1PU, England. The name(s) of the author(s) and the place where the work was done should be indicated clearly below the title of the article. Degrees and diplomas are not to be included.

It is understood that the article is offered to Leprosy Review alone, that it will be subject to editorial revision, and that its copyright becomes the property of LEPRA.

Format. Articles must be submitted *double spaced*, on one side of A4 (297 x 210 mm) paper, with wide margins (4 cm all round) in triplicate. On a separate sheet give the title, short title, name and postal address of the author, together with the name of the institution where the work was done. Abbreviations of titles of journals should follow the list of journals indexed in Index Medicus. References to books should include the editor(s), publisher and place of publication.

Electronic submission. Articles produced using a word-processor should also be supplied in electronic format (preferably Word or Wordperfect 5.1) if possible. Please submit a disk with the final revised version of the article. The electronic file should correspond exactly to the hard copy.

References. References should follow journal style [please contact the Assistant editor (e-mail sboobis@aol.com) for guidance if required], and journal title abbreviations should follow Index Medicus. Incorrectly styled references may lead to delay in publication.

Illustrations and Tables. Contributors must send the original artwork and two copies on separate sheets. In addition electronic versions may be submitted in the form of compressed tiffs, eps, jpegs or bitmaps. The recommended resolution is 600 dpi or above.

Units and Abbreviations. The journal recognizes the adoption of the Systeme International d'Unites (SI Units) proposed in Units, Symbols and Abbreviations (1972) published by the Royal Society of Medicine, 1 Wimpole Street, London W1M 8AE. Abbreviations should only be used for unwieldy names, and only when they occur frequently.

Proofs are submitted to authors for immediate return by air.

Copyright/Offprints. Authors submitting a manuscript do so on the understanding that if it is accepted for publication, copyright in the article for the United States of America shall be assigned to LEPRA. LEPRA will not put any limitation on the personal freedom of the author to use material contained in the article in other works which may be published in North America.

* * *

Typeset and printed in Europe by the Alden Group, Oxford.

CONTENTS

Editorial

Original papers