

# Comparative genomics of metabolic pathways in *Mycobacterium* species: gene duplication, gene decay and lateral gene transfer

Pradeep Reddy Marri,<sup>1</sup> John P. Bannantine<sup>2</sup> & Geoffrey B. Golding<sup>1</sup>

<sup>1</sup>Department of Biology, McMaster University, Hamilton, ON, Canada; and <sup>2</sup>National Animal Disease Center-USDA-Agricultural Research Service, Ames, IA, USA

**Correspondence:** Geoffrey B. Golding, Department of Biology, McMaster University, Hamilton, ON, Canada L8S 4K1. Tel.: +1 905 525 9140; fax: +1 905 522 6066; e-mail: golding@mcmaster.ca

Received 16 October 2005; revised 15 May 2006; accepted 11 June 2006.  
First published online 1 August 2006.

DOI:10.1111/j.1574-6976.2006.00041.x

Editor: Michael Galperin

## Keywords

*Mycobacterium*; comparative genomics; metabolic pathways; pathogenicity.

## Abstract

The genus *Mycobacterium* comprises significant pathogenic species that infect both humans and animals. One species within this genus, *Mycobacterium tuberculosis*, is the primary killer of humans resulting from bacterial infections. Five mycobacterial genomes belonging to four different species (*M. tuberculosis*, *Mycobacterium bovis*, *Mycobacterium leprae* and *Mycobacterium avium* ssp. *paratuberculosis*) have been sequenced to date and another 14 mycobacterial genomes are at various stages of completion. A comparative analysis of the gene products of key metabolic pathways revealed that the major differences among these species are in the gene products constituting the cell wall and the gene families encoding the acidic glycine-rich (PE/PPE/PGRS) proteins. *Mycobacterium leprae* has evolved by retaining a minimal gene set for most of the gene families, whereas *M. avium* ssp. *paratuberculosis* has acquired some of the virulence factors by lateral gene transfer.

## Introduction

The study of pathogenic bacteria is undergoing a paradigm shift. The enormous amount of data coming from sequencing projects and the availability of bioinformatics tools for faster analysis of the generated data are revolutionizing the science of bacterial pathogenesis. The availability of complete sequences of different species belonging to a single genus enables comparative studies to understand the differences and commonalities among a group of species. Studies involving comparisons of complete microbial genomes can reveal significant differences in gene content and genome organization between closely related bacteria, provide insights into physiology and pathogenesis, and can identify polymorphic sequences with potential relevance to pathogenesis, immunity and evolution (Schoolnik, 2002; Alsmark *et al.*, 2004; Bai *et al.*, 2004; Eppinger *et al.*, 2004; Ferretti *et al.*, 2004; Moreira *et al.*, 2004; Nascimento *et al.*, 2004; Prentice, 2004).

*Mycobacterium* is one of the most studied pathogenic genera owing to the severity of its impact on human populations. To date, five genomes of mycobacterial species have been sequenced [*Mycobacterium tuberculosis* H37Rv (Cole *et al.*, 1998; Camus *et al.*, 2002), *M. tuberculosis* CDC1551 (Fleischmann *et al.*, 2002), *Mycobacterium bovis* (Garnier *et al.*, 2003), *Mycobacterium leprae*

(Cole *et al.*, 2001) and *Mycobacterium avium* ssp. *paratuberculosis* (Li *et al.*, 2005)] and another 14 are at various stages of completion (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>). *Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) in humans, infects one-third of the world population, with a newly infected individual added every second (Snider *et al.*, 1994). Each year about 2 million people die from TB infection (<http://www.who.int/tb/en>). Moreover, TB is the leading cause of death among people who are HIV-positive, accounting for about 13% of AIDS deaths worldwide. Leprosy, caused by *M. leprae*, is one of the oldest recorded diseases and still remains a major public health problem (Cole *et al.*, 2001). *Mycobacterium bovis* and *M. avium* ssp. *paratuberculosis* are cattle pathogens causing TB and Johne's disease, respectively. Owing to a variety of diseases caused by these organisms and variations in their genome size, a comparative study of these species would generate a wealth of information providing insights into the regions responsible for pathogenesis and host-specificity.

Earlier comparative studies in mycobacteria have mostly involved different strains of the same species of *M. tuberculosis* to understand the variability of pathogenesis among different strains (Fleischmann *et al.*, 2002). Other similar studies either involved the comparison of *M. tuberculosis* with *M. bovis* or of *M. tuberculosis* with *M. leprae* (Brosch

*et al.*, 2000; Cole *et al.*, 2001; Garnier *et al.*, 2003). A comprehensive examination covering all the currently sequenced genomes has not been done. A comparative genome analysis of the five available genomes from the genus *Mycobacterium* will enable a better understanding of the genome structure of these bacteria and the horizontal gene transfer pattern, and help to identify the species-specific genes.

In the current study, we discuss the genes present in various metabolic pathways and compare these genes across the five sequenced genomes of *Mycobacterium*. The discussion is divided into the following sections: (1) genome features; (2) energy metabolism; (3) amino acid biosynthesis; (4) biosynthesis of cofactors, prosthetic groups; (5) degradation of carbon compounds, amino acids and amines; (6) central intermediary metabolism; (7) lipid metabolism; (8) PE and PPE gene families; (9) macromolecule metabolism and degradation; (10) regulatory genes; and (11) insights into pathogenesis.

The analysis is based on the annotations of these genomes and any errors in the annotation will be reflected in the results and discussion presented here.

## Genome features

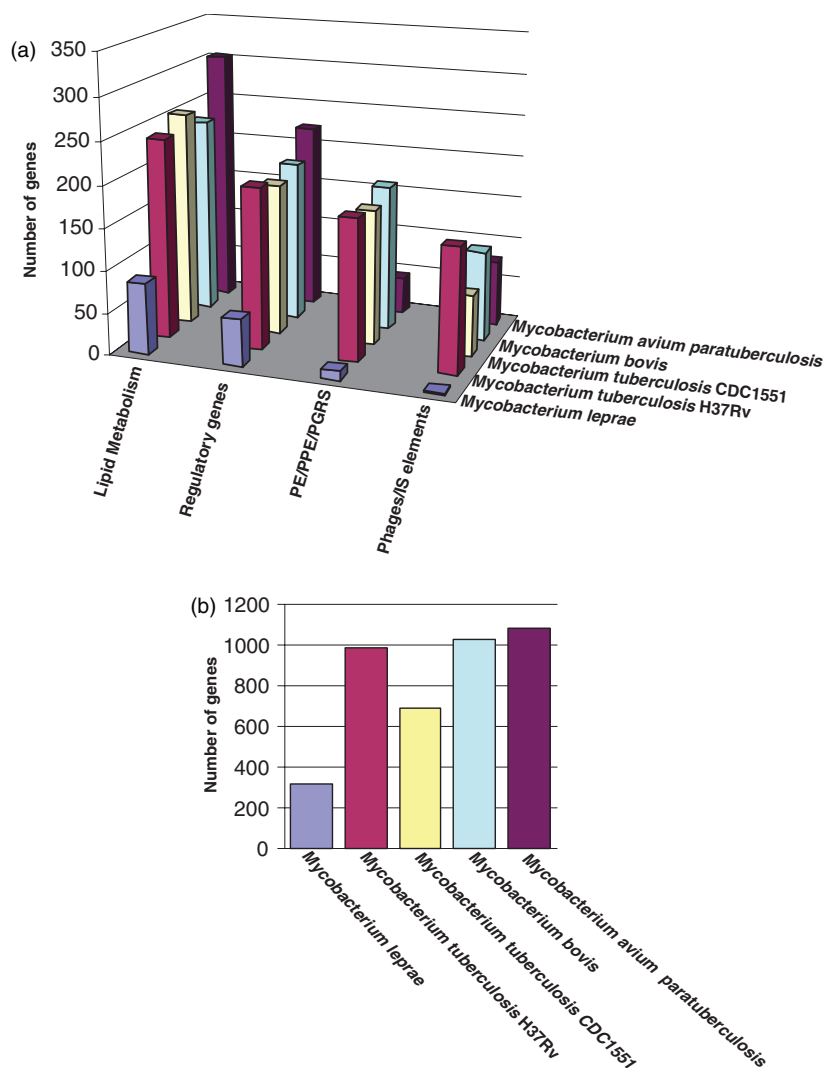
Mycobacteria typically have GC-rich sequences: the GC content of *M. tuberculosis* and *M. bovis* is around 65%; it is higher (69.3%) in *M. avium* ssp. *paratuberculosis* and lower (57.7%) in *M. leprae*. *Mycobacterium avium* ssp. *paratuberculosis* has a slightly higher percentage (91.5%) of the genome encoding proteins compared with *M. tuberculosis* (90.9%) and *M. bovis* (90.5%), whereas only half (49.6%) of the genome encodes functional proteins in *M. leprae*. The genomic features of these species are summarized in Table 1. There are also variations in the gene complement. *Mycobacterium tuberculosis* and *M. bovis* have ~3900 genes encoding proteins, *M. avium* ssp. *paratuberculosis* has 4350 genes and *M. leprae* has only ~1650 genes encoding functional proteins. The variations in the gene complement are reflected in the genes of lipid metabolism, PE/PPE gene family, insertion sequence (IS) elements and hypothetical proteins (Fig. 1). *Mycobacterium avium* ssp. *paratuberculosis* has an increased redundancy in the genes involved in lipid metabolism

probably resulting in a more robust cell wall compared with the other species owing to its colonization of the ruminant intestine, whereas *M. leprae* has evolved by having a minimal gene set for most of the pathways (Cole *et al.*, 2001; Vissa & Brennan, 2001; Li *et al.*, 2005). There are variations in the number of insertion sequences among these species with *M. tuberculosis* having 148 genes belonging to insertion sequences and phages compared with 107 of *M. bovis*, 87 of *M. avium* ssp. *paratuberculosis* and only two functional and 26 truncated copies of transposases of *M. leprae* (Fig. 1). The higher number of insertion element and phage-related genes might indicate greater intraspecies variability in *M. tuberculosis* compared with the other species.

Comparison of the proteins across the five genomes has revealed a common backbone of 1326 proteins (Appendix S1). The genetic closeness of these species is evident from the presence of a higher number of shared genes (estimated at 219) compared with the mycobacterial core (Charlebois & Dolittle, 2004; Marmiesse *et al.*, 2004). *Mycobacterium tuberculosis* and *M. bovis* share about 3700 genes between them, whereas they share a relatively lower number of genes (about 2600) with *M. avium* ssp. *paratuberculosis* (Table 2a). The genetic closeness of *M. tuberculosis* and *M. bovis* is also reflected in gene order, which is highly conserved in these two species; by contrast, there are many genomic rearrangements in *M. avium* ssp. *paratuberculosis* and *M. leprae* compared with *M. tuberculosis* (Fig. 2). Comparison of the protein sequences of the five organisms based on BLASTP analysis (Altschul *et al.*, 1997) identified 26, 47, 414, 155 and 966 unique proteins in *M. tuberculosis* H37Rv, *M. bovis*, *M. tuberculosis* CDC 1551, *M. leprae* and *M. avium* ssp. *paratuberculosis* respectively (Table S1). The presence of a higher number of unique proteins, especially in *M. avium* ssp. *paratuberculosis*, might possibly indicate that in spite of a pattern of reductional evolution (Cole *et al.*, 1998, 2001), these genomes might have also gained some additional genes during the course of their evolution (Koonin *et al.*, 2001; Kinsella *et al.*, 2003; Krzywinska *et al.*, 2004; Nakamura *et al.*, 2004). It should be noted that these genes are only 'unique' with respect to the other sequenced members of the genera and may be present in other bacteria or in other members of the genus that have not yet been sequenced. The number of unique genes includes some

**Table 1.** Genome features of mycobacterial species

	<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium bovis</i>	<i>Mycobacterium leprae</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
Genome size (bp)	4 411 529	4 403 836	4 345 492	3 268 203	4 829 781
Protein coding genes	3927	4186	3920	1604	4350
GC (mol%)	65.6	65.6	65.6	57.7	69.3
Average protein length	339	317	334	336	338
Protein coding (%)	90.9	90.7	90.5	49.6	91.5



**Fig. 1.** (a) Variations in lipid metabolism, phages/IS elements, regulatory and PE/PPE/PGRS gene families in *Mycobacterium* species. (b) Variations in the number of hypothetical proteins in *Mycobacterium* species.

**Table 2a.** Shared and unique genes among mycobacterial species\*

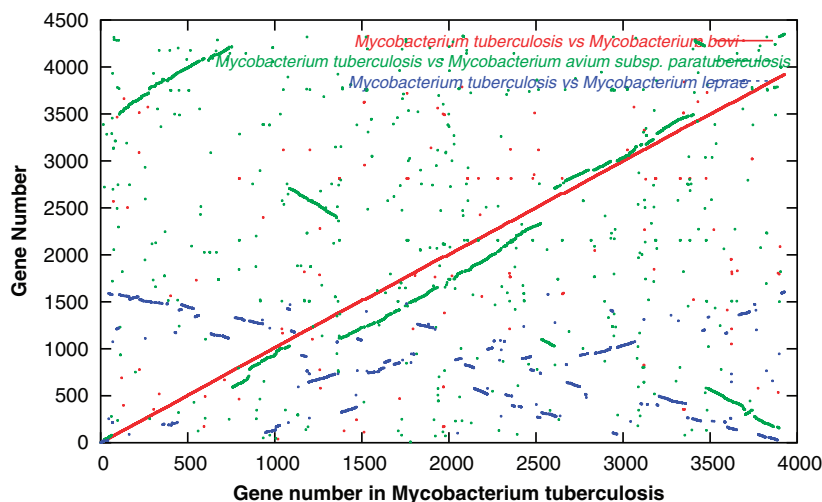
	<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium bovis</i>	<i>Mycobacterium leprae</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
<i>Mycobacterium tuberculosis</i> H37Rv	<b>26</b>	3644	3732	1396	2624
<i>Mycobacterium tuberculosis</i> CDC 1551	3644	<b>417</b>	3633	1384	2609
<i>Mycobacterium bovis</i>	3732	3633	<b>47</b>	1394	2622
<i>Mycobacterium leprae</i>	1396	1384	1394	<b>155</b>	1366
<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>	2624	2609	2622	1366	<b>966</b>

Numbers in bold indicate the genes uniquely present in the corresponding genome.

\*e-20, 35% length, based on genome annotations.

duplicate genes and portions of disrupted genes. When a similar analysis was performed using TBLASTN, the number of unique proteins was reduced to six, eight, 122, 149 and 872 in *M. tuberculosis* H37Rv, *M. bovis*, *M. tuberculosis* CDC 1551, *M. leprae* and *M. avium* ssp. *paratuberculosis*, respectively (Table 2b; Table S1). In *M. avium* ssp. *paratuberculosis*

and *M. leprae*, the number of genes is similar to those based on genome annotation, although the number of unique proteins is reduced in *M. tuberculosis* H37Rv, *M. bovis* and *M. tuberculosis* CDC 1551. This difference is largely a reflection of the use of different genome annotation programs for *M. tuberculosis* H37Rv (Krogh *et al.*, 1994) and *M.*



**Fig. 2.** Gene order comparison of *Mycobacterium tuberculosis* with *Mycobacterium bovis*, *Mycobacterium avium* ssp. *paratuberculosis* and *Mycobacterium leprae*. Gene homologs were determined by reciprocal BLAST hits and each match is represented by a dot.

**Table 2b.** Shared and unique genes among mycobacterial species\*

	<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium bovis</i>	<i>Mycobacterium leprae</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
<i>Mycobacterium tuberculosis</i> H37Rv	<b>6</b>	3827	3802	1396	2878
<i>Mycobacterium tuberculosis</i> CDC 1551	3827	<b>122</b>	3810	1396	2840
<i>Mycobacterium bovis</i>	3802	3810	<b>8</b>	1397	2851
<i>Mycobacterium leprae</i>	1396	1396	1397	<b>149</b>	1393
<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>	2878	2840	2851	1393	<b>872</b>

Numbers in bold indicate the genes uniquely present in the corresponding genome.

\*e-20, 35% length, using whole genome sequences.

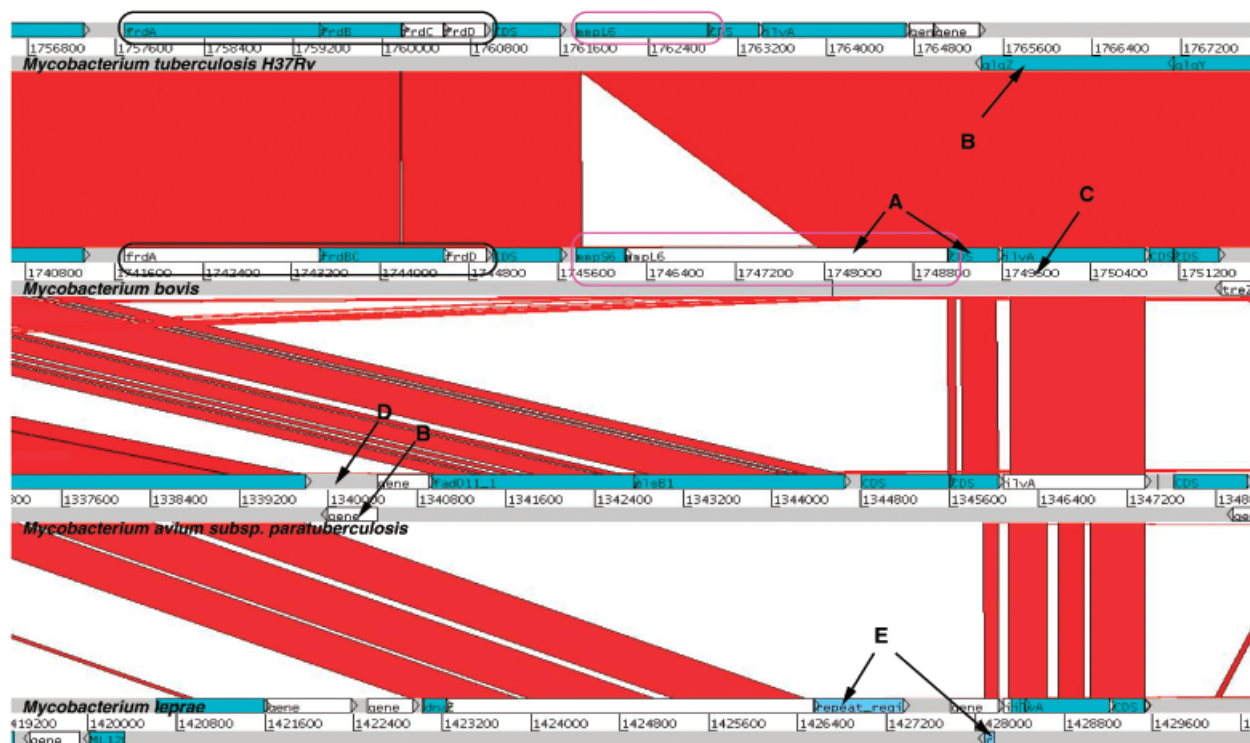
*tuberculosis* CDC 1551 (Salzberg *et al.*, 1998). Note that some homologous genes might be present but fail the criteria that the expected value cutoff is less than  $10^{-20}$  (the genes are evolutionarily distant) or have lengths that differ by more than 65% (the genes are truncated or fused). The unique genes are listed in Table S1. The unique genes mostly encode hypothetical proteins and do not enter further into the following discussions.

## Energy metabolism

All the genomes have a functional glycolytic pathway and tricarboxylic acid (TCA) cycle. *Mycobacterium tuberculosis* and *M. bovis* have an additional carbohydrate kinase gene (*pfkB*) that is absent in *M. leprae* and *M. avium* ssp. *paratuberculosis*. The substrate specificity of *pfkB* in *Mycobacterium* is unknown; however, the presence of this gene as part of an operon with the genes for universal stress protein A, a histidine kinase and an uncharacterized phosphoribosyltransferase suggests that *pfkB* is a stress-induced gene. The absence of *pfkB* might indicate a differential response to stress in *M. leprae* and *M. avium* ssp. *paratuberculosis*. *Mycobacterium avium* ssp. *paratuberculosis* has a metal-

independent form (class I) of fructose-bisphosphate aldolase (*fba*) whereas the other species have a Zn-dependent form (class II) of *fba* (Marsh & Lebherz, 1992). The presence of a class II *fba* in other actinobacterial genomes and the similarity of the *M. avium* ssp. *paratuberculosis* *fba* with a proteobacterial *fba* gene (69% identity with *fba* gene from *Roseovarius nubinhibens*) might indicate that *M. avium* ssp. *paratuberculosis* has acquired a copy of this gene by non-orthologous gene displacement. As the functions of both the classes of genes are interchangeable (Koonin & Galperin, 2003) this may not lead to any physiological differences between the species.

*Mycobacterium tuberculosis* CDC 1551, *M. bovis* and *M. avium* ssp. *paratuberculosis* have two isocitrate lyase homologues (*icl*, *aceA*), whereas *M. tuberculosis* H37Rv and *M. leprae* have only one functional copy of the isocitrate lyase gene. The *icl* gene is totally absent in *M. leprae* whereas *aceA* is split and nonfunctional in *M. tuberculosis* H37Rv (Cole *et al.*, 1998, 2001; Honer Zu Bentrup *et al.*, 1999). Isocitrate lyase is an essential anapleurotic enzyme of the glyoxylate cycle responsible for the growth of mycobacteria on acetate and palmitate and survival in the microaerophilic conditions inside the host (Rosenkrands *et al.*, 2000; Li *et al.*,



**Fig. 3.** Comparison of *frdABCD* genes in mycobacterial species (top to bottom): *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium avium* ssp. *paratuberculosis*, *Mycobacterium leprae* as seen in an ACT (Carver et al., 2005) genome browser. The ACT browser gives a comparative view of the genomes based on homology. Homologous regions between genomes are shown by red vertical lines. The white and blue rectangular boxes labeled a represent the genes on the positive strand, whereas the white and blue rectangular boxes labeled b represent the genes on the negative strand. The numbers labeled c represent the scale bar for genome length and the intergenic region is shown as gray regions labeled d. The light blue boxes labeled E represent repeat regions. The *frdABCD* operon (shown by a black oval) is present in *M. tuberculosis* and *M. bovis*, whereas it is absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. There is also variation in the membrane proteins in the four genomes. While *M. bovis* has two genes, *mmpS6* and *mmpL6* (shown by pink oval), *M. tuberculosis* has only *mmpL6* and it is shorter than its corresponding gene in *M. bovis*. *Mycobacterium avium* ssp. *paratuberculosis* and *M. leprae* do not have *frdABCD*, *mmpS6* and *mmpL6* genes.

2002). The presence of a single copy of this gene in *M. tuberculosis* and *M. leprae* might result in reduced virulence and survival inside macrophages. Moreover, the presence of a single copy of this gene in these species would make it an attractive antimycobacterial drug target as knocking-out of both the *icl* homologues leads to the rapid elimination of mycobacteria from lungs (Munoz-Elias & McKinney, 2005).

All the genes involved in aerobic respiration are conserved in *M. tuberculosis*, *M. bovis* and *M. avium* ssp. *paratuberculosis* whereas most are either lost or reduced to pseudogenes in *M. leprae* (Table S2). The pyruvate carboxylase in *M. tuberculosis* and *M. bovis* is replaced by phosphoenol pyruvate (PEP) carboxylase in *M. leprae* and *M. avium* ssp. *paratuberculosis*. In the absence of the NADH oxidase operon in *M. leprae*, PEP carboxylase might help in oxidation of NADH by converting PEP to fumarate or malate, whereas it might provide *M. avium* ssp. *paratuberculosis* with an additional option to produce ATP. In the case of anaerobic respiration, the essential genes *nirA* and *cysH* encoding nitrate and phosphate reductases, respectively, are

duplicated in *M. avium* ssp. *paratuberculosis* whereas the *narX* gene encoding a nitrate reductase is absent. A closer look at the duplicated genes (*nirA*, *cysH*) indicated that these genes are flanked by insertion elements probably indicating a recent duplication of these genes mediated by insertion elements. Whereas the genes *frdB* and *frdC* encoding proteins responsible for interconversion of fumarate and succinate are fused to form a single gene (*frdBC*) in *M. bovis*, *frdBCD* are absent in *M. avium* ssp. *paratuberculosis* (Fig. 3). The fumarate reductase complex, *frdABCD*, functions as an anaerobic phosphorylative electron transport chain in bacteria and plays a major role in metabolism of *M. tuberculosis* under starvation (Betts et al., 2002). The absence of *frdBCD* genes along with a nitrate reductase gene (*narX*), which is up-regulated during anaerobiosis in *M. tuberculosis* (Hutter & Dick, 1999), might possibly indicate a different mechanism of survival for *M. avium* ssp. *paratuberculosis* under anaerobic conditions. The fusion of *frdB* and *frdC* genes appears to have no effect on *M. bovis* as it is still able to use fumarate reductase as an electron acceptor. The *fdhF* gene

encoding formate dehydrogenase and the *fdxB* gene encoding a protein involved in electron transport in *M. avium* ssp. *paratuberculosis* have low sequence similarity with the genes of other mycobacterial species. *fdhF* is longer than its corresponding gene in other species and *fdxB* is shorter. There is an additional flavin adenine dinucleotide (FAD) binding domain in the *fdhF* gene of *M. avium* ssp. *paratuberculosis*, probably suggesting the dual function for this protein, whereas the *fdxB* gene in *M. avium* ssp. *paratuberculosis* has lost the NADH and FAD binding domains that are present in other mycobacteria but has an intact iron–sulfur domain. This might possibly indicate an alternate mechanism of action for these proteins resulting in an alternate survival strategy for *M. avium* ssp. *paratuberculosis* under anaerobic conditions as both genes are essential for the decomposition of formic acid under anaerobic conditions.

### Amino acid biosynthesis

Most of the genes involved in amino acid biosynthesis are highly conserved across all the species, emphasizing their role as essential genes. The genes in the glutamate, histidine and aromatic amino acid family biosynthesis pathway are conserved in all the genomes with the exception of *M. leprae*, which has lost some of these genes. The three essential genes of the aspartate family, *asnB*, *dapA* and *lysA*, are duplicated in *M. avium* ssp. *paratuberculosis* (Table S3). The gene *dapA* encodes dihydropicolinate synthase, which converts L-aspartate semialdehyde (ASA) to dihydropicolinate (DHDP) in one of the precursor steps in the formation of diaminopimelate (DAP) (Cirillo *et al.*, 1994; Pavelka & Jacobs, 1996), which in turn is a precursor for lysine biosynthesis, whereas *lysA* encodes DAP decarboxylase involved in the final step of lysine biosynthesis converting DAP to lysine (Gokulan *et al.*, 2003). The presence of duplicate copies of *lysA* and *dapA* might indicate an increased need for lysine in *M. avium* ssp. *paratuberculosis* owing to its role in cell-wall biosynthesis (Strominger, 1962). Increased lysine coupled with a higher number of lipid metabolism genes in *M. avium* ssp. *paratuberculosis* might result in a more robust cell wall that would lead to an enhanced protection against the acidic conditions prevailing in the ruminant gut. Additionally, it is also possible that *M. avium* ssp. *paratuberculosis*, in the absence of mycobactin, might use lysine as a precursor for the synthesis of siderophores *in vivo* as is seen in some *Streptomyces* species (Schupp *et al.*, 1988). The duplicate copy of *asnB*, a natural antibiotic resistance gene, might lead to increased levels of antibiotic resistance in *M. avium* ssp. *paratuberculosis* (Ren & Liu, 2006). The gene *cysA2* encoding a thiosulfate sulfurtransferase is duplicated in *M. tuberculosis* H37Rv whereas both *M. tuberculosis* and *M. bovis* have three copies (*cysM*, *cysM2*, *cysM3*) of the gene encoding cystathione  $\beta$ -synthase. These genes might help *M. tubercu-*

*losis* and *M. bovis* in the low-oxygen environments within the macrophages as thiosulfate sulfurtransferases have been found to have a role in the assembly of the iron–sulfur clusters that act as biosensors for oxygen and iron concentrations (Unden *et al.*, 1995; Florczyk *et al.*, 2001). *Mycobacterium avium* ssp. *paratuberculosis* and *M. leprae* have a single copy of the gene (*cysM2*) encoding cystathione  $\beta$ -synthase and also the *glyA* gene encoding serine hydroxymethyltransferase, which is present as two copies (*glyA*, *glyA2*) in *M. tuberculosis* and *M. bovis*. *Mycobacterium tuberculosis* might have acquired a duplicate copy of the gene given the major role of this gene in cell physiology (Chaturvedi & Bhakuni, 2003). The gene *ilvG*, encoding acetolactate synthase involved in isoleucine and valine biosynthesis, is disrupted and nonfunctional in *M. avium* ssp. *paratuberculosis* owing to a frameshift. However, it has a functional copy of the gene *ilvX*, which performs a similar function.

### Biosynthesis of cofactors, prosthetic groups and carriers

The genes involved in folic acid, pantothenate, pyridoxine and thiamine biosynthesis are highly conserved in all the species. *Mycobacterium tuberculosis* and *M. bovis* have some genes that are duplicated whereas *M. avium* ssp. *paratuberculosis* and *M. leprae* have evolved by having minimal genes required in all the pathways, indicating a metabolic streamlining. The genes *bioF2*, involved in the biosynthesis of biotin, *ribA*, involved in riboavin biosynthesis, *ggtB*, involved in glutathione degradation, and *idsB* and *grcC2*, involved in terpenoid biosynthesis, are duplicated in *M. tuberculosis* and *M. bovis*. The *bisC* gene encoding biotin sulfoxide reductase (Pierson & Campbell, 1990) and *trxA* encoding a thioredoxin are absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. As a result of the loss of the *bisC* gene, *M. avium* ssp. *paratuberculosis* and *M. leprae* may not be able to utilize biotin sulfoxide as a source for biotin biosynthesis whereas the absence of the thioredoxin gene might lead to variations in pathogenesis (Wieles *et al.*, 1995). The gene *cobL* encoding a methyl transferase involved in cobalamine biosynthesis is disrupted in *M. bovis* whereas it is intact in *M. tuberculosis* and *M. avium* ssp. *paratuberculosis*. There are significant differences in the genes involved in the biosynthesis of molybdopterin. Two gene clusters involved in the biosynthesis of molybdopterin are conserved in the two strains of *M. tuberculosis*. One of the gene clusters consisting of genes *moaA*, *moaB*, *moaC* and *moaD* is absent in *M. avium* ssp. *paratuberculosis*, only two genes (*moaB* and *moaE*) are functional in *M. leprae*, whereas *M. bovis* has a nonfunctional copy of *moaC2* and has lost the gene *moaE*, but has two additional genes in *moaA3* and *moaB3* (Table S4). The variations in the genes encoding molybdopterin, a

cofactor required for nitrate reductase activity, might lead to differences in the nitrate reductase activities in these species (Bertero *et al.*, 2003). Additionally, the loss of most of these genes along with the nitrate reductase gene cluster *narGHJI* in *M. leprae* indicates its inability to use nitrate as the final electron acceptor under anaerobic conditions.

### Degradation of carbon compounds, amino acids and amines

There are certain variations in the preference for carbon compounds among these species. The fundamental difference between *M. tuberculosis* and *M. bovis* is the inability of *M. bovis* to make pyruvate when glycerol is used as the sole carbon source (Garnier *et al.*, 2003; Hewinson *et al.*, 2006). The genome sequence indicates that all the genes of *M. bovis* encoding proteins required for the formation of pyruvate are nonfunctional. The pyruvate kinase gene (*pykA*) useful for the conversion of PEP to pyruvate is a pseudogene (Keating *et al.*, 2005). The gene *ald*, encoding L-alanine dehydrogenase, is also a pseudogene in the case of *M. bovis*, blocking the conversion of alanine to pyruvate. Additionally, *M. bovis* cannot utilize glycerol as a carbon source to form pyruvate owing to the disruption of the glycerol kinase (*glpK*) and also the *ugpA* involved in the import of glycerol-aldehyde-3-phosphate (Garnier *et al.*, 2003). The disruption in *M. tuberculosis* of *galT* encoding an enzyme involved in the conversion of alpha-D-galactose-1-phosphate to UDP-galactose indicates its inability to use galactose as a precursor for lactose biosynthesis. The gene *galE2* encoding an epimerase that converts UDP-glucose to UDP-galactose is nonfunctional as a result of being disrupted in *M. avium* ssp. *paratuberculosis*, whereas the gene *gabD1* encoding a dehydrogenase involved in the 4-aminobutyrate degradation pathway is totally absent. This may not affect functionality in this species as it has the genes *galE1* and *gabD2* encoding

proteins that perform functions similar to *galE2* and *gabD1*, respectively (Table 3). The important feature of the amino acid degradation pathway in *M. avium* ssp. *paratuberculosis* and *M. leprae* is the absence of the urease operon consisting of *ureA*, *ureB*, *ureC*, *ureD*, *ureF* and *ureG*. This indicates a low preference by *M. avium* ssp. *paratuberculosis* for ammonia as a nitrogen source probably as a result of the lower levels of ammonia in the intestine than in the lungs and might lead to differences in colonization and host-pathogen interactions (Clemens *et al.*, 1995; Burne & Chen, 2000). Two separate genes (*rocD1*, *rocD2*) encode ornithine aminotransferase in *M. bovis*, *M. leprae* and *M. tuberculosis* H37Rv whereas they are fused to form a single gene (*rocD1*) in *M. avium* ssp. *paratuberculosis* and *M. tuberculosis* CDC 1551.

### Central intermediary metabolism

There are minor differences in the genes encoding the proteins involved in central intermediary metabolism. *Mycobacterium avium* ssp. *paratuberculosis* lacks the gene *glpQ2*, encoding an esterase involved in the synthesis of glycerol-3 phosphate, and *sdA*, encoding a serine dehydratase involved in the conversion of serine to pyruvate. However, this might not have any effect as both losses seem to be compensated for by the presence of *glpQ1* and *ilvA*, which perform functions similar to *glpQ2* and *sdA*, respectively. All the taxa under study have the genes *rmlABCD* encoding proteins that are essential for the synthesis of rhamnose and the gene *wbbL* encoding a transferase that mobilizes rhamnose to the cell wall. In addition, all the organisms seem to have additional copies for some of the genes (*rmlA2*, *rmlB2*) in this pathway perhaps because rhamnose is an essential ingredient of the mycobacterial cell wall and there is no salvage pathway for rhamnose biosynthesis (Ma *et al.*, 2001). The gene *epiA* encoding a nucleotide-sugar epimerase is absent in *M. bovis*, and *epiB* is

**Table 3.** Variation in genes involved in the degradation of carbon compounds, amino acids and amines\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
<i>galT</i> , <i>galT</i>	Absent	Absent	<i>galT</i>	<i>galT</i>
<i>galE2</i>	MT0560	Absent	<i>galE3</i>	<i>MAP4031</i> , <i>4032</i>
<i>gabD1</i>	MT1772	Absent	<i>gabD2</i>	Absent
<i>glpK</i>	MT3798	<i>glpK</i>	<i>glpKa</i> , <i>glpKb</i>	<i>glpK</i>
<i>pykA</i>	MT1653	<i>pykA</i>	Pseudogene	<i>pykA</i>
<i>ugpA</i>	MT2901	Absent	<i>ugpAa</i> , <i>ugpAb</i>	Absent
<i>ureABC</i> <i>FGD</i>	MT1896-1901	Absent	<i>ureABC</i> <i>FGD</i>	Absent
<i>ald</i>	MT2850	<i>ald</i>	<i>alda</i> , <i>aldb</i>	<i>ald</i>
<i>rocD2</i> , <i>rocD1</i>	MT2384.1 <sup>†</sup>	ML1773, 1774	<i>rocD2</i> , <i>rocD1</i>	<i>rocD1</i> <sup>†</sup>

\*For a complete list of 40 genes see Table S5.

<sup>†</sup>Fused gene; genes in italics indicate split nonfunctional genes.

nonfunctional in *M. avium* ssp. *paratuberculosis*. *Mycobacterium tuberculosis* has two genes (*gca*, *gmdA*) encoding mannose dehydratase, *M. avium* ssp. *paratuberculosis* has only *gmdA* whereas *M. bovis* has only *gca* (Table 4). The genes *atsB*, *atsD*, *atsF* and *atsH* encoding arylsulfatases are absent in *M. avium* ssp. *paratuberculosis* and *M. leprae* whereas *atsA* is nonfunctional in *M. bovis*. The absence of the arylsulfatase genes in *M. avium* ssp. *paratuberculosis* and *M. leprae* coupled with the presence of minimal genes encoding cystathione  $\beta$ -synthase and thiosulfate sulfurtransferases might indicate a paucity of sulfated glycolipids in these species that will possibly result in differential host–pathogen interactions and reduced tolerance to stress (Mougous *et al.*, 2002). The genes involved in purine and pyrimidine nucleotide biosynthesis are highly conserved in all the species with an exception of two genes, *purT* (absent in *M. leprae*) and *purU* (absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*). The absence of *purU* might suggest that *M. avium* ssp. *paratuberculosis* and *M. leprae* will not be able to use *N*-formyl derivatives as precursors for the formation of formate. Most of the genes in the 2'-deoxyribonucleotide metabolism and nucleotide and nucleoside salvage pathways are either reduced to pseudogenes or lost in *M. leprae* whereas they are conserved in all the other species. The *treX* gene encoding a protein involved in trehalose metabolism is duplicated in *M. avium* ssp. *para-*

*tuberculosis* whereas the maltooligosyltrehalose synthase gene (*treY*) is disrupted in *M. bovis*. As a result, *M. bovis* may not be able to use glycogen as a precursor for the biosynthesis of trehalose, but this will not be critical as it has the genes encoding the enzymes of other two pathways (De Smet *et al.*, 2000).

## Lipid metabolism

Mycobacteria have a diverse array of molecules responsible for lipid metabolism. There are about 250 enzymes involved in this pathway, including homologs of those found in plants and animals. Similar to the pathways discussed above, most of the genes of *M. leprae* are either lost or reduced to pseudogenes, whereas *M. avium* ssp. *paratuberculosis* has a higher redundancy of genes in this pathway compared with *M. tuberculosis* (Li *et al.*, 2005). Because the cell wall is an interface between the pathogen and the host, the differences in the lipids constituting the cell wall might reflect the variations in pathogenesis among these species.

## Fatty acid biosynthesis

There are about 65 genes involved in the biosynthesis and modification of fatty acids, with the genes of mycolic acid biosynthesis conserved in all the species. Five genes consisting of three essential genes, *fabG2*, *accD4* and *kasB*, and two

**Table 4.** Variation in genes involved in central intermediary metabolism\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
glpQ2	MT0332	Absent	glpQ2	Absent
sdaA	MT0075	sdaA	sdaA	Absent
rmlB3	MT3574	Absent	rmlB2	Absent
wbbL2	MT1576	Absent	wbbL2	Absent
epiA	MT1562	Absent	Absent	epiA
epiB	MT3893	Absent	rfbB	Absent
gca	MT0121	Absent	gca	Absent
gmdA	MT1561	Absent	Absent	gmdA
atsA	MT0738	Pseudogene	<i>atsAb, atsAa</i>	atsA
atsB	MT3398	Absent	atsB	Absent
atsD	MT0692	Pseudogene	atsD	Absent
atsF	MT3162	Absent	Mb3104	Absent
atsH	MT3903	Absent	Mb3825	Absent
cysQ	MT2189	ML1301	cysQ	<b>cysQ1, cysQ2</b>
<b>sseC, sseC2</b>	MT3200	ML2199	sseC2	sseC
purU	MT3041	Absent	purU	Absent
nrdF	MT2033	Absent	nrdF1	Absent
nrdZ	MT0596	Absent	nrdZ	Absent
add	Absent	add	add	add
glgY	MT1614	Pseudogene	<i>treYa, treYb</i>	glgY
glgX	MT1615	Pseudogene	treX	<b>glgX1, glgX2</b>

Genes in bold type indicate multiple copies; genes in italics indicate split nonfunctional genes.

\*For a complete list of 46 genes see Table S5.

nonessential genes, *fabG3* and *fabG5* (Sasseti et al., 2003), are present as duplicates in the fatty acid biosynthesis pathway of *M. avium* ssp. *paratuberculosis* whereas the other taxa have a single copy (Table 5). The presence of multiple copies of these genes could be a possible mechanism to increase virulence as it has been demonstrated that mycobacteria can produce unique complex lipids by the combined action of fatty acid synthases and polyketide synthases (Kolattukudy et al., 1997). Mycobacteria have multiple copies of *fabG* (*fabG1*–*fabG5*), a gene encoding a  $\beta$ -ketoacyl carrier protein reductase that catalyses the first of the two reductive steps of the fatty acid synthesis cycle (Marrakchi et al., 2002). The presence of multiple copies of this gene might contribute to the virulence of mycobacteria (Banerjee et al., 1998). However, it will be interesting to see if all the copies of these genes are functional as it was recently demonstrated in *Lactococcus lactis* that only one of the two existing copies of *fabG* is actually functional (Wang & Cronan, 2004). The fatty acid biosynthesis genes *fabG1* and *inhA*, responsible for isoniazid resistance, are part of an operon in the case of *M. tuberculosis* and *M. bovis*, but they are expressed separately in the case of *M. avium* ssp. *paratuberculosis*, much as in *M. avium* and *Mycobacterium smegmatis* (Banerjee et al., 1998). The duplication of *accD4*, an essential gene of *M. tuberculosis* (Sasseti et al., 2003) that is involved in mycolic acid biosynthesis (Gande et al., 2004), might possibly indicate some additional and specific car-

boxylations in *M. avium* ssp. *paratuberculosis* that are probably involved in the synthesis of unusual lipids, whereas the duplication of *kasB*, a gene responsible for isoniazid resistance (Slayden & Barry, 2002), and *desA3*, a target for antimicrobial drug isoxyl (Phetsuksiri et al., 2003), might lead to an increased resistance of *M. avium* ssp. *paratuberculosis* against isoniazid and isoxyl. The gene Rv3472 encoding a hypothetical protein in *M. tuberculosis* H37Rv, the gene *acpA* encoding an acyl carrier protein, the gene Rv0914 encoding a lipid carrier protein and the gene *cdh* encoding a protein involved in phospholipid biosynthesis are absent in *M. avium* ssp. *paratuberculosis*. The genes Rv2261 and Rv2262 encoding hypothetical proteins in *M. tuberculosis* are fused to form a single protein in *M. bovis* and *M. tuberculosis* CDC 1551, whereas the corresponding genes are absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. These genes might possibly have some role in pathogenesis, as they are specifically present in TB-causing species.

### Fatty acid degradation

Genes involved in fatty acid degradation (*fad* genes) are fairly conserved in both strains of *M. tuberculosis* but show many differences with respect to *M. bovis* or *M. avium* ssp. *paratuberculosis*. The cholesterol oxidase gene, *choD*, is truncated in *M. avium* ssp. *paratuberculosis* whereas *fadH* is absent. The two genes *echA18* and *fadD11* are disrupted in

**Table 5.** Variation in genes of fatty acid metabolism and degradation\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC 1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
<i>fabG2</i>	MT1398	Pseudogene	<i>fabG2</i>	<i>fabG2_1, fabG2_2</i>
<i>fabG3</i>	MT2058	Absent	<i>fabG3</i>	<i>fabG3_1, fabG3_2</i>
<i>fabG5</i>	MT2836	Pseudogene	Mb2788c	<i>fabG5_1, fabG5_2</i>
<i>accD4</i>	MT3906	<i>accD4</i>	<i>accD4</i>	<i>accD4_1, accD4_2</i>
<i>fabG1</i>	MT1530	<i>fabG1</i>	<i>fabG1</i>	<b><i>fabG1</i></b>
<i>desA3</i>	MT3326	Pseudogene	Mb3258c	<i>desA3_1, desA3_2</i>
Rv3472	MT3578	Absent	Mb3501	Absent
Rv0033	MT0038	Absent	<i>acpA</i>	Absent
Rv0914c	MT0939	Pseudogene	Mb0938c	Absent
<i>Cdh</i>	MT2346	ML1417	<i>Cdh</i>	Absent
<i>Rv2261c, 2262c</i>	MT2322	Pseudogene	Mb2285c	Absent
<i>choD</i>	MT3517	<i>choD</i>	<i>choD</i>	<b><i>choD</i></b>
<i>fadH</i>	MT1212	Pseudogene	<i>fadH</i>	Absent
<i>echA18, echA18'</i>	Absent	Absent	<i>echA18</i>	Absent
<i>fadD11', fadD11</i>	MT1600	Absent	<i>fadD11</i>	<i>fadD11_1, fadD11_2</i>
<i>fadE22</i>	MT3147	Pseudogene	<i>fadE22a, fadE22b</i>	<i>fadE22</i>
<i>fadB3</i>	MT1754	Absent	<i>Fad3a, fad3b</i>	Absent
<i>echA3</i>	MT0660	Pseudogene	<i>echA3</i>	<b><i>echA3</i></b>
<i>Rv1136, 1137</i>	MT1169.1	Absent	Mb1168, 1169	<i>echA1_1</i>
<i>echA1</i>	MT0232	Absent	Absent	<i>echA1_2</i>

Genes in bold type indicate multiple copies; genes in italics indicate split nonfunctional genes.

\*For a complete list of 244 genes see Table S5.

**Table 6.** Variation in genes related to cell envelope\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
mmpL1	MT0412	Absent	mmpL1a,mmpL1b	Absent
mmpL2	MT0528	Absent	<b>mmpL2</b>	mmpL2
mmpL4	MT0466	ML2378	mmpL4	mmpL4_1- mmpL4_7
mmpL6	MT1608	ML2378	<b>mmpL6</b>	Absent
mmpL7	MT3012	mmpL7	mmpL7	Absent
mmpL8	MT3931	Absent	mmpL8	Absent
mmpL9	MT2402	Absent	mmpL9a,mmpL9b	Absent
mmpL12	MT1573	Absent	mmpL12	Absent
mmp13a, mmpL13b	MT1179.1	ML0971, ML0972	mmpL13	Absent
Absent	MT1802	Absent	mmpL14	Absent
mmpS1	MT0415	Absent	mmpS1,mmpS6	mmpS1
mmpS2	MT0527	Absent	mmpS2	Absent
lprM	MT2022	Absent	Absent	lprM
lprL	MT0623	Absent	Absent	lprL
lpqL	MT0432	Absent	lpqL	lpqL_1,lpq_2
lppS	MT2594	lppS	lppS	lppS_1,lppS_2

Genes in bold type indicate multiple copies; genes in italics indicate split nonfunctional genes.

\*For a complete list of 153 genes see Table S5.

*M. tuberculosis*. The gene *fadD11* is present in duplicate in *M. avium* ssp. *paratuberculosis* whereas *echA18* is absent. The genes *fadD27*, *fadE22* and *fadB3* are disrupted in *M. bovis*, whereas the gene *fadE27* is disrupted in *M. avium* ssp. *paratuberculosis*. *Mycobacterium avium* ssp. *paratuberculosis* has lost around 19 genes that are present in *M. tuberculosis* but has gained about 35 additional genes in the form of duplicates of existing genes. In a recent report on *M. tuberculosis*, it was shown that some of the fatty acid degradation (*fadD*) genes belong to a new class of fatty acyl-AMP ligases (FAALs) and that these *fadD* gene products can combine with *pkS* gene products in various ways to form complex hybrid metabolites (Trivedi *et al.*, 2004). The presence of multiple copies of various *fadD* genes in *M. avium* ssp. *paratuberculosis* might be seen as an adaptation of this organism to increase virulence and resistance by producing diverse metabolites compared with *M. tuberculosis* or *M. bovis*. The enoyl-CoA hydratase gene, *echA3*, in *M. avium* ssp. *paratuberculosis* is longer than in the other species owing to an insertion before the gene in the same reading frame. *Mycobacterium avium* ssp. *paratuberculosis* has two copies of the gene *echA1\_1* and *echA1\_2*; whereas *echA1\_2* is homologous to *echA1* of *M. tuberculosis*, it is absent in *M. bovis*. The second copy of the gene, *echA1\_1*, is split into two halves in *M. bovis* and *M. tuberculosis*. *echA1* is absent in *M. leprae*.

### Cell envelope

There are variations in the gene products constituting the cell envelope among the five taxa studied. These variations

might be due to the fact that the membrane proteins constitute important components of the mycobacterial cell wall and have a role in virulence and host specificity (Daffe & Etienne, 1999; Barry, 2001). All the genomes have the genes (*fbpA*, *fbpB*, *fbpC1*, *fbpC2*) belonging to the antigen 85 complex (Ag85). The Ag85 complex genes encode secreted proteins that have a role in pathogenesis and also catalyse the transfer of mycolates, leading to the formation of mycolated cell wall products such as  $\alpha, \alpha'$ -trehalose monomycolate (TMM) and  $\alpha, \alpha'$ -trehalose dimycolate (TDM) (Belisle *et al.*, 1997; Kremer *et al.*, 2002). The genes *mmpL1* and *mmpL9* encoding transmembrane proteins are split and appear to be nonfunctional in *M. bovis*, whereas they are totally absent in *M. avium* ssp. *paratuberculosis*. The gene *mmpL13* is disrupted in *M. tuberculosis* (Table 6). The deletion of a genomic region leads to the loss of the gene *mmpS6* and shortening of *mmpL6* in *M. tuberculosis* as compared with *M. bovis*, whereas both these genes are absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. Differences in the *mmpL* genes might have some effect on the transport of lipids to the membrane as some of these genes exist in close proximity with the polyketide synthesis (*pkS*) genes and are involved in the transport of lipids produced by its proximal *pkS* gene (Tekaiia *et al.*, 1999). Interestingly, in the case of the *M. bovis* *mmpL1* gene, even the downstream *pkS* gene, *pkS6*, is nonfunctional, indicating that the corresponding lipid produced by *pkS6* may not be required by *M. bovis*. However, the entire gene cluster is absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. *Mycobacterium avium* ssp. *paratuberculosis* also lacks another gene cluster, consisting of *mmpL7*, implicated in virulence in *M. tuberculosis*

(Camacho *et al.*, 1999, 2001; Jain & Cox, 2005). The *mmpL8* and *pks2* genes are also absent in *M. avium* ssp. *paratuberculosis* and *M. leprae*. These two genes are involved in sulpholipid biosynthesis, thereby contributing to virulence in *M. tuberculosis* (Converse *et al.*, 2003). All the regions appear to be precisely deleted in *M. avium* ssp. *paratuberculosis* and *M. leprae*, clearly suggesting alternative virulence factors. A similar case occurs with *mmpS6* and *mmpL6* genes where the corresponding region along with the upstream region containing *frdBCD* genes is lost in *M. avium* ssp. *paratuberculosis* and *M. leprae*. However, *M. avium* ssp. *paratuberculosis* has seven copies of the gene *mmpL4*. The presence of *mmpL4\_2* and *mmpL4\_3* downstream of *drABC* genes that encode proteins responsible for daunorubicin resistance suggests that these might encode membrane proteins that help in the efflux of this antibiotic. There is also a high degree of variation in the genes coding for lipoproteins that make up the cell envelope. Similar to the conserved membrane proteins, many of these are missing in *M. avium* ssp. *paratuberculosis* and *M. leprae* as compared with *M. tuberculosis* and *M. bovis*. While the gene *lprM* is absent in *M. bovis*, *lprP*, *lprQ* and *lprR* are present only in *M. bovis*. Whereas the *M. bovis* genes *lpqG* and *lpqL* are smaller than the corresponding genes in other mycobacteria, the genes *lppS* and *lpqL* are duplicated in *M. avium* ssp. *paratuberculosis*. *Mycobacterium leprae* has only a few functional lipoprotein genes. The presence of extensive variation in the cell envelope genes might lead to major differences in the virulence of these organisms.

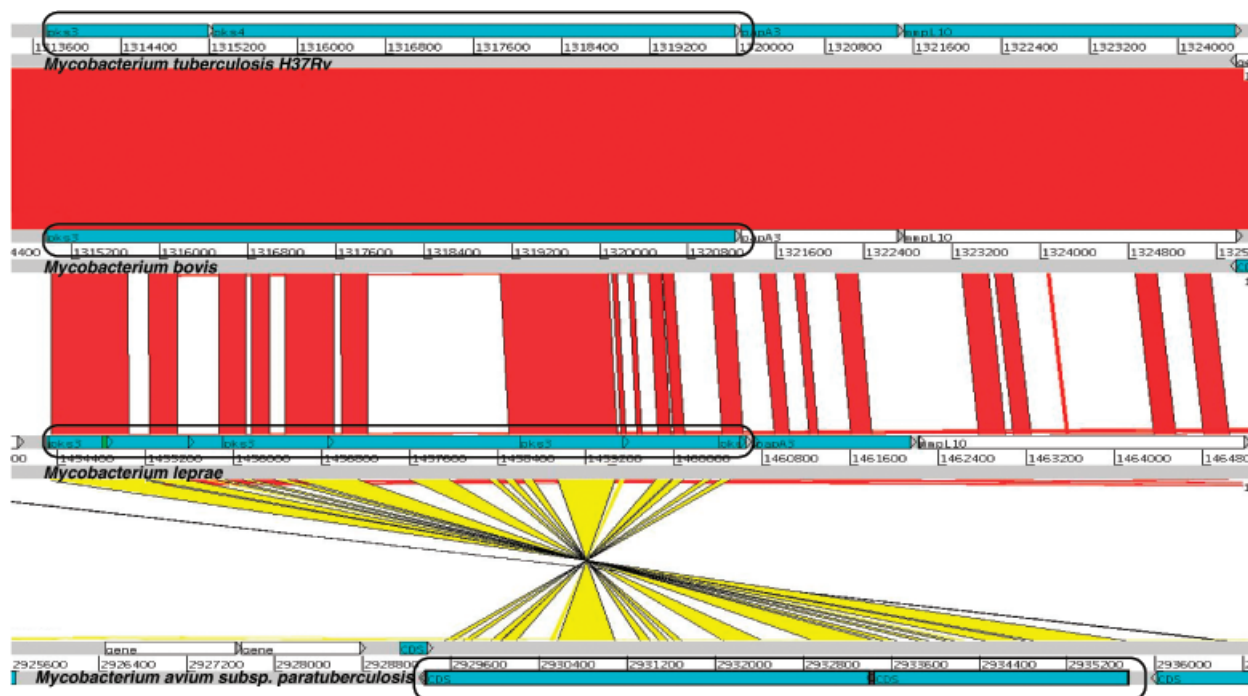
### Polyketide synthesis

There are many differences in the genes encoding polyketide synthases (*pks*) among the five organisms studied. Great variation occurs between *M. tuberculosis* and *M. bovis*, whereas many of the genes are lost in *M. leprae* and *M. avium* ssp. *paratuberculosis*. The gene *pks6* is split in *M. bovis*, and the genes *pks3* and *pks1* in *M. bovis* are fused genes with respect to *M. tuberculosis*. The gene *pks3* is a result of the fusion of *pks3* and *pks4* of *M. tuberculosis*, whereas *pks1* is a fusion of *pks1* and *pks15*. The genes *pks1* and *pks15* are absent in *M. avium* ssp. *paratuberculosis*, whereas the genes *pks3* (MAP2604) and *pks4* (MAP2603) are of a different length than in *M. tuberculosis* H37Rv (Fig. 4). The splitting or fusion of these genes might have an effect on virulence under *in vivo* conditions. Additionally, this might explain the differences in the virulence of mycobacterial species as these gene products can combine with the fatty acid degradation (*fadD*) gene products in different combinations to produce a variety of hybrid metabolites (Trivedi *et al.*, 2004). The gene cluster (*mbtA–mbtJ*) encoding the biosynthetic enzymes responsible for assembly of the virulence-conferring siderophore mycobactin and essential for growth

of *M. tuberculosis* in macrophages (Quadri *et al.*, 1998; De Voss *et al.*, 2000) is truncated in *M. avium* ssp. *paratuberculosis* (Li *et al.*, 2005). As a result, *M. avium* ssp. *paratuberculosis* needs to be supplemented with mycobactin in any growth medium (Quadri *et al.*, 1998; Li *et al.*, 2005). In spite of the split in the genes *mbtA* and *mbtB*, the order of the first eight genes (*mbtA–mbtH*) is maintained but an insertion in the genomic region between *trpE2* (*mbtJ*) and *lipK* (*mbtI*) and another between *lipK* (*mbtI*) and *mbtH* have led to the disruption of the operon. It will be interesting to see if this disruption in the operon structure of *M. avium* ssp. *paratuberculosis* has any bearing on its survival in macrophages. The gene cluster constituting the genes *ppsA–ppsE* encoding the polyketide phenolphthiocerol and the gene *mas* encoding mycocerosic acid, which together form pthiocerol dimycocerosate (DIM), are absent in *M. avium* ssp. *paratuberculosis*. This suggests a variation in the mode of virulence in *M. avium* ssp. *paratuberculosis* compared with other mycobacteria as the *pps* operon was implicated to have a role in virulence in pathogenic mycobacteria (Daffe & Laneelle, 1988; Azad *et al.*, 1997). The *mbt* operon is absent in *M. leprae*, whereas the virulence gene cluster comprising the *pps* operon is intact.

### PE and PPE gene families

The genome of *M. tuberculosis* contains two large families of acidic, glycine-rich proteins, the PE and the PPE gene families. These gene families, constituting about 10% of the genome in *M. tuberculosis* H37Rv, are the major source of divergence between the genomes of *M. tuberculosis* and *M. bovis*, which are otherwise > 99% similar (Garnier *et al.*, 2003). *Mycobacterium tuberculosis* has around 100 genes belonging to the PE gene family, consisting of PE and PE-PGRS genes, and about 60 genes belonging to the PPE gene family. Some of the characterized genes indicate that they are expressed based on the changing microenvironments encountered by the pathogen and play an important role in survival and multiplication of mycobacteria in their chosen environment (Brennan & Delogu, 2002; Voskuil *et al.*, 2004). Additionally, they represent a source of antigenic variation (Cole *et al.*, 1998; Choudhary *et al.*, 2003; Chakhaiyar *et al.*, 2004) and might interfere with the immune responses by inhibiting antigen processing (Talarico *et al.*, 2005). *Mycobacterium leprae* has almost no genes belonging to this family, whereas *M. avium* ssp. *paratuberculosis* has seven PE and 37 PPE genes. The absence of PE-PGRS genes in *M. avium* ssp. *paratuberculosis* might suggest a limited variation in the cell envelope proteins and altered colony morphology, as these proteins have been reported to be components of the cell envelope leading to extensive variation among mycobacteria (Espitia *et al.*, 1999; Brennan *et al.*, 2001; Banu *et al.*, 2002; Delogu *et al.*, 2004). The lack of PE-PGRS genes



**Fig. 4.** Diagrammatic representation of the *pks3* gene in *Mycobacterium* species. Mycobacterial species have many variations in the genes involved in polyketide synthesis (*pks*), providing them with variations in virulence factors. *Mycobacterium bovis* and *Mycobacterium leprae* have a single *pks3* gene (encircled by black oval) whereas it is split into two genes (*pks3*, *pks4*) in *Mycobacterium tuberculosis*. In *Mycobacterium avium* ssp. *paratuberculosis*, the genes *pks3* and *pks4* are of different sizes than in *M. tuberculosis* and are inverted (indicated by yellow criss-cross lines).

in *M. avium* ssp. *paratuberculosis* that are responsible for the survival of *M. tuberculosis* in macrophages (Ramakrishnan *et al.*, 2000) might even suggest variations in the survival mechanism of *M. avium* ssp. *paratuberculosis* inside macrophages. These variations in the PE/PPE gene families might lead to predominant differences in the pathogenesis between the mycobacterial species.

## Macromolecule metabolism and degradation

The genes encoding aminoacyl tRNA synthetases, ribosomal subunit proteins and proteins involved in translation are highly conserved across all the species. *Mycobacterium avium* ssp. *paratuberculosis* and *M. leprae* have lost a single gene, *rpmB*. The genes *hdsI*, *hdsM* and *mrr* encoding proteins involved in DNA restriction, *dinF* encoding a protein involved in DNA repair and *helZ* coding for DNA helicase are absent in *M. avium* ssp. *paratuberculosis* (Table 7). The absence of genes for DNA restriction might possibly indicate a higher rate of gene transfer in *M. avium* ssp. *paratuberculosis* (Marri *et al.*, 2006), whereas the absence of *dinF* and *helZ* might have an effect on the DNA repair mechanisms in this organism. The genes *alkA* and *recB* encoding proteins involved in DNA replication and repair are intact in the two *M. tuberculosis* strains. The gene *alkA* is

split in *M. bovis* whereas *recB* is split in both *M. avium* ssp. *paratuberculosis* and *M. bovis*. The truncation of the *M. bovis* *alkA* gene might not be critical for the survival of *M. bovis* but could affect the induction of an effective DNA repair response under nitrosative stress (Durbach *et al.*, 2003). However, the presence of a functional *recD* might suppress the defects in recombination caused as a result of the inactivation of *recB* (Amundsen *et al.*, 2000). *Mycobacterium leprae* lacks most of the genes of DNA replication and repair along with the *recBCD* operon. Most of the genes encoding proteins involved in DNA transcription are lost in *M. leprae* whereas they are conserved across the other three species. *Mycobacterium tuberculosis* has as many as 13 genes coding for sigma factors (Manganelli *et al.*, 1999). All the genes except *sigM* are conserved in *M. bovis* whereas *M. leprae* has only four functional genes (*sigA*, *sigB*, *sigC* and *sigE*). The *M. avium* ssp. *paratuberculosis* *sigC* gene is longer than the corresponding gene in other mycobacterial species due to an insertion before the gene, whereas the *sigK* gene is absent (Table 7). As *sigC* regulates the expression of virulence-related genes in *M. tuberculosis* (Sun *et al.*, 2004), it will be interesting to see if the *M. avium* ssp. *paratuberculosis* *sigC* regulates the same set of genes, as this might lead to variation in pathogenesis. The duplication of *sigF* encoding a transcription factor that controls the expression of genes responsible for mycobacterial persistence during

**Table 7.** Variation in genes of macromolecule metabolism and degradation\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
rpmB2	MT2118	Absent	rpmB2	Absent
rpmB	MT0114	Absent	rpmB1	Absent
mrr	MT2603	Absent	mrr	Absent
hsdS'	MT2825	Absent	hsdS'	Absent
hsdM	MT2826	Pseudogene	hsdM	Absent
dinF	MT2902	Absent	dinF	Absent
helZ	MT2160.1, MT2161	Absent	helZ	Absent
alkA	MT1358	Pseudogene	<i>alkAa, alkAb</i>	alkA
recB	MT0658	Absent	<i>recBa, recBb</i>	MAP4092c, 4093c
sigC	MT2129	ML1448	sigC	MAP1814
sigF	MT3385	Pseudogene	sigF	<b>sigF_1, sigF_2</b>
sigK	MT0461	Pseudogene	sigK	Absent
sigM	MT4030	Pseudogene	<i>sigMa, sigMb</i>	sigM
pcp	MT0334	Absent	pcp	Absent
Rv0198c	MT0208	ML2613	Mb0204c	Absent
Rv1977	MT2029	Absent	Absent	Absent
Rv3883c	MT3998	ML0041	Mb3913c	Absent
<i>ptrBa, ptrBb</i>	MT0805	ptrB	ptrB	ptrBa
plcC	MT2414	Absent	Absent	Absent
plcB	MT2415	Absent	Absent	Absent
plcA	MT2416	Pseudogene	Absent	Absent
lipS	MT3265	Absent	<i>mesTa, mesTb</i>	Absent
lipD	MT1974	Absent	lipD	Absent
lipF	MT3591	Absent	lipF	Absent
lipJ	MT1951	Pseudogene	lipJ	Absent
lipR	MT3169	Absent	lipR	Absent
Rv1105	Absent	Absent	Mb1135	Absent
mhpE	MT3575	Absent	mhpE	Absent

Genes in bold type indicate multiple copies; genes in italics indicate split nonfunctional genes.

\*For a complete list of 300 genes see Table S5.

chemotherapy might suggest an increased persistence of *M. avium* ssp. *paratuberculosis* in the host during chemotherapy (Michele *et al.*, 1999).

The genes encoding proteins involved in the degradation of DNA and RNA are highly conserved across all the species whereas the genes encoding proteins involved in the degradation of glycopeptides, polysaccharides, esterases and lipases have significant differences. The genes encoding pyrrolidone carboxyl peptidase (*pcp*) and Mycosin-I (Rv3883) are absent in *M. avium* ssp. *paratuberculosis*. The absence of the *pcp* gene might have some effect on protein folding and degradation (Kim *et al.*, 2001) while the absence of mycosin, a cell-wall-associated serine protease expressed after the infection of macrophages, might lead to variations in pathogenesis (Dave *et al.*, 2002). The gene *ptrB* encoding a protease involved in the degradation of oligopeptides is split in *M. tuberculosis* H37Rv whereas it is intact in other taxa. The genes *plcC*, *plcB* and *plcA* encoding phospholipase C are present only in *M. tuberculosis* strains. The presence of these genes specifically in *M. tuberculosis* and their role in virulence might make them potential targets for anti-TB drugs

(Johansen *et al.*, 1996; Raynaud *et al.*, 2002). The *lipS* gene is split into two in *M. bovis*, whereas it is completely absent in *M. avium* ssp. *paratuberculosis*. The other genes of the lipid degradation pathway absent in *M. avium* ssp. *paratuberculosis* include *lipD*, *lipf*, *lipJ*, *lipR* and Rv1105, probably indicating a lesser role of lipid catabolism.

## Regulatory genes

Mycobacteria are expected to have a wide array of regulatory proteins owing to the complexity of the environmental and metabolic choices for these organisms. *Mycobacterium tuberculosis* and *M. bovis* have 190 regulatory proteins, most of which are conserved between both species, whereas *M. avium* ssp. *paratuberculosis* has a higher number (235). *Mycobacterium avium* ssp. *paratuberculosis*, unlike *M. tuberculosis* and *M. bovis*, can survive in the environment outside the host and expansion of the regulatory gene repertoire could possibly help in its survival under a wide range of environmental conditions (Whittington *et al.*, 2004, 2005). The gene *sirR* encoding an iron-dependent repressor (Gupta

*et al.*, 1999) is nonfunctional in *M. avium* ssp. *paratuberculosis* and *M. leprae*, whereas the genes *virS* and *nadR* are absent (Table 8). The *virS* gene encoding a bacterial virulence-regulating protein is also absent in *M. smegmatis* and *M. avium*. The absence of *virS* further confirms its role as a regulator of genes that differentiate the *M. tuberculosis* complex from other mycobacterial species (Raffaelli *et al.*, 1999), whereas the possible effects of the absence of *nadR*, a gene encoding a bifunctional protein that has a role in the transport of nicotinamide mononucleotide, remain to be tested. Moreover, the *nadR* homolog in *M. tuberculosis* and *M. bovis* is shorter (323 amino acids) than the corresponding gene in *Escherichia coli* (417 amino acids), losing a portion of the N-terminal region, indicating a lack of repressor function (Hill *et al.*, 1998). *Mycobacterium avium* ssp. *paratuberculosis* has laterally acquired a second copy of the *narL* gene encoding a nitrate/nitrite regulatory protein (Marri *et al.*, 2006). The presence of the second copy of *narL* along with duplicate copies for some of the genes encoding nitrate/nitrite reductase and its inability to use ammonia might possibly indicate a preference of *M. avium* ssp. *paratuberculosis* for nitrate/nitrite as a nitrogen source. The presence of a duplicate copy of the gene *oxyS* might possibly make *M. avium* ssp. *paratuberculosis* more susceptible to organic hydroxyperoxide stress, as overproduction of *oxyS* reduces the expression levels of the gene *ahpC* encoding alkyl hydroperoxide reductase (Domenech *et al.*, 2001). The presence of *lysR*, a regulatory gene of *lysA* (absent in *M. tuberculosis* and *M. bovis*), suggests an *E. coli*-like expression of *lysA*, regulated by *lysR*, in the case of *M. avium* ssp. *paratuberculosis* (in *M. tuberculosis* and *M. bovis* the expression of *lysA* is constitutive) (Stragier *et al.*, 1983). *Mycobacterium bovis* and *M. avium* ssp. *paratuberculosis* have an

additional copy of the gene *embR* encoding a transcriptional activator of *embAB* genes. The *embAB* genes encode proteins that are involved in cell-wall arabinan biosynthesis and are the target for antimycobacterial drug ethambutol (Belanger *et al.*, 1996). The duplicate copy of *embR* might lead to an increased level of resistance in *M. bovis* and *M. avium* ssp. *paratuberculosis* against ethambutol. The genes Rv0600 and Rv0601 of *M. tuberculosis* H37Rv encoding proteins belonging to a two-component regulatory system are fused in *M. tuberculosis* CDC1551 whereas they are absent in *M. avium* ssp. *paratuberculosis* along with another gene, Rv2027.

The characteristic feature of the regulatory system of mycobacteria is the presence of eukaryotic-like Ser/Thr protein kinases (STPKs). *Mycobacterium tuberculosis* has 11 members of this protein kinase family which act as regulators of metabolic processes such as cell development, interaction with host cells and transcription (Av-Gay & Everett, 2000). The gene *pknD* is split into two in *M. bovis* (Peirs *et al.*, 2000), whereas the gene *pknH* has a deletion compared with that in *M. tuberculosis*. Although the putative active sites in *pknH* are conserved in *M. bovis* in spite of the deletion, this might have a bearing on its specificity for substrate (Garnier *et al.*, 2003). The nonfunctional *pknD* might affect phosphate transfer across the membrane in *M. bovis*, as *pknD* encodes a protein responsible for phosphate transfer (Av-Gay & Everett, 2000). While *M. leprae* has only four of the 11 STPKs, most of these are either disrupted or completely absent in *M. avium* ssp. *paratuberculosis*. The truncation of the gene *pknA* and loss of *pknI* might have an effect on the *in vivo* growth of *M. avium* ssp. *paratuberculosis*, as the products of these genes are involved in cell division (Chaba *et al.*, 2002). This might possibly be one of the reasons for a slower growth rate of *M. avium* ssp.

**Table 8.** Variation in genes involved in regulation\*

<i>Mycobacterium tuberculosis</i> H37Rv	<i>Mycobacterium tuberculosis</i> CDC1551	<i>Mycobacterium leprae</i>	<i>Mycobacterium bovis</i>	<i>Mycobacterium avium</i> ssp. <i>paratuberculosis</i>
sirR	MT2858	Pseudogene	sirR	<i>MAP2894,2895</i>
nadR	MT0222	Absent	nadR	Absent
narL	MT0866	Absent	narL	narL_1,narL_2
oxyS	MT0125.1	oxyS	oxyS	oxyS_1,oxyS_2
embR	MT1305	Absent	embR, embR2	embR_1,embR_2
<i>Rv0600c,0601c</i>	MT0630	Absent	<i>Mb0616c, 0617c</i>	Absent
pknA	MT0018	pknA	pknA	<b>pknA</b>
pknD	MT0958	Pseudogene	<i>pknDb,pknDa</i>	pknD
pknE	MT1785	Absent	pknE	Absent
pknI	MT2982	Pseudogene	pknI	Absent
pknJ	MT2149	Absent	pknJ	<b>pknJ</b>
pknK	MT3165	Absent	pknK	Absent

Genes in bold type indicate multiple copies; genes in italics indicate split nonfunctional genes; a single gene in bold type indicates one that has a different length in other organisms.

\*For a complete list of 190 genes see Table S5.

*paratuberculosis*. The disruption of *pknH* in *M. avium* ssp. *paratuberculosis* might have implications for the survival of this organism under heat or acid stress, as *pknH* phosphorylates a regulatory protein (product of the *embR* gene) involved in arabinan biosynthesis (Molle *et al.*, 2003; Sharma *et al.*, 2004). The genes *pknE* and *pknK* involved in membrane transport are totally absent. However, the losses in some of the STPKs appear to have been compensated for by the presence of a higher number (41 vs. 30 in *M. tuberculosis* H37Rv) of two-component regulator genes in *M. avium* ssp. *paratuberculosis*.

## Insights into pathogenesis

The cell wall serves as the interface between pathogen and host where all initial events of infection occur. With this in mind, the primary differences in the pathogenesis between these sequenced mycobacterial species are likely attributed to variations in the proteins and lipids constituting the cell wall. Compared with *M. tuberculosis* and *M. bovis*, *M. leprae* has a very limited variation in the lipids of the cell wall owing to fewer genes encoding lipid biosynthesis whereas *M. avium* ssp. *paratuberculosis* has a higher number and redundancy of these genes, indicating greater genetic ability to produce variation in the cell-wall lipid composition (Cole *et al.*, 1998, 2001; Garnier *et al.*, 2003; Li *et al.*, 2005). Moreover, the ability of mycobacteria to generate distinct lipids from various combinations of fatty acid degradation (*fadD*) genes and polyketide synthase (*pks*) genes (Trivedi *et al.*, 2004) would be of further use in generating diversity. This is especially true for *M. tuberculosis* and *M. bovis*, which have a similar number of genes encoding lipid biosynthesis, yet each species produces distinct lipids, especially sulfated lipids (Brodin *et al.*, 2004). However, the lipoarabinomanan (LAM) structure is the same among the species discussed herein (Rivera-Marrero *et al.* 2002).

Cell-surface proteins can mediate adherence to host cells or invasion of host tissues and are key players during these initial infection events. For example, it has been demonstrated previously that MAP2121c (also known as MMP for major membrane protein) is a surface protein that plays a role in invasion of bovine epithelial cells (Nigou *et al.*, 2003). This surface protein is coded for in the *M. avium* ssp. *paratuberculosis* and *M. leprae* genomes, but is absent in *M. bovis* and the two *M. tuberculosis* strains. One possible explanation for this distribution of MMP may lie in a shared host receptor that may be present in the Schwann's cells of the peripheral nerves, which *M. leprae* infects, and the intestinal epithelial cells that line the cattle gut, which *M. avium* ssp. *paratuberculosis* infects. This shared receptor, which is currently unknown, would not be present in the trachea or alveolar cells that *M. bovis* and *M. tuberculosis* are known to infect. A similar case can be made for the *mmpL5*

gene encoding a protein involved in lipid transport. This gene was specifically present in the cattle strains of *M. avium* ssp. *paratuberculosis* and absent in sheep strains, possibly indicating that some of these *mmpL* gene products could also help in host specificity (Marsh & Whittington, 2005).

Secondly, the PE/PPE/PGRS genes have been found to be a major source of antigenic difference between *M. tuberculosis* CDC 1551 and *M. tuberculosis* H37Rv (Fleischmann *et al.*, 2002) and also between *M. tuberculosis* and *M. bovis* (Garnier *et al.*, 2003). Some of the variations in virulence could also be attributed to the differential presence of genes in *M. tuberculosis* (Bannantine *et al.*, 2003) or *M. bovis* (Constant *et al.*, 2002) or differential expression of the virulence-related genes (Charlet *et al.*, 2005). The variation is again limited in *M. leprae*, which has lost most/all of the PE/PPE/PGRS genes. *Mycobacterium avium* ssp. *paratuberculosis* has only seven PE and 32 PPE genes and it has lost the genes *mmpL7*, *mmpL8*, *pks2*, *mas* and Rv3883 along with the genes in the *pps* operon that were shown to have a role in virulence in *M. tuberculosis* (Smith, 2003). However, the paucity of PE/PPE genes and the absence of some of these virulence genes is compensated for by the acquisition of virulence factors as a result of lateral gene transfer (Paustian *et al.*, 2005; Marri *et al.*, 2006).

Finally, each of the genomes has about one-third of the genes encoding proteins with unknown function. Some of these proteins might also contribute to pathogenesis and host specificity.

## Concluding remarks

The availability of five genome sequences of *Mycobacterium* species has provided a better understanding of the evolution of these species. The streamlining of most of the metabolic pathways and the presence of numerous pseudogenes in *M. leprae* suggests an evolutionary process towards its specialized growth in Schwann's cells, whereas an increased gene repertoire correlates with the ability of *M. avium* ssp. *paratuberculosis* to survive under diverse environmental conditions outside the host.

All the genomes have similar functional pathways for energy metabolism, amino acid biosynthesis, cofactor biosynthesis, nucleotide metabolism and macromolecule metabolism, with a majority of the genes conserved among the five genomes. Some of the key differences include the loss of the nitrate reductase, nitrite reductase, fumarate reductase, urease and NADH oxidase operons by *M. leprae*, resulting in curtailed growth under anaerobic and microaerophilic conditions. The loss of fumarate reductase and the urease operons in *M. avium* ssp. *paratuberculosis* indicates that nitrate is a major source of energy under anaerobic conditions. Furthermore, *M. avium* ssp. *paratuberculosis* has lost many of the genes (*mmpL7*, *mmpL8*, *pks2*, *mas*, *ppsABCD*)

implicated to have a role in virulence in *M. tuberculosis*, but has acquired some additional novel virulence factors by lateral gene transfer, suggesting different pathogenic pathways. The higher redundancy in *M. avium* ssp. *paratuberculosis* of the gene products that contribute to the cell-wall structure might enhance its ability to survive in the environment of the ruminant gut. Additionally, the presence of duplicate copies of *sigF* (a gene responsible for mycobacterial persistence during chemotherapy), *embR* (a regulator of genes involved in arabinan biosynthesis), *kasB* (a target for isoniazid) and *desA3* (a target for isoxyl) might indicate an increased resistance of *M. avium* ssp. *paratuberculosis* to some antimicrobial drugs. The similar number of genes in *M. tuberculosis* and *M. bovis* indicates that the differences in the pathogenesis between these species could result from variations in the PE/PPE/PGRS genes and the ability of these mycobacteria to produce a diversity of lipids from the combination of *fad* and *pks* gene products *in vivo*.

There remain about one-third of the genes in each of these genomes that are not functionally characterized to date. Characterization of these genes will not only help increase our understanding of the physiology and molecular biology of these pathogens, but will also enable us to gain insights into pathogenesis and host specificity, and help in creating improved drugs.

## Acknowledgements

We thank Weilong Hao for the gene comparison program and Drs Christian Baron, Marie Elliot and Justin Nodwell for their review of the manuscript. We thank the Editor and three anonymous reviewers for their insightful comments. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to G.B.G. We acknowledge funding by the USDA-CSREES and USDA-Agricultural Research Service to J.P.B.

## References

- Alsmark CM, Frank AC, Karlberg EO *et al.* (2004) The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc Natl Acad Sci USA* **101**: 9716–9721.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Amundsen SK, Taylor AF & Smith GR (2000) The RecD subunit of the *Escherichia coli* RecBCD enzyme inhibits RecA loading, homologous recombination, and DNA repair. *Proc Natl Acad Sci USA* **97**: 7399–7404.
- Av-Gay Y & Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* **8**: 238–244.
- Azad AK, Sirakova TD, Fernandes ND & Kolattukudy PE (1997) Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem* **272**: 16741–16745.
- Bai X, Zhang J, Holford IR & Hogenhout SA (2004) Comparative genomics identifies genes shared by distantly related insect-transmitted plant pathogenic mollicutes. *FEMS Microbiol Lett* **235**: 249–258.
- Banerjee A, Sugantino M, Sacchettini JC & Jacobs WR Jr (1998) The *mabA* gene from the *inhA* operon of *Mycobacterium tuberculosis* encodes a 3-ketoacyl reductase that fails to confer isoniazid resistance. *Microbiology* **144**: 2697–2704.
- Bannantine JP, Huntley JF, Miltner E, Stabel JR & Bermudez LE (2003) The *Mycobacterium avium* ssp. *paratuberculosis* 35 kDa protein plays a role in invasion of bovine epithelial cells. *Microbiology* **149**: 2061–2069.
- Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC & Cole ST (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* **44**: 9–19.
- Barry CE III (2001) Interpreting cell wall 'virulence factors' of *Mycobacterium tuberculosis*. *Trends Microbiol* **9**: 237–241.
- Belanger AE, Besra GS, Ford ME, Mikusova K, Belisle JT, Brennan PJ & Inamine JM (1996) 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* **93**: 11919–11924.
- Belisle JT, Vissa VD, Sievert T, Takayama K, Brennan PJ & Besra GS (1997) Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**: 1420–1422.
- Bertero MG, Rothery RA, Palak M *et al.* (2003) Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nat Struct Biol* **10**: 681–687.
- Betts JC, Lukey PT, Robb LC, McAdam RA & Duncan K (2002) Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43**: 717–731.
- Brennan MJ & Delogu G (2002) The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends Microbiol* **10**: 246–249.
- Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M & Jacobs WR Jr (2001) Evidence that mycobacterial PE PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* **69**: 7326–7333.
- Brodin P, Rosenkrands I, Andersen P, Cole ST & Brosch R (2004) ESAT-6 proteins: protective antigens and virulence factors. *Trends Microbiol* **12**: 500–508.
- Brosch R, Gordon SV, Eiglmeier K, Garnier T & Cole ST (2000) Comparative genomics of the leprosy and tubercle bacilli. *Res Microbiol* **151**: 135–142.
- Burne RA & Chen YY (2000) Bacterial ureases in infectious diseases. *Microbes Infect* **2**: 533–542.

- Camacho LR, Ensergueix D, Perez E, Gicquel B & Guilhot C (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* **34**: 257–267.
- Camacho LR, Constant P, Raynaud C *et al.* (2001) Analysis of the phthiocerol dimycocerosate locus of *Mycobacterium tuberculosis*. Evidence that this lipid is involved in the cell wall permeability barrier. *J Biol Chem* **276**: 19845–19854.
- Camus JC, Pryor MJ, Medigue C & Cole ST (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**: 2967–2973.
- Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG & Parkhill J (2005) ACT: the Artemis comparison tool. *Bioinformatics* **21**: 3422–3423.
- Chaba R, Raje M & Chakraborti PK (2002) Evidence that a eukaryotic-type serine/threonine protein kinase from *Mycobacterium tuberculosis* regulates morphological changes associated with cell division. *Eur J Biochem* **269**: 1078–1085.
- Chakhaiyar P, Nagalakshmi Y, Aruna B, Murthy KJ, Katoch VM & Hasnain SE (2004) Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. *J Infect Dis* **190**: 1237–1244.
- Charlebois RL & Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res* **14**: 2469–2477.
- Charlet D, Mostowy S, Alexander D, Sit L, Wiker HG & Behr MA (2005) Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in *sigK*. *Mol Microbiol* **56**: 1302–1313.
- Chaturvedi S & Bhakuni V (2003) Unusual structural, functional, and stability properties of serine hydroxymethyltransferase from *Mycobacterium tuberculosis*. *J Biol Chem* **278**: 40793–40805.
- Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Murthy KJ, Katoch VM & Hasnain SE (2003) PPE antigen Rv2430c of *Mycobacterium tuberculosis* induces a strong B-cell response. *Infect Immun* **71**: 6338–6343.
- Cirillo JD, Weisbrod TR, Banerjee A, Bloom BR & Jacobs WR Jr (1994) Genetic determination of the meso-diaminopimelate biosynthetic pathway of mycobacteria. *J Bacteriol* **176**: 4424–4429.
- Clemens DL, Lee BY & Horwitz MA (1995) Purification, characterization, and genetic analysis of *Mycobacterium tuberculosis* urease, a potentially critical determinant of host–pathogen interaction. *J Bacteriol* **177**: 5644–5652.
- Cole ST, Brosch R, Parkhill J *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537–544.
- Cole ST, Eiglmeier K, Parkhill J *et al.* (2001) Massive gene decay in leprosy bacillus. *Nature* **409**: 1007–1011.
- Constant P, Perez E, Malaga W, Laneelle MA, Saurel O, Daffe M & Guilhot C (2002) Role of the *pkc15/1* gene in the biosynthesis of phenolglycolipids in the *M. tuberculosis* complex: evidence that all the strains synthesize glycosylated *p*-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbour a frameshift mutation in *pkc15/1* gene. *J Biol Chem* **277**: 38148–38158.
- Converse SE, Mougous JD, Leavell MD, Leary JA, Bertozzi CR & Cox JS (2003) MmpL8 is required for sulfolipid-1 biosynthesis and *Mycobacterium tuberculosis* virulence. *Proc Natl Acad Sci USA* **100**: 6121–6126.
- Daffe M & Etienne G (1999) The capsule of *Mycobacterium tuberculosis* and its implications for pathogenicity. *Tuber Lung Dis* **79**: 153–169.
- Daffe M & Laneelle MA (1988) Distribution of phthiocerol diester, phenolic mycosides and related compounds in mycobacteria. *J Gen Microbiol* **134**: 2049–2055.
- Dave JA, Gey van Pittius NC, Beyers AD, Ehlers MR & Brown GD (2002) Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. *BMC Microbiol* **2**: 30.
- Delogu G, Pusceddu C, Bua A, Fadda G, Brennan MJ & Zanetti S (2004) Rv1818c-encoded PE PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol* **52**: 725–733.
- De Smet KA, Weston A, Brown IN, Young DB & Robertson BD (2000) Three pathways for trehalose biosynthesis in mycobacteria. *Microbiology* **146**: 199–208.
- De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y & Barry CE (2000) The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages. *Proc Natl Acad Sci USA* **97**: 1252–1257.
- Domenech P, Honore N, Heym B & Cole ST (2001) Role of OxyS of *Mycobacterium tuberculosis* in oxidative stress: overexpression confers increased sensitivity to organic hydroperoxides. *Microbes Infect* **3**: 713–721.
- Durbach SI, Springer B, Machowski EE *et al.* (2003) DNA alkylation damage as a sensor of nitrosative stress in *Mycobacterium tuberculosis*. *Infect Immun* **71**: 997–1000.
- Eppinger M, Baar C, Raddatz G, Huson DH & Schuster SC (2004) Comparative analysis of four Campylobacteriales. *Nat Rev Microbiol* **2**: 872–885.
- Espitia C, Lalette JP, Mondragon-Palomino M *et al.* (1999) The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiology* **145**: 3487–3495.
- Ferretti JJ, Ajdic D & McShan WM (2004) Comparative genomics of streptococcal species. *Indian J Med Res* **119**: 1–6.
- Fleischmann RD, Alland D, Eisen JA *et al.* (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184**: 5479–5490.
- Florczyk MA, McCue LA, Stack RF, Hauer CR & McDonough KA (2001) Identification and characterization of mycobacterial proteins differentially expressed under standing and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. *Infect Immun* **69**: 5777–5785.

- Gande R, Gibson KJ, Brown AK *et al.* (2004) Acyl-CoA carboxylases (accD2 and accD3), together with a unique polyketide synthase (Cg-pks), are key to mycolic acid biosynthesis in *Corynebacteriaceae* such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*. *J Biol Chem* **279**: 44847–44857.
- Garnier T, Eiglmeier K, Camus JC *et al.* (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci USA* **100**: 7877–7882.
- Gokulan K, Rupp B, Pavelka MS Jr, Jacobs WR Jr & Sacchettini JC (2003) Crystal structure of *Mycobacterium tuberculosis* diaminopimelate decarboxylase, an essential enzyme in bacterial lysine biosynthesis. *J Biol Chem* **278**: 18588–18596.
- Gupta S, Jain S & Tyagi AK (1999) Analysis, expression and prevalence of the *Mycobacterium tuberculosis* homolog of bacterial virulence regulating proteins. *FEMS Microbiol Lett* **172**: 137–143.
- Hewinson RG, Vordermeier HM, Smith NH & Gordon SV (2006) Recent advances in our knowledge of *Mycobacterium bovis*: a feeling for the organism. *Vet Microbiol* **112**: 127–139.
- Hill PJ, Cockayne A, Landers P, Morrissey JA, Sims CM & Williams P (1998) SirR, a novel iron-dependent repressor in *Staphylococcus epidermidis*. *Infect Immun* **66**: 4123–4129.
- Honer Zu Bentrup K, Miczak A, Swenson DL & Russell DG (1999) Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J Bacteriol* **181**: 7161–7167.
- Hutter B & Dick T (1999) Up-regulation of narX, encoding a putative 'fused nitrate reductase' in anaerobic dormant *Mycobacterium bovis* BCG. *FEMS Microbiol Lett* **178**: 63–69.
- Jain M & Cox JS (2005) Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in *M. tuberculosis*. *PLoS Pathog* **1**: e2.
- Johansen KA, Gill RE & Vasil ML (1996) Biochemical and molecular analysis of phospholipase C and phospholipase D activity in mycobacteria. *Infect Immun* **64**: 3259–3266.
- Keating LA, Wheeler PR, Mansoor H, Inwald JK, Dale J, Hewinson RG & Gordon SV (2005) The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactivative pyruvate kinase: implications for *in vivo* growth. *Mol Microbiol* **56**: 163–174.
- Kim JK, Kim SJ, Lee HG *et al.* (2001) Molecular cloning and characterization of *Mycobacterium bovis* BCG pcp gene encoding pyrrolidone carboxyl peptidase. *Mol Cells* **12**: 347–352.
- Kinsella RJ, Fitzpatrick DA, Creevey CJ & McNerney JO (2003) Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. *Proc Natl Acad Sci USA* **100**: 10320–10325.
- Kolattukudy PE, Fernandes ND, Azad AK, Fitzmaurice AM & Sirakova TD (1997) Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol* **24**: 263–270.
- Koonin EV & Galperin MY (2003) Evolution of central metabolic pathways: the playground of non-orthologous gene displacement. *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics*, pp. 295–355. Kluwer Academic Publisher, Dordrecht.
- Koonin EV, Makarova KS & Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709–742.
- Kremer L, Maughan WN, Wilson RA, Dover LG & Besra GS (2002) The *M. tuberculosis* antigen 85 complex and mycolyltransferase activity. *Lett Appl Microbiol* **34**: 233–237.
- Krogh A, Miam IS & Haussler D (1994) A hidden Markov model that finds genes in *E. coli* DNA. *Nucleic Acids Res* **22**: 4768–4778.
- Krzywinska E, Krzywinski J & Schorey JS (2004) Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. *Microbiology* **150**: 1707–1712.
- Li J, Zhu D, Yi Z, He Y, Chun Y, Liu Y & Li N (2002) DNazymes targeting the icl gene inhibit ICL expression and decrease *Mycobacterium tuberculosis* survival in macrophages. *Oligonucleotides* **15**: 215–222.
- Li L, Bannantine JP, Zhang Q *et al.* (2005) The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci USA* **102**: 12344–12349.
- Ma Y, Stern RJ, Scherman MS *et al.* (2001) 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* **45**: 1407–1416.
- Manganelli R, Dubnau E, Tyagi S, Kramer FR & Smith I (1999) Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol* **31**: 715–724.
- Marmiesse M, Brodin P, Buchrieser C *et al.* (2004) Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**: 483–496.
- Marrakchi H, Ducasse S, Labesse G *et al.* (2002) MabA (FabG1), a *Mycobacterium tuberculosis* protein involved in the long-chain fatty acid elongation system FAS-II. *Microbiology* **148**: 951–960.
- Marri PR, Bannantine JP, Paustian ML & Golding GB (2006) Lateral gene transfer in *Mycobacterium avium* subspecies *paratuberculosis*. *Can J Microbiol* **52**: 560–569.
- Marsh IB & Whittington RJ (2005) Deletion of an mmpL gene and multiple associated genes from the genome of the S strain of *Mycobacterium avium* ssp. *paratuberculosis* identified by representational difference analysis and *in silico* analysis. *Mol Cell Probes* **19**: 371–384.
- Marsh JJ & Leberer HG (1992) Fructose-bisphosphate aldolases: an evolutionary history. *Trends Biochem Sci* **17**: 110–113.
- Michele TM, Ko C & Bishai WR (1999) Exposure to antibiotics induces expression of the *Mycobacterium tuberculosis* sigF gene: implications for chemotherapy against mycobacterial persistors. *Antimicrob Agents Chemother* **43**: 218–225.

- Molle V, Kremer L, Girard-Blanc C, Besra GS, Cozzone AJ & Prost JF (2003) An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* **42**: 15300–15309.
- Moreira LM, de Souza RF, Almeida NF Jr et al. (2004) Comparative genomics analyses of citrus-associated bacteria. *Annu Rev Phytopathol* **42**: 163–184.
- Mougous JD, Green RE, Williams SJ, Brenner SE & Bertozzi CR (2002) Sulfotransferases and sulfatases in mycobacteria. *Chem Biol* **9**: 767–776.
- Munoz-Elias EJ & McKinney JD (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. *Nat Med* **11**: 638–644.
- Nakamura Y, Itoh T, Matsuda H & Gojobori T (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**: 760–766.
- Nascimento AL, Ko AI, Martins EA et al. (2004) Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J Bacteriol* **186**: 2164–2172.
- Nigou J, Gilleron M & Puzo G (2003) Lipoarabinomannans: from structure to biosynthesis. *Biochimie* **85**: 153–166.
- Paustian ML, Kapur V & Bannantine JP (2005) Comparative genomic hybridizations reveal genetic regions within the *Mycobacterium avium* complex that are divergent from *Mycobacterium avium* subspecies *paratuberculosis* isolates. *J Bacteriol* **187**: 2406–2415.
- Pavelka MS Jr & Jacobs WR Jr (1996) Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*. *J Bacteriol* **178**: 6496–6507.
- Peirs P, Parmentier B, De Wit L & Content J (2000) The *Mycobacterium bovis* homologous protein of the *Mycobacterium tuberculosis* serine/threonine protein kinase MbK (PknD) is truncated. *FEMS Microbiol Lett* **188**: 135–139.
- Phetsuksiri B, Jackson M, Scherman H et al. (2003) Unique mechanism of action of the thiourea drug isoxyl on *Mycobacterium tuberculosis*. *J Biol Chem* **278**: 53123–53130.
- Pierson DE & Campbell A (1990) Cloning and nucleotide sequence of bisC, the structural gene for biotin sulfoxide reductase in *Escherichia coli*. *J Bacteriol* **172**: 2194–2198.
- Prentice MB (2004) Bacterial comparative genomics. *Genome Biol* **5**: 338.
- Quadri LE, Sello J, Keating TA, Weinreb PH & Walsh CT (1998) Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol* **5**: 631–645.
- Raffaelli N, Lorenzi T, Mariani PL, Emanuelli M, Amici A, Ruggieri S & Magni G (1999) The *Escherichia coli* NadR regulator is endowed with nicotinamide mononucleotide adenylyltransferase activity. *J Bacteriol* **181**: 5509–5511.
- Ramakrishnan L, Federspiel NA & Falkow S (2000) Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* **288**: 1436–1439.
- Raynaud C, Guilhot C, Rauzier J et al. (2002) Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **45**: 203–217.
- Ren H & Liu J (2006) AsnB is involved in natural resistance of *Mycobacterium smegmatis* to multiple drugs. *Antimicrob Agents Chemother* **50**: 250–255.
- Rivera-Marrero CA, Ritzenthaler JD, Newburn SA, Roman J & Cummings RD (2002) Molecular cloning and expression of a novel glycolipid sulfotransferase in *Mycobacterium tuberculosis*. *Microbiology* **148**: 783–792.
- Rosenkrands I, King A, Weldingh K, Moniatte M, Moertz E & Andersen P (2000) Towards the proteome of *Mycobacterium tuberculosis*. *Electrophoresis* **21**: 3740–3756.
- Salzberg SL, Delcher AL, Kasif S & White O (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**: 544–548.
- Sasseti CM, Boyd DH & Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77–84.
- Schoolnik GK (2002) Functional and comparative genomics of pathogenic bacteria. *Curr Opin Microbiol* **5**: 20–26.
- Schupp T, Toupet C & Divers M (1988) Cloning and expression of two genes of *Streptomyces pilosus* involved in the biosynthesis of the siderophore desferrioxamine B. *Gene* **64**: 179–188.
- Sharma K, Chandra H, Gupta PK et al. (2004) PknH, a transmembrane Hank's type serine/threonine kinase from *Mycobacterium tuberculosis* is differentially expressed under stress conditions. *FEMS Microbiol Lett* **233**: 107–113.
- Slayden RA & Barry CE (2002) The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in *Mycobacterium tuberculosis*. *Tuberculosis* **82**: 149–160.
- Smith I (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* **16**: 463–496.
- Snider DE Jr, Raviglione M & Kochi A (1994) Global Burden of Tuberculosis. *Tuberculosis: Pathogenesis, Protection, and Control* (Bloom B, ed), pp. 3–12. ASM Press, Washington, DC.
- Stragier P, Borne F, Richaud F, Richaud C & Patte JC (1983) Regulatory pattern of the *Escherichia coli* lysA gene: expression of chromosomal lysA–lacZ fusions. *J Bacteriol* **156**: 1198–1203.
- Strominger JL (1962) Biosynthesis of bacterial cell walls. *Fed Proc* **21**: 134–143.
- Sun R, Converse PJ, Ko C, Tyagi S, Morrison NE & Bishai WR (2004) *Mycobacterium tuberculosis* ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol Microbiol* **52**: 25–38.
- Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L & Yang Z (2005) Variation of the *Mycobacterium tuberculosis* PE PGRS33 gene among clinical isolates. *J Clin Microbiol* **43**: 4954–4960.

- Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG & Cole ST (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* **79**: 329–342.
- Trivedi OA, Arora P, Sridharan V, Tickoo R, Mohanty D & Gokhale RS (2004) Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* **428**: 441–445.
- Uden G, Becker S, Bongaerts J, Holighaus J, Schirawski J & Six S (1995) O<sub>2</sub>-sensing and O<sub>2</sub>-dependent gene regulation in facultatively anaerobic bacteria. *Arch Microbiol* **164**: 81–90.
- Vissa VD & Brennan PJ (2001) The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol* **2**.
- Voskuil MI, Schnappinger D, Rutherford R, Liu Y & Schoolnik GK (2004) Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis* **84**: 256–262.
- Wang H & Cronan JE (2004) Only one of the two annotated *Lactococcus lactis* fabG genes encodes a functional beta-ketoacyl-acyl carrier protein reductase. *Biochemistry* **43**: 11782–11789.
- Whittington RJ, Marshall DJ, Nicholls PJ, Marsh IB & Reddacliff LA (2004) Survival and dormancy of *Mycobacterium avium* ssp. *paratuberculosis* in the environment. *Appl Environ Microbiol* **70**: 2989–3004.
- Whittington RJ, Marsh IB & Reddacliff LA (2005) Survival of *Mycobacterium avium* ssp. *paratuberculosis* in dam water and sediment. *Appl Environ Microbiol* **71**: 5304–5308.
- Wieles B, Nagai S, Wiker HG, Harboe M & Ottenhoff TH (1995) Identification and functional characterization of thioredoxin of *Mycobacterium tuberculosis*. *Infect Immun* **63**: 4946–4948.

## Supplementary material

The following supplementary material is available for this article online:

**Appendix S1.** Methods.

**Table S1.** List of unique genes in the *Mycobacterium* species.

**Table S2.** Variations in genes involved in energy metabolism.

**Table S3.** Variations in genes involved in amino acid biosynthesis.

**Table S4.** Variations in genes involved in biosynthesis of cofactors, prosthetic groups and carriers.

**Table S5.** Comprehensive list of genes involved in all the pathways in *Mycobacterium* species.

This material is available as part of the online article from <http://www.blackwell-synergy.com>