

Lateral gene transfer in *Mycobacterium avium* subspecies *paratuberculosis*

Pradeep Reddy Marri, John P. Bannantine, Michael L. Paustian, and G. Brian Golding

Abstract: Lateral gene transfer is an integral part of genome evolution in most bacteria. Bacteria can readily change the contents of their genomes to increase adaptability to ever-changing surroundings and to generate evolutionary novelty. Here, we report instances of lateral gene transfer in *Mycobacterium avium* subsp. *paratuberculosis*, a pathogenic bacteria that causes Johne's disease in cattle. A set of 275 genes are identified that are likely to have been recently acquired by lateral gene transfer. The analysis indicated that 53 of the 275 genes were acquired after the divergence of *M. avium* subsp. *paratuberculosis* from *M. avium* subsp. *avium*, whereas the remaining 222 genes were possibly acquired by a common ancestor of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* after its divergence from the ancestor of *M. tuberculosis* complex. Many of the acquired genes were from proteobacteria or soil dwelling actinobacteria. Prominent among the predicted laterally transferred genes is the gene *rsbR*, a possible regulator of sigma factor, and the genes designated *MAP3614* and *MAP3757*, which are similar to genes in eukaryotes. The results of this study suggest that like most other bacteria, lateral gene transfers seem to be a common feature in *M. avium* subsp. *paratuberculosis* and that the proteobacteria contribute most of these genetic exchanges.

Key words: mycobacteria, *M. avium* subsp. *paratuberculosis*, lateral gene transfer, unique genes, phylogeny.

Résumé : Le transfert génique latéral fait partie intégrante de l'évolution du génome chez la plupart des bactéries. Les bactéries peuvent couramment changer le contenu de leur génome afin d'augmenter l'adaptabilité à des environnements en changement constant et pour générer des nouveautés évolutives. Nous rapportons ici des cas de transferts de gènes latéraux chez *Mycobacterium avium* subsp. *paratuberculosis*, une bactérie pathogène causant la maladie de Johne chez la vache. Un ensemble de 275 gènes a été identifié comme pouvant avoir été récemment acquis par transfert génique latéral. Une analyse a indiqué que 53 des 275 gènes ont été acquis après la divergence de *M. avium* subsp. *paratuberculosis* de *M. avium* subsp. *avium* alors que les 222 gènes restants ont été potentiellement acquis par un ancêtre commun de *M. avium* subsp. *paratuberculosis* et *M. avium* subsp. *avium* après sa divergence de l'ancêtre du complexe *M. tuberculosis*. Plusieurs des gènes acquis provenaient de protéobactéries ou d'actinobactéries du sol. Parmi les gènes prédits transférés latéralement, on remarque *rsbR*, un régulateur potentiel de facteurs sigma, et les gènes désignés *MAP3614* et *MAP3757* qui sont semblables à des gènes chez les eucaryotes. Les résultats de cette étude indiquent que, comme la plupart des autres bactéries, les transferts géniques latéraux semblent être une caractéristique commune chez *M. avium* subsp. *paratuberculosis* et que les protéobactéries contribuent à la majorité de ces échanges génétiques.

Mots clés : mycobactéries, *M. avium* subsp. *paratuberculosis*, transfert génique latéral, gènes uniques, phylogénie.

[Traduit par la Rédaction]

Introduction

The number of genes in bacterial genomes is not stable (Doolittle 2000; Mira et al. 2001). Bacteria can readily change the contents of their genomes to increase adaptability to ever-changing surroundings and to generate evolutionary novelty. Lateral gene transfer (LGT) is an important feature

in the evolution of bacteria and provides a ready-to-use repertoire of genes that enables them to gain novel phenotypic and physiological characteristics (Hochhut et al. 2005; Ochman et al. 2000). Although the generation of evolutionary novelty can also be achieved by the modification and (or) duplication of existing genes, the acquisition of new genes by LGT appears to be a more frequent event (Dutta and Pan 2002; Gogarten et al. 2002; Gogarten and Townsend 2005; Lan and Reeves 1996).

The *Mycobacterium* genus contains over 70 species of human and animal pathogens as well as nonpathogenic saprophytes. Genetic studies to date indicate that unlike other bacteria, the *Mycobacterium tuberculosis* complex does not seem to exchange genetic material frequently (Smith et al. 2003; Sreevatsan et al. 1997; Supply et al. 2003). In spite of so many reports on LGT in other bacteria, there have been only a few instances of LGT in the chromosomal DNA in mycobacteria (Gamielien et al. 2002; Kinsella et al. 2003;

Received 28 November 2005. Revision received 4 January 2006. Accepted 16 January 2006. Published on the NRC Research Press Web site at <http://cjm.nrc.ca> on 14 June 2006.

P.R. Marri and G.B. Golding.¹ Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada.

J.P. Bannantine and M.L. Paustian. National Animal Disease Center, United States Department of Agriculture – Agriculture Research Service (USDA–ARS), Ames, IA 50010, USA.

¹Corresponding author (e-mail: golding@mcmaster.ca).

Koonin et al. 2001; Krzywinska et al. 2004; Nakamura et al. 2004; Paustian et al. 2005; Sassanfar et al. 1996; Tizard et al. 1998). It is generally believed that mycobacteria have very low instances of lateral transfer and that this phenomena appears to be comparatively rare in these species (Nakamura et al. 2004). In part, this is due to the parasitic life-style of these bacteria and the belief that this might limit opportunities for lateral transfer. Indeed, this understanding led Krzywinska et al. (2004) to state that LGT “has not been clearly demonstrated” in *Mycobacterium* spp., but they were able to demonstrate the limited transfer of a small 4100 bp region between two serotypes of *M. avium* subsp. *avium*.

The availability of the genome sequence of *M. avium* subsp. *paratuberculosis* (Li et al. 2005) provides a platform to test the extent and pattern of LGT in this pathogen. Unlike the other pathogenic mycobacteria, *M. tuberculosis* and *M. bovis*, which are spread directly from host to host, *M. avium* subsp. *paratuberculosis* spends part of its life-cycle in fecal material in pastures or in barns. This situation puts the organism in an environment perhaps more conducive for gene transfer events.

The objective of the present study was to identify the extent of LGT in *M. avium* subsp. *paratuberculosis* since it diverged from the ancestor of the *M. tuberculosis* complex. A comparative genomics approach supplemented with BLAST and phylogenetic clustering was employed to identify the laterally transferred genes. The study predicts that *M. avium* subsp. *paratuberculosis* has recently acquired 275 genes.

Materials and methods

Sequences used

The study involved eight mycobacterial genomes constituting five completely annotated genomes, *M. tuberculosis* H37Rv (NC_000962), *M. tuberculosis* CDC1551 (NC_002755), *M. leprae* TN (NC_002677), *M. bovis* AF212297 (NC_002945), and *M. avium* subsp. *paratuberculosis* (NC_002944); and three unannotated genomes, *M. marinum*, *M. smegmatis*, and *M. avium* subsp. *avium*. The sequences were downloaded from the NCBI (<ftp://ftp.ncbi.nih.gov/genomes/Bacteria/>).

Identification of unique genes

All the protein sequences from *M. avium* subsp. *paratuberculosis* were compared with that of *M. tuberculosis*, *M. bovis*, and *M. leprae*, using the BLASTP algorithm (Altschul et al. 1997). Comparisons were done with the expect value cutoff set at 1.0×10^{-10} and the filter query sequence option set to “false”. An additional criteria of a minimum match length (length of the query sequence that matches the target) was set at 30%. All proteins that did not have a corresponding match in any of the other four congeneric genomes were classified as unique. The uniqueness of each protein was further confirmed by TBLASTN and Artemis Comparison Tool (Carver et al. 2005).

Lateral gene transfer

The unique proteins of *M. avium* subsp. *paratuberculosis* were compared with the NCBI nr database using BLASTP with the expect value cutoff set at 1.0×10^{-10} to identify possible homologs in other organisms. For each protein, the

first 60 hits with an expect value less than 1.0×10^{-10} were chosen for further analysis regarding LGTs. The complete protein sequences of these 60 hits were extracted from the GenBank database and a multiple alignment was done using CLUSTALW (Thompson et al. 1994). This multiple alignment was used to generate a phylogenetic tree using the neighbor-joining method (Saitou and Nei 1987), as implemented in PHYLIP (Felsenstein 2004). An analysis of 100 bootstraps was done to validate the tree obtained. To identify the possible time scale of the acquisitions, all the genes that were identified as LGT were compared with the unannotated genomes of *M. marinum* (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/m_marinum), *M. smegmatis*, and *M. avium* subsp. *avium* (<http://tigrblast.tigr.org/cmr-blast/>) using TBLASTN.

The genes that had homologs in *M. marinum*, *M. smegmatis*, and *M. avium* subsp. *avium* and cluster within actinobacteria were assumed to be possible multiple deletions. The genes that had no hits in the NCBI database but are present in *M. marinum* and *M. smegmatis* were also assumed to be multiple deletions. A unique gene in *M. avium* subsp. *paratuberculosis* was labelled as laterally transferred (and unlikely to be multiply deleted) if (i) it had no hits in the NCBI database and was absent in *M. marinum*, *M. smegmatis*, and *M. avium* subsp. *avium*; (ii) other than itself, it had only a single hit in the database to a taxa outside of the *Mycobacterium*; (iii) it had multiple hits in the database but had two or fewer hits within the actinobacteria and clusters within the actinobacteria with a bootstrap value of 70% or above; or (iv) it clustered outside the actinobacteria with a bootstrap value of 70% or above.

Primer design and PCR

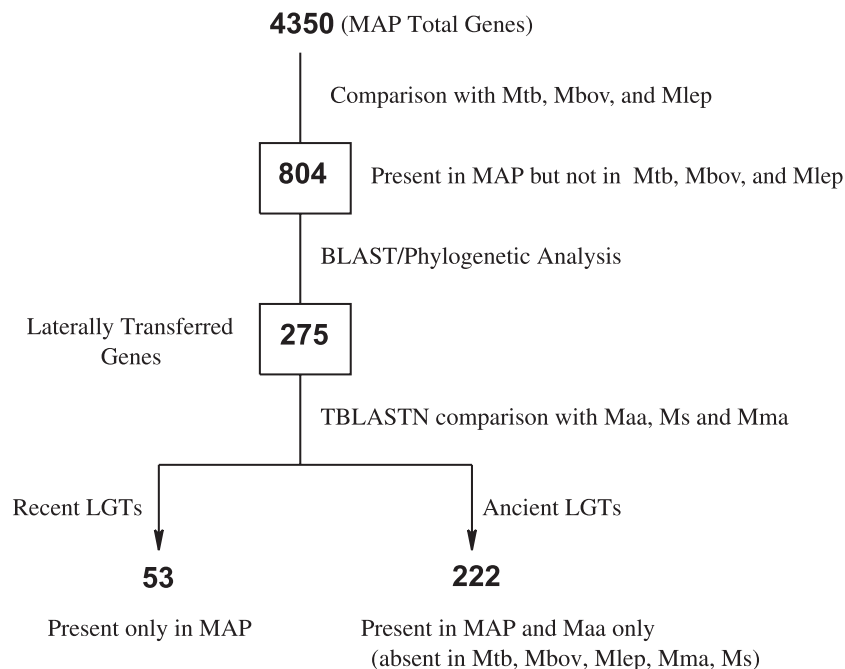
Six genes were randomly selected from those that are potentially LGTs, and primers were designed using Primer3 software (Rozen and Skaletsky 2004). These primers were used in PCR reactions to amplify the corresponding genes from 35 taxa comprising 11 distant mycobacterial species. PCRs consisted of 1× PCR Buffer II (Perkin-Elmer, Boston, Massachusetts), 0.2 mmol deoxynucleotide triphosphate mix/L, 3 mmol MgCl₂/L, 5% dimethyl sulfoxide (Sigma Chemical Co., St. Louis, Missouri), 0.04 U of Ampli-Taq Gold DNA polymerase (Applied Biosystems, Foster City, California)/μL, and 0.06 μmol/L (each) primer. The conditions used for PCR were an initial denaturation at 94 °C for 5 min followed by 35 cycles (each) of 45 s at 94 °C, 1 min at 58 °C, and 2 min at 72 °C.

Results

The initial comparison involving *M. tuberculosis*, *M. bovis*, and *M. leprae* resulted in the identification of 804 possible unique genes in *M. avium* subsp. *paratuberculosis*. The total gene complement in *M. avium* subsp. *paratuberculosis* is estimated to be 4350 genes (Li et al. 2005). This complement of genes is comparable to that observed in other proteobacteria. These genes were further analyzed to identify any cases of LGTs and multiple deletions. However, it should be noted that the list of unique genes is relative to the genomes used in this study and might change as new mycobacterial genomes become available. It is also likely that some of the

Table 1. Selected unique laterally transferred genes in *Mycobacterium avium* subsp. *paratuberculosis*.

Gene	Function	Organism in cluster	Bootstrap value (%)
MAP0083	Hypothetical protein	<i>Bacillus subtilis</i>	78
MAP0388	Hypothetical protein	<i>Nostoc</i> sp.	100
MAP0685	Hypothetical protein	<i>Erwinia carotovora</i>	100
MAP0956	Hypothetical protein	<i>Polaromonas</i> sp.	91
MAP1083	Hypothetical protein	<i>Bradyrhizobium japonicum</i>	92
MAP1086	Hypothetical protein	<i>Erwinia carotovora</i>	99
MAP1104	Hypothetical protein	<i>Novosphingobium aromaticivorans</i>	86
MAP1637	Hypothetical protein	<i>Sulfolobus tokadaii</i>	75
MAP1864	Hypothetical protein	<i>Cytophaga hutchinsonii</i>	99
MAP1989	Hypothetical protein	<i>Bradyrhizobium japonicum</i>	83
MAP2085	Hypothetical protein	<i>Brucella suis</i>	100
MAP2660	Hypothetical protein	<i>Xanthomonas</i> sp.	73
MAP3163	Hypothetical protein	<i>Photobacterium profundum</i>	100
MAP3614	Hypothetical protein	<i>Cryptococcus neoformans</i>	96
MAP3757	Hypothetical protein	Fungi	92
<i>bpoC_1</i>	Peroxidase	<i>Bacillus</i> sp.	77
<i>fliH</i>	Decarboxylase	<i>Rhodococcus</i> sp.	100
<i>rsbR</i>	Regulatory gene	<i>Nocardia farcinica</i>	100

Fig. 1. Flowchart for the identification of lateral gene transfers (LGTs) in *Mycobacterium avium* subsp. *paratuberculosis*. MAP, *M. avium* subsp. *paratuberculosis*; Maa, *M. avium* subsp. *avium*; Mtb, *M. tuberculosis*; Mbov, *M. bovis*; Mlep, *M. leprae*; Mma, *M. marinum*; Ms, *M. smegmatis*.

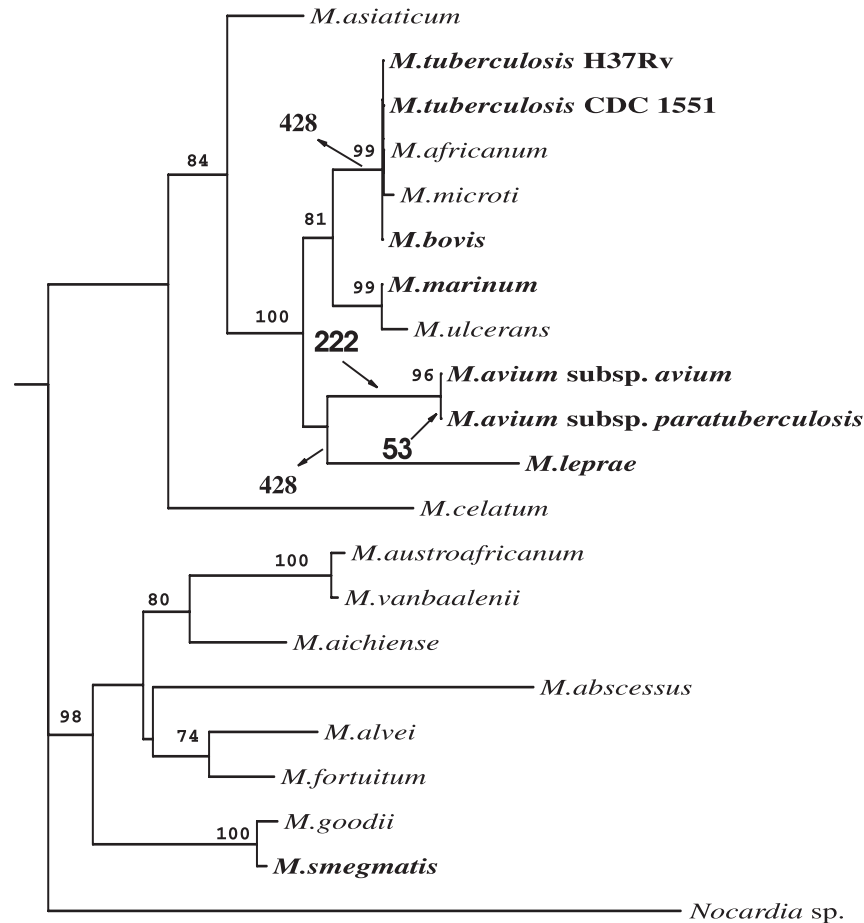
genes that are not identified as unique in *M. avium* subsp. *paratuberculosis* might be orthologous LGTs.

The analysis of the 804 genes using BLAST and phylogenetic methods identified 275 genes that are possibly acquired by *M. avium* subsp. *paratuberculosis* by LGT. The entire list of the 275 LGTs can be found at http://evol.mcmaster.ca/Mycobacterium_unique.html. Of the remaining 529 genes, 428 were present in more than two actinobacterial genomes and, hence, were thought to be possible cases of multiple deletions. The remaining 101 genes clustered outside actino-

bacteria, suggesting that they could be possible LGTs, but their phylogenetic clustering was not supported by high enough ($\geq 70\%$) bootstrap values. These genes were not considered for further analysis. For a selected group of laterally transferred genes, the taxa from which they potentially originated is shown in Table 1.

To possibly identify the evolutionary time scale of the transfers, we compared the 275 laterally transferred genes with the unannotated genomes of *M. avium* subsp. *avium*, *M. smegmatis*, and *M. marinum*. The comparison revealed

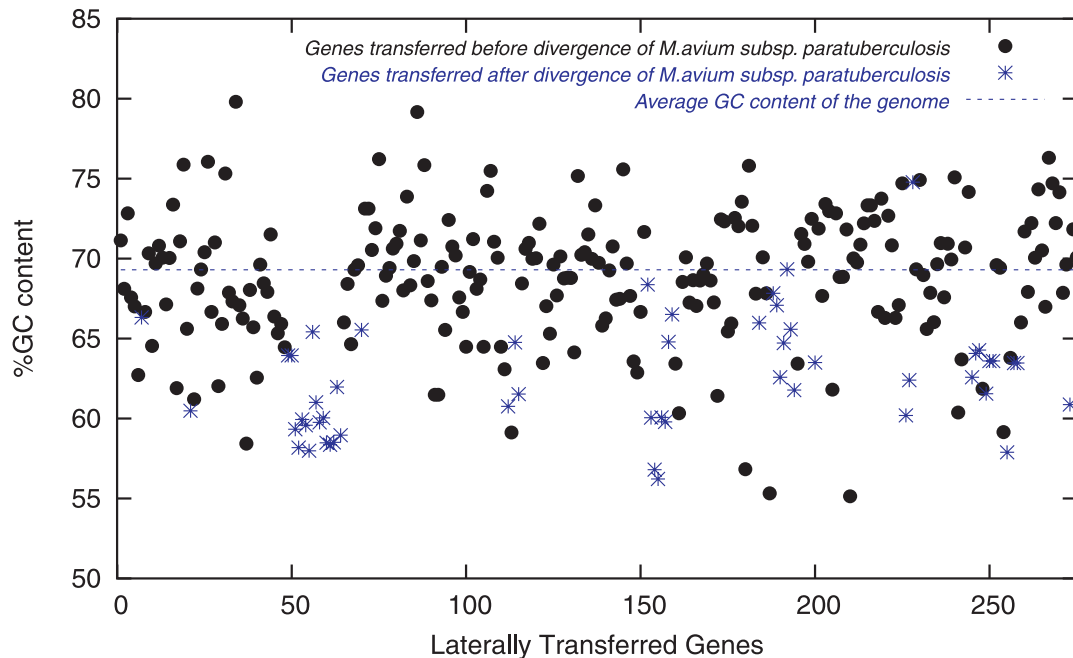
Fig. 2. Phylogenetic tree of mycobacterial species based on 16s rRNA generated using the neighbor-joining method of PHYLIP. The tree is rooted using *Nocardia* sp. The numbers on the branches indicate bootstrap values above 70%. The taxa in bold are used in the present study to identify lateral gene transfers (LGTs). Inward arrows indicate LGTs. Outward arrows indicate possible multiple deletions.



that all the genes were absent in *M. marinum* and *M. smegmatis*. While 53 genes were present only in *M. avium* subsp. *paratuberculosis*, 222 were also present in *M. avium* subsp. *avium* (Fig. 1). This suggests that 53 of the 275 laterally transferred genes might have been acquired after the very recent divergence of *M. avium* subsp. *paratuberculosis* from *M. avium* subsp. *avium*, whereas the remaining 222 that are shared by *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* (and absent in *M. tuberculosis*, *M. bovis*, *M. leprae*, *M. smegmatis*, *M. marinum*, and other actinobacteria) suggest that they might have been acquired by the common ancestor of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* after its divergence from the *M. tuberculosis* complex (Fig. 2). Most of the 53 *M. avium* subsp. *paratuberculosis* specific genes have a significantly low GC content (58%–61%) compared with the average GC content (69.3%) of the *M. avium* subsp. *paratuberculosis* genome (Fig. 3), a feature that has been used to suggest that such genes might be recent acquisitions by lateral transfer (Lawrence and Ochman 1998). The majority of the remaining 222 genes have an average GC content similar to that of the *M. avium* subsp. *paratuberculosis* genome, suggesting that they may have been acquired earlier and have been ameliorated during the evolution of *M. avium* subsp. *paratuberculosis* (Lawrence and Ochman 1997).

To further explore the history of the possible LGTs, we randomly PCR amplified six genes, *MAP0859*, *MAP1635*, *rsbR* (*MAP0379c*), *MAP0119*, *MAP0771*, and *MAP3508*, in 35 mycobacterial genomes comprising 11 divergent species. Of the six genes studied, *MAP0859* amplified in all the *M. avium* subsp. *paratuberculosis* strains used in the study but did not amplify in any other species. This suggests that the gene is either absent in the other species or too divergent to amplify under our PCR conditions. It is probably a lateral acquisition after the divergence of *M. avium* subsp. *paratuberculosis* from *M. avium* subsp. *avium* (Fig. 4A). The remaining five genes amplified in *M. avium* subsp. *paratuberculosis* along with all the subspecies of *M. avium*, suggesting a possible acquisition of these genes by their common ancestor (Fig. 4B; Table 2). The PCR analysis indicated that the gene *MAP1635* was also present in *M. smegmatis*, whereas the gene *MAP3508* was present in *M. smegmatis*, *M. scrofulaceum*, and *M. goodii*. However, the clustering of the gene *MAP1635* with the proteobacteria and *MAP3508* with *Sphingomonas* with a high bootstrap value of 81 and 100, respectively, suggests an orthologous LGT in the case of these two genes. Similarly, the gene *MAP0771*, while amplified in *M. goodii*, is identified as laterally transferred owing to its phylogenetic clustering with *Burkholderia* sp.

Fig. 3. Average GC content of the laterally transferred genes in *Mycobacterium avium* subsp. *paratuberculosis*. x axis, genes arranged in the increasing order of gene number.



Two of the 275 predicted laterally transferred genes, *MAP3614* and *MAP3757*, clustered with the eukaryotes. The gene *MAP3614* clustered with *Cryptococcus neoformans* with a bootstrap value of 96%, whereas *MAP3757* clustered with a group of fungi with a bootstrap support of 92%. Forty-eight genes clustered with proteobacteria, 11 clustered with firmicutes, and five clustered with cyanobacteria (Table 3). The gene *MAP1637* clustered with an Archaea, *Sulfolobus tokadaii*, with a bootstrap value of 75%. As an example, the analysis of the tree for *MAP1759* is presented in Fig. 5. It indicates that this hypothetical protein from *M. avium* subsp. *paratuberculosis* clusters with the genes from *B. japonicum* and *M. magnetotacticum*, with a clade credibility support of 95%. There were 42 genes that clustered with the genes of actinobacteria; *Streptomyces avermitilis* and *Nocardia farcinica* being the prominent members from this group. Interestingly, two of the laterally transferred genes *MAP1334* and *MAP3755* had homologs only in *M. ulcerans*. Though the presence of these genes only in *M. avium* subsp. *paratuberculosis* and *M. ulcerans* could suggest possible multiple deletions, their low GC content (61.49% and 61.56, respectively) and their absence in *M. smegmatis* and *M. marinum* indicate that they might have been recently acquired by *M. avium* subsp. *paratuberculosis* and *M. ulcerans* separately.

Analysis of the laterally transferred genes using the Artemis Comparison Tool (Carver et al. 2005) revealed that ten of the potential lateral transfer events contained gene clusters of three or more genes. Twenty-five of the 53 genes acquired after the divergence of *M. avium* subsp. *paratuberculosis* were a part of three large regions of about 13.5, 4.8, and 5.4 kb consisting of 14, 4, and 7 genes, respectively. The 13.5 kb cluster (*MAP0852*–*MAP0865*) is flanked by an insertion element upstream and an integrase downstream of the cluster.

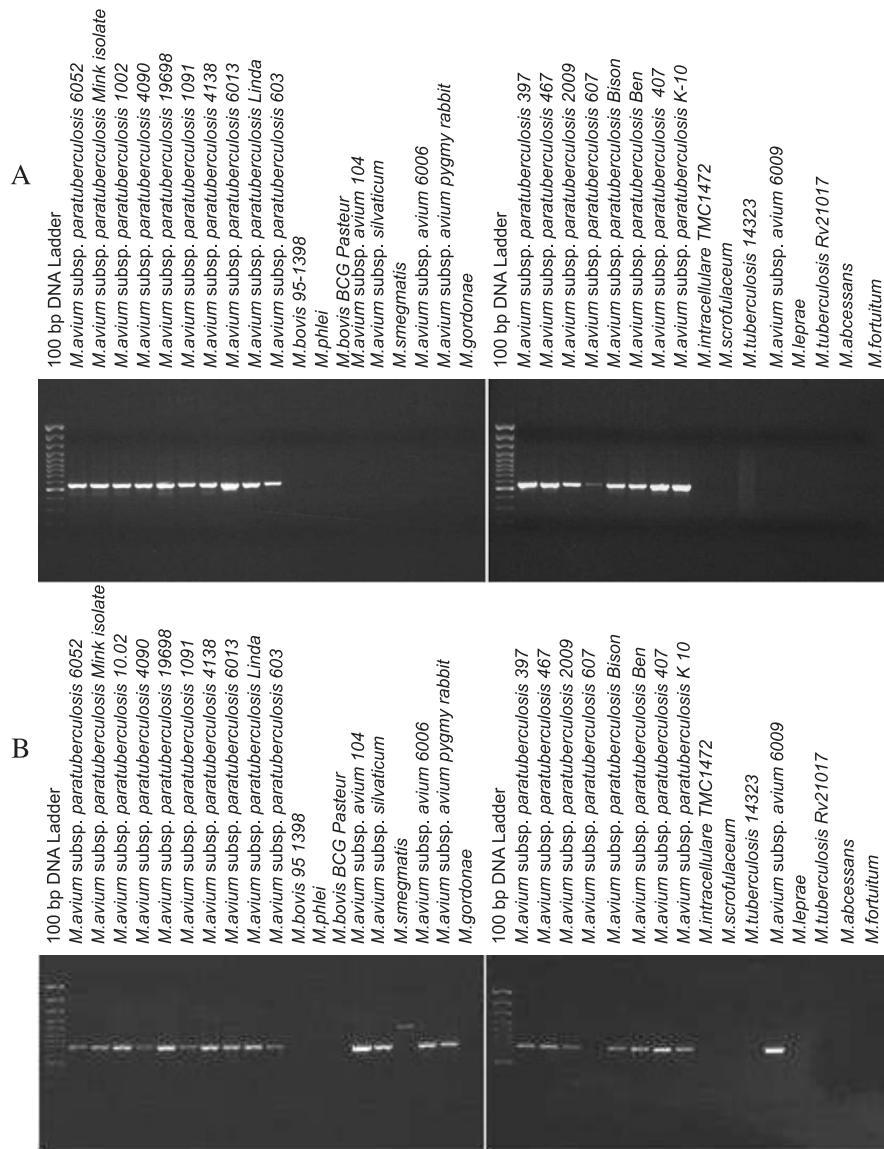
The 4.8 kb cluster (*MAP2151*–*MAP2154*) is flanked by insertion elements on both sides, providing further evidence for their potential acquisition by lateral transfer.

Only 30 out of the 275 predicted laterally transferred genes are functionally characterized to date, whereas the others are all hypothetical proteins. A closer look at the 30 functionally characterized genes revealed no clear indication about the specific kind of genes being transferred. While, five of them were informational genes, the rest were all operational genes with diverse functionalities.

Discussion

The acquisition of new genes by LGT accelerates the process of genome evolution (Jain et al. 2003). With the availability of large numbers of sequenced genomes, many studies are concentrating on the identification of laterally transferred genes among organisms (Garcia-Vallve et al. 2000).

There are several methods to detect LGT in bacteria. All methods rely on some unusual feature(s) that distinguish the laterally transferred genes from the remainder of the genes in the genome. Initial methods relied primarily on variations in codon usage and GC content. While these methods could successfully identify recent gene transfers from organisms with strongly different codon usage or GC content, they fail to identify more ancient transfers (because of the amelioration of genes over a period of time) or transfers from organisms with a similar codon usage or GC content (Lawrence and Ochman 1998). Moreover, codon usage and GC content are not always sufficiently reliable to detect LGTs (Koski et al. 2001). Other methods employ a BLAST-based approach to identify LGTs. These methods might also give erroneous results, as the best BLAST hit may not always be the closest (Koski and Golding 2001). Still other methods rely on a

Fig. 4. Polymerase chain reaction amplification of the *MAP0859* (A) and *rsbR* (B) genes in selected mycobacterial species.

phylogenetic approach, whereby, they construct a phylogenetic tree for genes from closely related organisms. Phylogenetic methods, which analyze a gene's relationship to genes from other organisms, are the most intuitive way of identifying laterally acquired genes (Lawrence and Ochman 2002). Here, we combine gene presence or absence information from a small taxonomic grouping and supplement it with BLAST, GC content, and phylogenetic information as a simple, practical, and reliable means to identify candidate LGT events.

The prediction of 275 laterally transferred genes indicates that *M. avium* subsp. *paratuberculosis* has acquired a considerable fraction (188 kb or 3.9%) of its genome since it diverged from its common ancestor with *M. tuberculosis* (a plausible history of these changes is given by the arrows in Fig. 2). Moreover, we presume that the results reported here are probably a lower estimate of the extent of LGT occurring in this organism, as there are another 101 genes that cluster with organisms outside the mycobacterial cluster but are not supported by high bootstrap values. This might be because

the donor species of these genes are not sequenced yet. With the sequencing of new genomes, the number of laterally transferred genes is bound to increase.

The 53 genes specific to *M. avium* subsp. *paratuberculosis* that are possibly acquired most recently could be used as diagnostic markers for *M. avium* subsp. *paratuberculosis*. The genes specific to *M. avium* subsp. *paratuberculosis* include *MAP0284*, *MAP0865*, and *MAP3774* that have also been identified as specific to *M. avium* subsp. *paratuberculosis* based on computational and microarray-based analysis (Semret et al. 2005). Moreover, the three gene clusters specific to *M. avium* subsp. *paratuberculosis* identified in this study were also shown to be specific to *M. avium* subsp. *paratuberculosis* based on a comparative genomics study involving species belonging to the *M. avium* subsp. *avium* complex (Paustian et al. 2005). Another gene *MAP2182* identified as laterally acquired in this study was found to encode a protein (HspX) responsible for cell attachment and was shown to be specific to *M. avium* subsp. *paratuberculosis*

Table 2. Polymerase chain reaction analysis of the randomly selected laterally transferred genes.

Organism (<i>Mycobacterium</i> sp.)	MAP1635	MAP0859	MAP0119	MAP0771	rsbR	MAP3508	16s rRNA
MAP K-10	+	+	+	+	+	+	+
MAP 1002	+	+	+	+	+	+	+
MAP 2009	+	+	+	+	+	+	+
MAP 4090	+	+	+	+	+	+	+
MAP 1091	+	+	+	+	+	+	+
MAP 19698	+	+	+	+	+	+	+
MAP 6013	+	+	+	+	+	+	+
MAP Ben	+	+	+	+	+	+	+
MAP Linda	+	+	+	+	+	+	+
MAP sheep 397	+	+	+	+	+	+	+
MAP sheep 407	+	+	+	+	+	+	+
MAP sheep 467	+	+	+	+	+	+	+
MAP sheep 603	+	+	+	+	+	+	+
MAP sheep 607	+	+	+	+	-	+	+
MAP bison 6012	+	+	+	-	+	+	+
MAP 6052	+	+	+	+	+	+	+
MAP milk isolate	+	+	+	+	+	+	+
MAP 4138	+	+	+	+	+	+	+
Maa 6009	+	-	+	+	+	+	+
Maa 104	-	-	+	+	+	+	+
Maa pygmy rabbit	-	-	+	+	+	+	+
<i>M. avium</i> subsp. <i>silvaticum</i>	+	-	+	+	+	+	+
Maa wood pigeon 6006	+	-	+	+	+	+	+
<i>M. smegmatis</i>	+	-	-	-	-	+	+
<i>M. tuberculosis</i> Rv21017	-	-	-	-	-	-	+
<i>M. tuberculosis</i> 14323	-	-	-	-	-	-	+
<i>M. leprae</i>	-	-	-	-	-	-	+
<i>M. scrofulaceum</i>	-	-	-	+	-	+	+
<i>M. gordonae</i>	-	-	-	-	-	+	+
<i>M. intracellulare</i>	-	-	-	-	-	-	+
<i>M. abscessans</i>	-	-	-	-	-	-	+
<i>M. fortuitum</i>	-	-	-	-	-	-	+
<i>M. bovis</i> 95-1398	-	-	-	-	-	-	+
<i>M. bovis</i> BCG	-	-	-	-	-	-	+
<i>M. phlei</i>	-	-	-	-	-	-	+

Note: MAP, *Mycobacterium avium* subsp. *paratuberculosis*; Maa, *Mycobacterium avium* subsp. *avium*. +, gene present; -, gene absent.

Table 3. Distribution of laterally transferred genes.

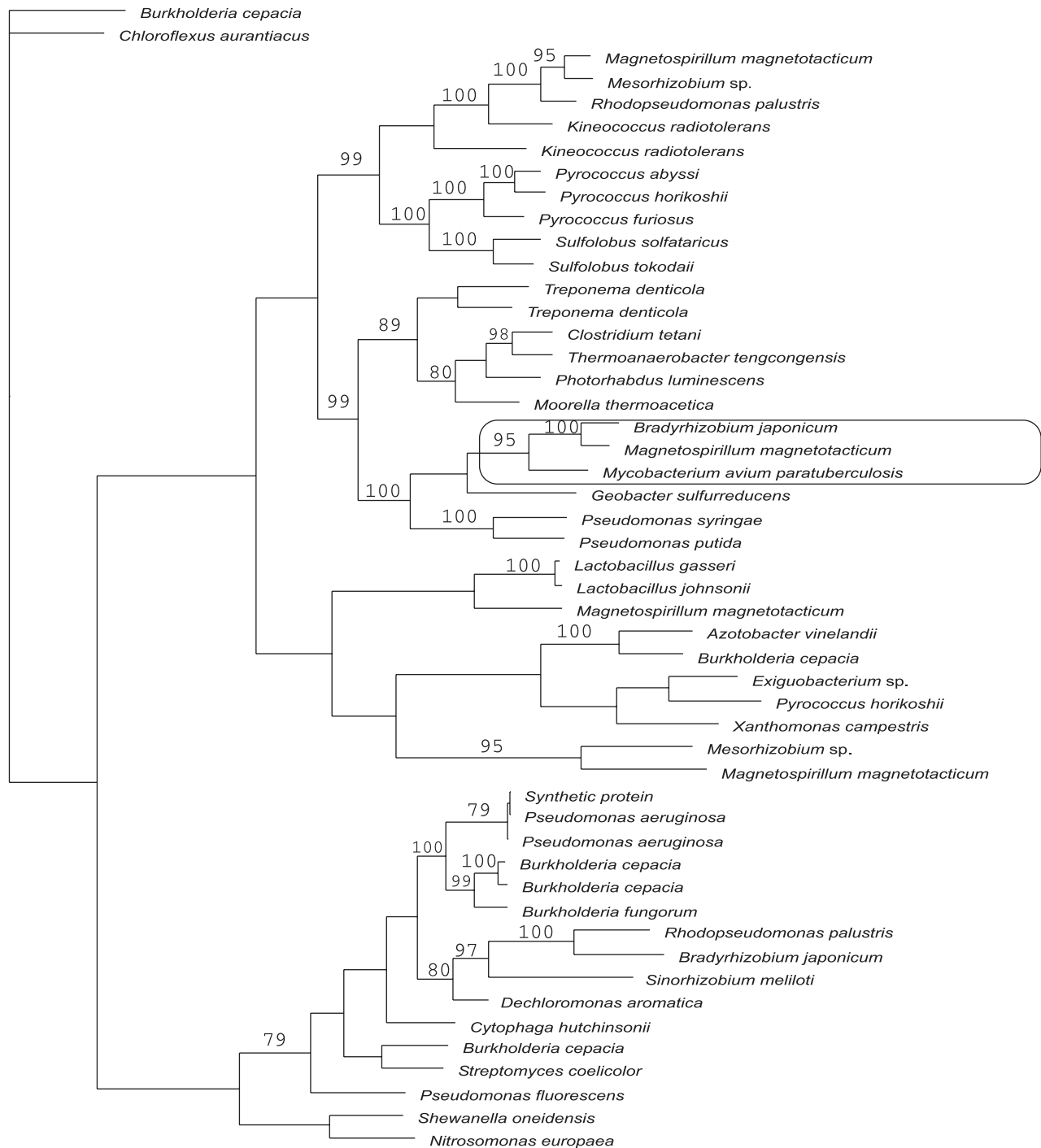
Donor group	No. of genes
Proteobacteria	48
Actinobacteria	42
Firmicutes	11
Cyanobacteria	5
Eukaryotes	2
Archaea	1
No hits	166
Total	275

(Ellingson et al. 1998). The specificity of these genes to *M. avium* subsp. *paratuberculosis* and their low GC content suggest that these could be pathogenicity islands that might

have been most recently acquired by *M. avium* subsp. *paratuberculosis* (after its divergence from *M. avium* subsp. *avium*) and may have a role in virulence and survival (Hacker and Kaper 2000; Hochhut et al. 2005; Oelschlaeger and Hacker 2004). The remaining 222 genes that are shared by *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* and that are absent in other mycobacterial and actinobacterial genomes constitute the set of genes that may have been acquired before the divergence of *M. avium* subsp. *paratuberculosis* from *M. avium* subsp. *avium*.

The GC content analysis indicated that some of the laterally transferred genes, consisting mostly of the genes acquired after the divergence of *M. avium* subsp. *paratuberculosis*, showed deviation from the average GC content of the *M. avium* subsp. *paratuberculosis* genome (Fig. 3), clearly indicating their recent acquisition by lateral transfer (Lawrence and Ochman 1998). However, a majority (183 of 275) of the laterally transferred genes had an average GC content similar to that in *M. avium* subsp. *paratuberculosis* genome,

Fig. 5. Phylogenetic tree for *MAP1759* generated using the neighbor-joining method of PHYLIP. The box indicates the cluster containing *Mycobacterium avium* subsp. *paratuberculosis*. The numbers on the branches indicate bootstrap values above 70%.



indicating that they may have been acquired early and ameliorated over time, or they might have been acquired from genomes having a similar GC content.

Only a fraction (10.9%) of the identified LGTs have been functionally characterized to date. A functional categorization of these genes did not give a clear indication of the specific kinds of genes transferred and why they were retained. While five genes belonged to informational pathways, the others were operational genes with diverse functionalities. The presence of a gene homologous to *rsbR*, a possible regulatory

gene of sigma B, might be seen as one of the possible mechanisms to tide over stress, as this gene has been found to be a positive regulator of sigma factor B under conditions of stress in *Bacillus subtilis* (Akbar et al. 1997). The majority of the LGTs identified in this study are hypothetical proteins that are yet to be characterized, but the fact that all these are absent in *M. tuberculosis* and that most of these genes are acquired from proteobacteria and soil-dwelling actinobacteria might indicate that some of these genes might aid the survival of *M. avium* subsp. *paratuberculosis* in soil. The

characterization of these hypothetical proteins will be important for a better understanding of the biology of *M. avium* subsp. *paratuberculosis*.

Most of the genes identified as LGTs and further analyzed have been acquired from proteobacteria. The transfers from the proteobacteria to *M. avium* subsp. *paratuberculosis* seem plausible because most of the originating species of proteobacteria are also soil dwelling (Brosch et al. 2000), in accord with the suggestion that lateral transfers are thought to be influenced by physical proximity rather than by phylogenetic proximity (Matte-Tailliez et al. 2002). *Mycobacterium avium* subsp. *paratuberculosis* has the ability to survive in soil (Whittington et al. 2004, 2005) and might be in close physical proximity with these taxa and hence have a greater opportunity to exchange genetic material.

These data prove that LGT is as common a feature in *M. avium* subsp. *paratuberculosis* as in most other bacteria. The fact that *M. avium* subsp. *paratuberculosis* exists in close proximity with proteobacteria might be one of the reasons for the high incidence of gene transfers in this organism. These findings are a precursor for similar studies in *M. tuberculosis*, and it will be interesting to know if such transfers are common in *M. tuberculosis* even though it is strictly intracellular.

Acknowledgements

We thank Dr. Richard Morton for his review and critical comments on the manuscript and Weilong Hao for the gene comparison program and useful discussions during the study. We thank the two anonymous reviewers for their helpful comments. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant to GBG. Portions of this study were supported by the John's Disease Integrated Program (JDIP) grant from USDA-CSREES to JPB.

References

- Akbar, S., Kang, C.M., Gaidenko, T.A., and Price, C.W. 1997. Modulator protein RsbR regulates environmental signalling in the general stress pathway of *Bacillus subtilis*. *Mol. Microbiol.* **24**: 567–578.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Brosch, R., Gordon, S.V., Eiglmeier, K., Garnier, T., and Cole, S.T. 2000. Comparative genomics of the leprosy and tubercle bacilli. *Res. Microbiol.* **151**: 135–142.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics*, **21**: 3422–3423.
- Doolittle, W.F. 2000. Uprooting the tree of life. *Sci. Am.* **282**: 90–95.
- Dutta, C., and Pan, A. 2002. Horizontal gene transfer and bacterial diversity. *J. Biosci.* **27**: 27–33.
- Ellingson, J.L., Bolin, C.A., and Stabel, J.R. 1998. Identification of a gene unique to *Mycobacterium avium* subspecies *paratuberculosis* and application to diagnosis of paratuberculosis. *Mol. Cell Probes*, **12**: 133–142.
- Felsenstein, J. 2004. Phylip: phylogeny inference package. University of Washington, Seattle, Wash.
- Gamielien, J., Ptitsyn, A., and Hide, W. 2002. Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. *Trends Genet.* **18**: 5–8.
- Garcia-Vallve, S., Romeu, A., and Palau, J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Res.* **10**: 1719–1725.
- Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**: 2226–2238.
- Gogarten, J.P., and Townsend, J.P. 2005. Horizontal gene transfer, genome innovation and evolution. *Natl. Rev. Microbiol.* **3**: 679–687.
- Hacker, J., and Kaper, J.B. 2000. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**: 641–679.
- Hochhut, B., Dobrindt, U., and Hacker, J. 2005. Pathogenicity islands and their role in bacterial virulence and survival. *Contrib. Microbiol.* **12**: 234–254.
- Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**: 1598–1602.
- Kinsella, R.J., Fitzpatrick, D.A., Creevey, C.J., and McInerney, J.O. 2003. Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 10320–10325.
- Koonin, E.V., Makarova, K.S., and Aravind, L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**: 709–742.
- Koski, L.B., and Golding, G.B. 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52**: 540–542.
- Koski, L.B., Morton, R.A., and Golding, G.B. 2001. Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol. Biol. Evol.* **18**: 404–412.
- Krzywinska, E., Krzywinski, J., and Schorey, J.S. 2004. Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. *Microbiology (Reading, U.K.)*, **150**: 1707–1712.
- Lan, R., and Reeves, P.R. 1996. Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* **13**: 47–55.
- Lawrence, J.G., and Ochman, H. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**: 383–397.
- Lawrence, J.G., and Ochman, H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 9413–9417.
- Lawrence, J.G., and Ochman, H. 2002. Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* **10**: 1–4.
- Li, L., Bannantine, J.P., Zhang, Q., Amonsin, A., May, B.J., Alt, D., Banerji, N., Kanjilal, S., and Kapur, V. 2005. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 12344–12349.
- Matte-Tailliez, O., Brochier, C., Forterre, P., and Philippe, H. 2002. Archaeal phylogeny based on ribosomal proteins. *Mol. Biol. Evol.* **19**: 631–639.
- Mira, A., Ochman, H., and Moran, N.A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**: 589–596.
- Nakamura, Y., Itoh, T., Matsuda, H., and Gojobori, T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat. Genet.* **36**: 760–766.
- Ochman, H., Lawrence, J.G., and Groisman, E.A. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature (London)*, **405**: 299–304.

- Oelschlaeger, T.A., and Hacker, J. 2004. Impact of pathogenicity islands in bacterial diagnostics. *APMIS*, **112**: 930–936.
- Paustian, M.L., Kapur, V., and Bannantine, J.P. 2005. Comparative genomic hybridizations reveal genetic regions within the *Mycobacterium avium* complex that are divergent from *Mycobacterium avium* subsp. *paratuberculosis* isolates. *J. Bacteriol.* **187**: 2406–2415.
- Rozen, S., and Skaletsky, H.J. 2004. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, N.J.
- Saitou, N., and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–422.
- Sassanfar, M., Kranz, J.E., Gallant, P., Schimmel, P., and Shiba, K. 1996. A eubacterial *Mycobacterium tuberculosis* tRNA synthetase is eukaryote-like and resistant to a eubacterial-specific anti-synthetase drug. *Biochemistry*, **35**: 9995–10003.
- Semret, M., Alexander, D.C., Turenne, C.Y., de Haas, P., Overduin, P., van Soolingen, D., Cousins, D., and Behr, M.A. 2005. Genomic polymorphisms for *Mycobacterium avium* subsp. *paratuberculosis* diagnostics. *J. Clin. Microbiol.* **43**: 3704–3712.
- Smith, N.H., Dale, J., Inwald, J., Palmer, S., Gordon, S.V., Hewinson, R.G., and Smith, J.M. 2003. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc. Natl. Acad. Sci. U.S.A.* **100**: 15271–15275.
- Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., and Musser, J.M. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 9869–9874.
- Supply, P., Warren, R.M., Banuls, A.L., Lesjean, S., Van Der Spuy, G.D., Lewis, L.A., Tibayrenc, M., Van Helden, P.D., and Locht, C. 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol. Microbiol.* **47**: 529–538.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Tizard, M., Bull, T., Millar, D., Doran, T., Martin, H., Sumar, N., Ford, J., and Hermon-Taylor, J. 1998. A low G + C content genetic island in *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* with homologous genes in *Mycobacterium tuberculosis*. *Microbiology*, **144**: 3413–3423.
- Whittington, R.J., Marshall, D.J., Nicholls, P.J., Marsh, I.B., and Reddacliff, L.A. 2004. Survival and dormancy of *Mycobacterium avium* subsp. *paratuberculosis* in the environment. *Appl. Environ. Microbiol.* **70**: 2989–3004.
- Whittington, R.J., Marsh, I.B., and Reddacliff, L.A. 2005. Survival of *Mycobacterium avium* subsp. *paratuberculosis* in dam water and sediment. *Appl. Environ. Microbiol.* **71**: 5304–5308.