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

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

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

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

Summary

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

Introduction

Repetitive sequences are common constituents of the genomes of all living organisms although they are far more prominent in higher eukaryotes where they can account for a substantial percentage of the chromosomal DNA. There are two principal forms of repetitive DNA in bacterial genomes: dispersed and tandem repeats. Dispersed repetitive sequences can correspond to duplicated genes, or to mobile genetic elements present in several copies like insertion sequences (IS). IS are often an important component of bacterial genomes and as a result of their ability to transpose have mutational potential based on their ability to locate within coding or regulatory regions. Hundreds of individual IS have been described and grouped into 17 families on the basis of their genetic organization, sequence similarities in their recombinases/transposases, the similarity of their ends (direct or terminal inverted repeats) and their target sites which are often duplicated during transposition. The genome sequence of *Mycobacterium tuberculosis* H37Rv contains more than 56 IS, belonging to eight
families and these are an important source of plasticity and genetic variability. A novel repeated sequence, the REP13E12 family, is present in seven copies on the chromosome and contains a probable phage attachment site. A large portion of the genome has also evolved from gene duplication events, followed by sequence divergence, leading to functional redundancy and expansion of the biological potential of the tubercle bacillus.7

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

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

Materials and methods

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

To investigate MIRU-based polymorphism, PCR primers were designed using the Oligo 5.0 software (National Biosciences, Plymouth, MN, USA), and the sequences are summarized in Table 1. M. leprae DNA was prepared by the freeze-boiling method. For PCR reactions, 5 μl of DNA solutions was added to a final volume of 25 μl containing 10% DMSO, 0.5 mmol/l of each dATP, dCTP, dGTP and dTTP, 0.2 μmol/l of primers, 2.5 μl of PCR buffer [170 mmol/l (NH4)2SO4, 600 mmol/l Tris-HCL (pH 8.8), 20 mmol/l MgCl2, 100 mmol/l β-mercaptoethanol] and 1.25 IU of Taq polymerase (Gibco-BRL). The PCR was performed using a PTC-100 (MJ Research, Inc.) for 35 cycles of 1 min at 94°C, 2 min at 59°C, 2 min at 72°C. The reactions were terminated by incubating for 10 min at 72°C and analysed by agarose gel electrophoresis using the appropriate controls.

Results and discussion

DIRECT REPEATS

On examination of the M. leprae genome sequence a series of perfect direct repeats was found ranging in size from 2 to 52 bp. All repeat sequences of >20 bp that were present in
| PRIMER F | SEQUENCE | PRIMER R | SEQUENCE | GENES |
|----------|----------|----------|----------|-------|
| B937-MIRU1F | GTGCTGACCACGCTATCCTGA | B937-MIRU1R | CCCCGGACCAGATTTCTAC | MLO534...carA |
| B1308-MIRU1F | GTTCTTGTGTCCGCGGTAGT | B1308-MIRU1R | TTACGAGCTGTTATGAAACTG | MLO719...aldB |
| B2235-MIRU1F | GGTGGCCGCCTGAGTCAC | B2235-MIRU1R | GAGGGGATGCGCACCATTG | miaA...dapF |
| B1764-MIRU1F | GGGCTTTATCCTGTAACAG | B1764-MIRU1R | GCCGTTAGAGCCACCAAC | dut...MLO1029 |
| B1764-MIRU2F | TTCACGCGAATCCAGGTCAGAC | B1764-MIRU2R | CTAGCAGGGAGGAGGCAAAAC | MLO1042...MLO1041 |
| L471-MIRU1F | CCCAGGACCCGACGAG | L471-MIRU1R | GACCGCTGATGTGTTG | MLO1135...MLO1136 |
| B1549-MIRU2F | TACCAGGGAGCCGAGTCAT | B1549-MIRU2R | CGGACGCTGCTGACCAC | cysM...MLO1171 |
| B1549-MIRU1F | GTTCAGGATACACAGGCAGCA | B1549-MIRU1R | TCAAGGGACTGGTAGGG | rplA...MLO1175 |
| B1133-MIRU1F | TGACCCTCGTTTTTGT | B1133-MIRU1R | GTGCCGCGAGGTCTGTTC | argD...argF |
| B2266-MIRU2F | GGAAGGCACTGGCAAGTCGT | B2266-MIRU2R | TACCGGCCACAGAC CCTG | MLO2199...MLO2200 |
| B2266-MIRU1F | TGACGCTGACACTAGACACAG | B2266-MIRU1R | CGTTCACGCGACTCAC | pabCM12203 |
| B2168-MIRU2F | CGGCGGTTGCTGCTAGAAGA | B2168-MIRU2R | TGACCGCGCAAGCGACTTTTG | MLO2412...MLO2413 |
| B2168-MIRU3F | TGGGCTCAAACCTCCTTTGC | B2168-MIRU3R | GGGCTGGCATTGCTCAAAC | MLO2439...MLO2440 |
| B2168-MIRU1F | GATGCGCTGCGTTTCTGAG | B2168-MIRU1R | GCACTTTGTGTGCTGGAAT | MLO2442...MLO2443 |
two or more identical copies were annotated together with all perfect repeats present in three or more copies. There are far fewer tandem repeats in the genome of *M. leprae* compared to that of *M. tuberculosis*, mainly as a result of the much smaller number of PE-PGRS genes that are composed of such motifs. No tetranucleotide repeats are found in *M. leprae* nor in *M. tuberculosis* and dinucleotide repeats were only observed in the leprosy bacillus. A trinucleotide repeat (TTC) displaying copy number differences has been described in some isolates of the leprosy bacillus, and this has 21 repetitions in the TN strain used for genome sequencing, while a hexanucleotide repeat in the *sigA* (rpoT) gene has been found recently to be present in three copies in most strains of *M. leprae*, including TN, but in four copies in others. Promising discriminatory tests have been devised that target these polymorphisms, and this encourages us to examine di- and trinucleotide repeats for variability in different isolates of *M. leprae*.

**MYCOBACTERIAL INTERSPERSED REPEITIVE UNITS (MIRUS)**

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

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

**INSERTION SEQUENCES AND DUPLICATED GENES**

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

Two identical copies of 16 dispersed repeats of >700 bp were detected and examined (Table 3). Three of these (1329, 1261 and 1179 bp) probably correspond to extinct IS, although it is interesting to note that despite loss of function their sequences are perfectly conserved, whereas two others (1054 and 753 bp) appear to be counterparts of the REP13E12 repeats described in the tubercle bacilli. The remaining duplicated sequences correspond to genes or more rarely to pseudogenes (Table 3), and again it is unusual to find perfect conservation of the sequence in bacteria. This suggests that these duplication events may have occurred very recently or that sequence divergence occurs at an exceptionally slow rate in *M. leprae*. 
Table 2. Features of MIRU in *M. leprae* and comparison with *M. tuberculosis*

| Class | Position | Bases | Genes* | *M. tuberculosis* | Class (No.)* | Comments |
|-------|----------|-------|--------|-------------------|--------------|----------|
| MIRU2 153610...153666 | 57 | *rbfE*...ML0112 | *rbfE*...3781 | – | – |
| MIRU2 534710...534766 | 57 | *scoA*...*scoB* | *scoA*...*scoB* | – | – |
| MIRU1 648342...648425 | 84 | ML0534...carA | RV1382...carA | 3 (2) | Degenerate |
| MIRU1 862062...862114 | >53 | ML0719...*aldB* | RV3292...*aldB* | – | – |
| MIRU3 877171...877223 | 53 | *purK*...*purE* | *purK*...*purE* | 3 | Embedded in *purK* |
| MIRU2 1164265...1164320 | 57 | *miaA*...*dapF* | *miaA*...*dapF* | – | – |
| MIRU1 1194835...1194886 | 52 | *dut*...ML1029 | *dut*...RV2696c | 1 | Out-of-frame |
| MIRU2 1207116...1207172 | 57 | ML1042...ML1041 | RV2680...*echA15* | 2 (3) | – |
| MIRU2 1327764...1327823 | 60 | ML1135...ML1136 | RV1300...RV1301 | 2 (3) | – |
| MIRU1 1368453...1368519 | 67 | *cysM*...ML1171 | *cysM*...RV1337 | – | In-frame with ML1171 |
| MIRU1 1371749...1371830 | 83 | *rphA*...ML1175 | *rphA*...RV1341 | 2 | Out-of-frame |
| MIRU2 1629692...1629746 | 55 | ML1368...ML1369 | RV1709...RV1710 | 2 | Out-of-frame |
| MIRU2 177816...177872 | 57 | ML1476...ML1475 | RV2454c...RV2455c | – | – |
| MIRU1 1692100...1692180 | 81 | argD...argF | argD...argF | – | – |
| MIRU2 1915421...1915483 | 66 | *pyrH*...frr | *pyrH*...frr | 2 (2) | No ATG |
| MIRU2 2044283...2044339 | 57 | *ilvB*...*ilvN* | *ilvB*...*ilvN* | – | No ATG |
| MIRU1 2613356...2613460 | 105 | ML2199...ML2200 | RV0813c...RV0814c | – | – |
| MIRU1 2617793...2618783 | 81 | *pabC*...ML2203 | *pabC*...RV0811c | – | – |
| MIRU2 2884042...2884095 | 54 | ML2412...ML2413 | RV0525...RV0526 | 2 | No ATG |
| MIRU1 2917923...2918003 | 81 | ML2442...ML2443 | RV0486...RV0487 | – | – |

* Underlining indicates pseudogenes.

* –, denotes MIRU absent.
Table 3. Identical duplicated genes and sequences of >700 bp.

| Sequence | Genes* | Description |
|----------|--------|-------------|
| 1329 bp  | ML0040 | Possible transposase remnant |
| 1261 bp  | ML1749 | Possible transposase remnant |
| 1179 bp  | ML0444 | Pseudogene similar to group II intron maturase |
| 1054 bp  | ML1290/ML1850 | Pseudogenes orthologous to REP13EI2 proteins |
| 753 bp   | ML1118/ML2286 | Pseudogenes orthologous to REP13EI2 proteins |
| 1551 bp  | ansP1/ansP2 | L-asparagine transport proteins |
| 1391 bp  | ML2356/ML2357 | Part of polyketide synthase |
| 1219 bp  | enoyl_2-oxalo | ENO family transport proteins |
| 1186 bp  | MLE0580/ML2602 | Methyl-acceptor-3-phosphate synthase |
| 1063 bp  | ML0125/ML0128 | Putative glycosyl transferase |
| 879 bp   | ML0105/ML1180 | QLS5 family |
| 879 bp   | ML0105/ML1181 | ESAT-6 family |
| 871 bp   | ML0105/ML1943 | Pseudogene orthologous to Rv3714c |
| 847 bp   | fadD5/fadD8 | Acyl-CoA synthase pseudogenes |
| 740 bp   | MLE2155/MLE2159 | Similar to region of cytochrome P450s |
| 704 bp   | umpA2/umpA1j | Mycolic acid synthase and pseudogene |

* Underlining indicates pseudogenes.

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

RLEP

The RLEP element was initially detected as a repetitive sequence in *M. leprae* by means of Southern blotting and subsequently characterized at the molecular level by Woods.

![Figure 1. Organization of repeated loci encoding PE, PPE, and ESAT-6 proteins. Gene names are given and repeat sizes indicated in bp.](image)
et al. These authors estimated that there were at least 28 copies of RLEP in the genome and demonstrated that there was a central portion, common to all copies of RLEP, flanked by additional sequences whose presence was variable. With the complete genome sequence at our disposal we were able to perform the definitive bioinformatic analysis and this revealed that the TN strain of \textit{M. leprae} contains 37 copies of RLEP, one of which, RLEP\_29, lacks part of the central domain and will not be discussed further here. There is a conserved segment of 488 bp found in all intact copies of RLEP and this is flanked by additional sequences present in two or more independent RLEP elements. Consequently, the total length can vary from 601 to 1075 bp (Figure 2) and, as described previously, no open reading frames capable of coding for transposases, resolvases or other IS-associated functions could be found. Further comparisons uncovered six polymorphic sites in the 488 bp conserved segment, three of which occurred only once while the remainder were found in numerous copies of RLEP. All of these polymorphisms can be accounted for by C-T transitions. On construction of a tree of RLEP sequences (Figure 2) by phylogenetic analysis using parsimony routines (PAUP), three large branches were established together with several outliers. However, there was little clear association between the length of the RLEP element or the presence of particular polymorphic nucleotides and its position in the tree (Figure 2). This is consistent with the complex organization of these sequences.

Roughly 1\% of the chromosome is composed of RLEP DNA and these elements are distributed fairly randomly. It is clear, however, that RLEP has contributed extensively to the remodelling of the \textit{M. leprae} genome as copies are often found at breaks in synteny with \textit{M. tuberculosis}. This will be discussed further below. There is a marked over-representation of RLEP elements, in either orientation, at the 3'-ends of genes, since \textasciitilde 30 of the copies are within 80bp of the stop codon of the nearest gene. In several instances, RLEP is situated within the coding sequence at the 3'-end or overlapping the stop codon. Examples of this may be found in the \textit{truA}, \textit{truB} and \textit{polA} genes. In two cases, important genes such as \textit{glnA} and \textit{polA} are flanked by inverted pairs of RLEPs in a configuration resembling that of a transposon. Data have been published that show that this composite \textit{polA} structure is polymorphic between isolates of \textit{M. leprae}\textsuperscript{28} and it is conceivable that similar variability may also be associated with \textit{glnA}. Transcription of \textit{polA} may also have been impaired by RLEP,\textsuperscript{29} as RLEP\_22 is situated 6bp upstream of the \textit{polA} initiation codon. Some copies of RLEP are found within the sequences of pseudogenes that have intact functional orthologues in \textit{M. tuberculosis}, notable examples are RLEP\_8 in the spermidine biosynthetic gene, \textit{speE}, RLEP\_29 in the phosphoglucomutase gene, \textit{pgmA}, and RLEP\_28 in ML1722, a pseudogene orthologous to Rv3037c, a conserved hypothetical gene of \textit{M. tuberculosis}. These observations suggest that RLEP may have been capable of transposition at one time although it is quite unclear how this was mediated. Furthermore, attempts to detect restriction fragment length polymorphisms linked to RLEP have revealed no diversity suggesting that RLEP is no longer capable of movement.\textsuperscript{30}

\begin{center}
\textbf{REPLEP}
\end{center}

There are 13 essentially intact sequences belonging to the REPLEP family and two large fragments (Figure 3). The largest elements are 881 bp long, with extensive complementarity between bases 1–95 and 783–880 (68\% identity). REPLEP is bounded in most cases by an 8bp inverted repeat (5'-GTTTGAGG) and contains no open reading frames. In several cases the inverted repeats continue past this octamer with certain REPLEP elements...
Figure 2. Phylogenetic tree of RLEP elements established using the phylogenetic analysis using parsimony routine of the GCG package. RLEP identifiers, sequence lengths and the sequence present at the concatenated polymorphic sites are indicated on the right. The sequence gaa signifies the presence of G, A and A at positions 501, 583 and 592 in the multiple alignment.

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

There are five intact sequences belonging to the LEPREP family and three fragments (Figure 4). These display near identical sequences with only three base differences being detected, two C-T transitions in REPLEP5 and one in REPLEP3. Unlike RLEP, REPLEP and LEPRPT, LEPREP displays a number of features commonly associated with IS elements and most probably corresponds to a degenerate version. The complete LEPREP sequence is 2383 bp long, contains a 54 bp palindromic inverted repeat and has a 6 bp inverted repeat (5'-CTAGTG) at its ends. Although there are no open reading frames that could code for
Figure 5. Repetitive elements and genome discontinuities. The three main repetitive elements in the *M. leprae* genome are shown together with examples of flanking genes and their counterparts in *M. tuberculosis*. The *M. tuberculosis* genes are designated with Rv prefixes, and the *M. leprae* genes with ML prefixes. Note the breaks in continuity of number of *M. tuberculosis* genes that indicate translocation event.

A. **REPLEP**, 12 complete copies of ~875 bp, plus 2 fragments;

B. **RLEP**, 36 complete copies of 545–700 bp plus 1 fragment;

C. **LEPREP**, 6 complete copies of 2,400 bp plus 3 fragments.

D. Example of gene loss by deletion following transposition of a **REPLEP** element and homologous recombination between the two copies. The gene organization in *M. tuberculosis* H37Rv is shown at the top followed by the corresponding region in an ancestor of *M. leprae*. The present situation in the TN strain of *M. leprae* is shown at the bottom.
Riportato DNA in M. leprae

Functional proteins, BLASTX searches revealed extensive sequence similarity to parts of transposases from Pseudomonas putida (EMBL:AI245436) and Agrobacterium tumefaciens (EMBL:AJ218270), and to putative group II intron maturase-related proteins such as that of the fungus, Cryptococcus parasiticus (EMBL:AI218567). Copies 4, 5 and 8 of LEPREP have been truncated and in two cases this appears to have resulted from the insertion of another IS element of 1261 bp that is now degenerate but still shows extensive similarity to IS1549 from Mycobacterium smegmatis. Copies 4 and 5 are truncated at their 3' and 5' ends, respectively, share a 14 bp residual sequence and are followed by the IS1549-like element (Figure 4). They may once have comprised part of the same LEPREP element. Four copies of LEPREP are followed by the same 21 (or 17) bp sequence whereas two copies are preceded by common sequences of 195, 146 and 33 bp, respectively. It is conceivable that these sequences represent preferential sites of insertion for LEPREP.

LEPRPT

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

REMODELLING THE GENOME

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

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

Acknowledgements

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

1. Malhi JS, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998; 62: 725–774.
2. Cole ST, Brosch R, Parkhill J et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature, 1998; 393: 537–544.
3. Gordon SV, Huyyn M, Parkhill J et al. New insertion sequences and a novel repeated sequence in the genome of Mycobacterium tuberculosis H37Rv: Microbiology, 1999; 145: 881–892.
4. Brosch R, Gordon SV, Eiglmeier K et al. Genomics, biology, and evolution of the Mycobacterium tuberculosis complex. In Hoffnung GF, Jacobs WR Jr (eds) Molecular genetics of mycobacteria. ASM Press, Washington D.C., 2000; pp. 19–36.
5. Brosch R, Philippe W, Stavropoulos E et al. Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra. Infect Immun, 1999; 67: 5768–5774.
6. Fang Z, Ding C, Kenna DT et al. IS6110-mediated deletions of wild-type chromosomes of Mycobacterium tuberculosis. J Bacteriol, 1999; 181: 1014–1020.
7. Tekaia F, Gordon SV, Garnier T. Analysis of the proteome of Mycobacterium tuberculosis in situ. Tubercle Lung Dis, 1999; 79: 329–342.
8. Lupsa JR, Roth JR, Weinstock GM. Chromosomal duplications in bacteria, fruit flies, and humans. Am J Hum Genet, 1996; 68: 21–27.
9. Brosch R, Gordon SV, Buchrieser C et al. Comparative genomics uncovers tandem chromosomal duplications in some strains of Mycobacterium bovis BCG: implications for vaccination. Comp Funct Genom (Oxf), 2000; 17: 111–123.
10. Supply F, Magdalena J, Himpero S, Locati C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol, 1997; 26: 991–1003.
11. Mazars E, Lejean S, Barus AL et al. High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci USA, 2001; 98: 1901–1906.
12. Supply F, Mazars E, Lejean S et al. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol Microbiol, 2000; 36: 762–771.
13. Cole ST, Eiglmeier K, Parkhill J et al. Massive gene decay in the leprosy bacillus. Nature, 2001; 409: 1007–1011.
14. Altschul SF, Boguski MS, Gish W, Wooton J. A basic local alignment search tool. J Mol Biol, 1997; 300: 110–130.
15. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. J Mol Biol, 1990; 215: 403–410.
16. Rutherford K, Parkhill J, Crook J et al. Artemis: sequence visualization and annotation. Bioinformatics, 2000; in press.
17. Reinsch G. Tandem repeats finder. a program to analyze DNA sequences. Nucleic Acids Res, 1999; 27: 573–580.
18. Jones LM, Cole ST, Klopper L. Leprosy: a Mycobacterium leprae genome browser. Lepr Rev, 2001; 72: 470–477.
19. Woods SA, Cole ST. A rapid method for the detection of potentially viable Mycobacterium leprae in human biopsies: a novel application of PCR. FEMS Microb Lett, 1999; 68: 305–310.
20. Shin YC, Lee H, Walsh GP et al. Variable numbers of TTC repeats in Mycobacterium leprae DNA from leprosy patients and use in strain differentiation. J Clin Microbiol, 2000; 38: 4535–4538.
21. Matsuoka M, Kai M, Maeda S et al. Massive gene decay in the leprosy bacillus. Nature, 2001; 409: 1007–1011.
22. Shin YC, Lee H, Walsh GP et al. Variable numbers of TTC repeats in Mycobacterium leprae DNA from leprosy patients and use in strain differentiation. J Clin Microbiol, 2000; 38: 4535–4538.
23. Borst P, Moszer I. Artemis: sequence visualization and annotation. Bioinformatics, 2000; in press.
24. Altschul SF, Boguski MS, Gish W, Wooton J. A basic local alignment search tool. J Mol Biol, 1997; 300: 110–130.
25. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. J Mol Biol, 1990; 215: 403–410.
26. Rutherford K, Parkhill J, Crook J et al. Artemis: sequence visualization and annotation. Bioinformatics, 2000; in press.
27. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. J Mol Biol, 1990; 215: 403–410.
28. Rutherford K, Parkhill J, Crook J et al. Artemis: sequence visualization and annotation. Bioinformatics, 2000; in press.
Fahi H, Cole ST. The Mycobacterium leprae genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel polA locus associated with genomic variability. Mol Microbiol, 1995; 16: 909–919.

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

Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of Mycobacterium leprae exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. Mol Microbiol, 1999; 4: 1653–1659.