© Indian Academy of Sciences

REVIEW ARTICLE

The genetics of *Mycobacterium tuberculosis*

S. VIJAYA

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India

Abstract

Gene manipulation in *Mycobacterium tuberculosis* has been slow in coming of age owing to the inherent difficulties associated with working on this aerosol-transmitted pathogen, in addition to the paucity of molecular tools such as plasmids and transposons. One of the early approaches to overcome these difficulties was the development of phasmids, which combined the properties of phages and plasmids and allowed introduction of recombinant genes into mycobacteria. The lone plasmid pAL5000 of mycobacteria has been exploited to its fullest potential in the construction of a plethora of vectors. Above all, the single most important achievement has been the development of elegant and innovative approaches to overcome the problem of illegitimate recombination which threatened the success of allelic-exchange mutagenesis in the slow-growing pathogenic mycobacterial species. In this review I discuss the current status of conditionally replicating plasmid and transposon vectors and their application in generating targeted mutations in mycobacteria.

[Vijaya S. 1998 The genetics of *Mycobacterium tuberculosis*. *J. Genet.* 77, 123–128]

The genus *Mycobacterium* includes a number of clinically important pathogens such as *M. tuberculosis*, *M. leprae*, *M. bovis* and *M. avium*. Tuberculosis has been and continues to be a major public health problem of the world, a situation that has been further aggravated by the advent of HIV-AIDS. Tuberculosis has thus played the role of a great leveller, including among its victims citizens of developed and developing countries alike. This has in turn spurred a flurry of research activity that endeavours to understand the basis of the success of this organism in surviving in the human host, despite four decades of use of antitubercular drugs. In India tuberculosis has always been a major killer, making this country one of the natural testing grounds for the highly promising *M. bovis*-BCG, the only vaccine available for TB even today. The near-complete failure of the BCG vaccination programme in India (Bloom and Fine 1994), combined with the emergence of drug resistance in *M. tuberculosis* in recent times, has led to the sobering realization that newer approaches for vaccination and identification of drug targets needed for developing the next generation of chemotherapeutic agents are an obligation one cannot escape from.

Molecular-genetic manipulation of microorganisms remains one of the most powerful approaches for understanding the basic biology of the microbe. A genetic system that permits the transfer, mutation and expression of specific genes is a prerequisite for a detailed understanding of the molecular basis of survival, virulence and pathogenicity of a bacterium. However, application of these techniques to *M. tuberculosis* was retarded owing to several unique features of this organism. Primarily, the pathogenic nature of the organism and its easy transmission by aerosols, with the consequent need for expensive containment facilities, discouraged all but the most intrepid from pursuing research into its biochemistry and molecular biology. The tough lipid and carbohydrate-rich cell wall of *M. tuberculosis* was a further impediment as it made disruption of cells for obtaining macromolecules extremely inefficient by most commonly employed methods. These factors resulted in delay in development of vectors and methods for gene transfer into mycobacteria; when they did finally arrive, the organism was found to prefer illegitimate recombination, and the frequency of homologous recombination was disappointingly low. Despite these drawbacks, research on *M. tuberculosis* flourished in a few select pockets, such as the University of Wisconsin, led by Donald Smith; University of Colorado, led by Alfred Crowle; University College, London, led by Graham Rook; Colorado State University, led by Patrick Brennan; Case Western Reserve University, led by Thomas Daniel and Jerold Ellner; Albert Einstein College of

Keywords. *Mycobacterium tuberculosis*; molecular genetics; transposons; temperate mycobacteriophages; phasmids; gene disruption.

Medicine, led by Barry Bloom; and the Indian Institute of Science, led by T. Ramakrishnan.

History of mycobacterial genetics

The early attempts to develop genetic systems for mycobacteria focussed on tools parallel to those that had revolutionized the study of other, more amenable bacteria such as *E. coli*. Foremost among these were the use of temperate bacteriophages that integrate into the host genome and can be coaxed to serve the interests of the geneticist by transducing segments of DNA, as well as the phenomenon of conjugative transfer of genetic material with attendant recombination. The fast-growing nonpathogen *M. smegmatis* was the natural candidate for standardizing these protocols, which could then be applied to the more dangerous slow-growing pathogenic members of the genus. In a systematic effort, several mycobacteriophages of *M. smegmatis* were isolated from soil by the Bangalore team headed by T. Ramakrishnan. It is gratifying to note that the first ever successful demonstration of transduction in mycobacteria was from the Indian Institute of Science (Sundar Raj and Ramakrishnan 1970). In this study, one of the mycobacteriophages, I3, which formed turbid plaques on *M. smegmatis* strain SN2, was used to demonstrate reversion of a variety of auxotrophs to prototrophy following infection with I3 propagated on wild-type SN2. The failure of deoxyribonuclease treatment to bring about a reduction in number of revertants ruled out transformation as the cause of the observed genetic alterations. This report was followed by another, three years later, from Japanese workers, of a similar reversion of auxotrophic markers when different pairs of *M. smegmatis* strains were mixed and plated (Tokunaga *et al.* 1973). No sex factors that could bring about high-frequency transfer were, however, detected. It was suggested that recombination between the two alleles brought together by conjugation could explain the observations. The scientists were, however, unable to build a linkage map owing to 'ordering ambiguities'.

Mycobacterial plasmid vectors

The advent of molecular genetics, with the attendant tools such as plasmid vectors, gave impetus to the search for mycobacterial plasmids that could replicate in mycobacteria and could be used to transform them. This search has till date yielded limited success, with pAL5000 (Labidi *et al.* 1985) being the lone plasmid that has been manipulated for use in fast-growing and slow-growing mycobacteria. Also, the dearth of selectable markers for mycobacteria was a major limitation in exploitation of this plasmid, although it was completely sequenced (Rauzier *et al.* 1988) and subjected to gene disruption and deletion analysis (Ranes *et al.* 1990). A truncated version of pAL5000 was cloned into pUC19 containing Tn*903*, which encodes resistance to kanamycin (Ranes *et al.* 1990), to yield the mini *Mycobacterium*–*E. coli* shuttle vector pRR3, which has origins of replication for both systems and can transform mycobacteria to kanamycin resistance. Indeed, resistance to kanamycin has proved to be the most reliable selection marker for mycobacteria. In another attempt to develop a plasmid vector for mycobacteria, pAL5000 was cloned into the *E. coli* plasmid pIJ666 to generate shuttle plasmids which, upon transformation of *M. smegmatis* or *M. bovis* BCG with it, conferred resistance to kanamycin (Snapper *et al.* 1988). In a somewhat different approach, an integrating plasmid with only the *E. coli* origin of replication was successfully deployed in *M. smegmatis* (Husson *et al.* 1990), where it was used to express the *M. leprae hsp65* gene. An interesting feature of this vector was its use of a nutritional marker, *pyrF*, which allowed both positive and negative selection. An intact *pyrF* gene, which encodes orotidine monophosphate decarboxylase, made mycobacteria and *E. coli* susceptible to 5-fluoroorotic acid while its absence resulted in an auxotroph requiring uracil supplementation in the growth medium. The advent of electroporation as a method to introduce foreign DNA into all types of cells revolutionized transformation of mycobacteria, where getting DNA across the tough lipid-rich cell walls of these organisms had been a major bottleneck.

The advent of phasmids

As an alternative to plasmids therefore, mycobacteriophages, of which quite a few had been identified, were explored as the basis of developing vectors for introducing genes into mycobacteria. A pathbreaking report describing the success of such an approach appeared in 1987 (Jacobs *et al.* 1987). Two temperate mycobacteriophages, D29 and TM4, were used for the purpose, both of which have a genome size of approximately 50 kb. The phage DNA was subjected to partial *Sau*3A digestion, and fragments of 30 kb to 50 kb were purified and ligated to the 6.4-kb *Bam*HI-digested cosmid vector pHC79. The property of the cosmid component of the recombinant was used to package the recombinant phasmid into lambda phage heads and propagate them in *E. coli*. It was hoped that those recombinants where the pHC79 had got inserted into a nonessential region of D29 or TM4 would still remain viable phages. This was tested by introducing DNA obtained from the *E. coli* transformants into *M. smegmatis* by electroporation. Packaged phages were indeed obtained from the *M. smegmatis* and, in addition, these phages could be successfully propagated in other slow-growing mycobacteria such as BCG. The strategy used in this approach is outlined in figure 1. The ability of these recombinant mycobacteriophages to infect and replicate in BCG raised hopes of introducing foreign genes into these mycobacteriophage vectors and expressing them in BCG, which, as a vaccine strain, has been found to be safe

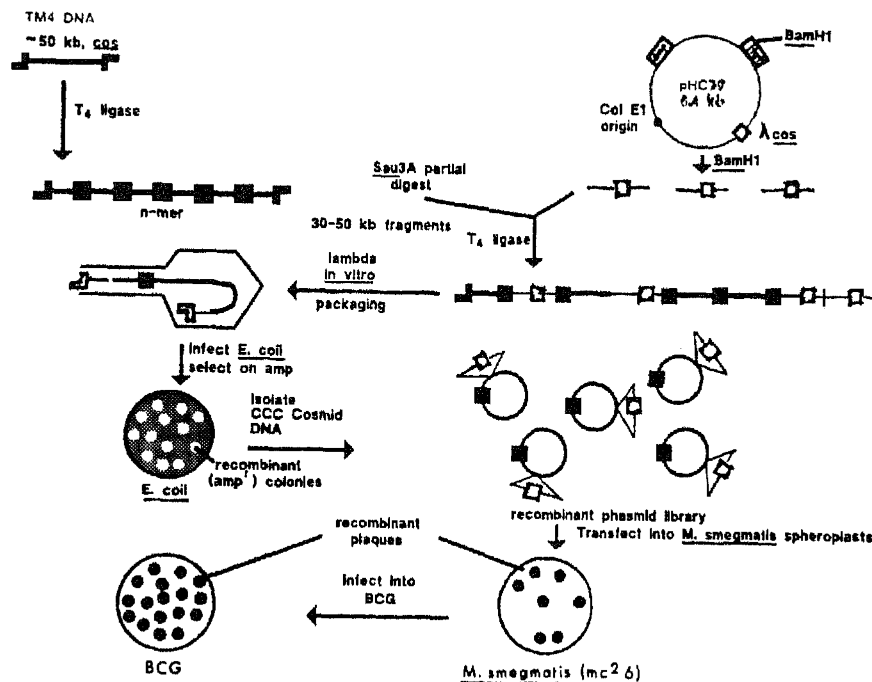

Figure 1. Diagram of the construction of the shuttle phasmids. TM4 phage DNA was ligated at a concentration of 250 μg ml^{-1} and then partially digested with *Sau*3A to obtain fragments that averaged 30–50 kb. These fragments were ligated at a 1 : 2 molar ratio of TM4 fragments to pHC79 that had been cleaved with *Bam*HI and alkaline phosphatased. The packaging of an aliquot of this ligation with *in vitro* packaging mix (Gigapack Plus, Stratagene) and subsequent transduction into ER1381 (*hsdR mcrA$^+$ mcrB$^+$*) yielded 10^6 ampicillin colonies per μg of TM4 DNA insert, when plated on L-agar containing ampicillin at 50 μg ml^{-1}. A pool of 40,000 ampicillin-resistant clones was prepared by homogenizing colonies in L-broth from the selection plates with a glass spreader. Plasmid DNA was isolated from the pool of clones by alkaline-SDS treatment followed by phenol–chloroform extraction and concentration with ethanol. Covalently closed circular plasmid DNA was transfected into mc^{2}6 spheroplasts. The plaques were screened for the presence of pHC79 by performing plaque lifts using Biotrans nylon membranes (ICN Radiochemicals). The membranes were hybridized with pHC79 DNA nick-translated with ^{32}P-dCTP, and autoradiography was carried out. [Reprinted with permission from *Nature* (Jacobs Jr W. R., Tuckman M. and Bloom B. R. 1987 Introduction of foreign DNA into mycobacteria using a shuttle plasmid. **327**, 532–534), copyright 1987, Macmillan Magazines Ltd]

for human use. However, the recombinant phasmids derived from D29 and TM4 only gave rise to lytic infection and did not lysogenize the mycobacteria. These workers therefore resorted to phasmids derived from the temperate mycobacteriophage L1 (Snapper *et al.* 1988), which they showed could stably lysogenize *M. smegmatis* to kanamycin resistance when the Tn*903*-derived *aph* gene was cloned into the phasmid. It was also claimed that these L1-derived phasmids could lysogenize BCG. Phasmids expressing the luciferase gene as reporter (Jacobs *et al.* 1993) and able to grow in *M. tuberculosis* have subsequently been successfully exploited for rapid diagnosis and drug susceptibility testing of *M. tuberculosis* clinical isolates.

Targeted gene disruption

The disruption of specific genes by targeted recombination with a cloned defective copy of the same gene is another powerful approach to understanding the function of each gene in an organism. With the gradual availability of plasmids and selection markers for mycobacteria, gene disruption was attempted, again first with the fast-growing nonpathogen *M. smegmatis*. In one of the early approaches (Husson *et al.* 1990), a shuttle vector that can replicate to high copy number in *E. coli* but has to integrate into the mycobacterial chromosome to survive was constructed, utilizing a nutritional selection based on the *M. smegmatis pyrF* gene into which the kanamycin resistance gene was inserted. This plasmid vector was shown to integrate into the genomic *pyrF* locus by homologous recombination, conferring resistance to 5-fluoroorotic acid and to kanamycin. Electroporation was used to introduce the DNA into cells, but efficiencies were distressingly low, as low as 10 to a few hundred transformants per microgram of DNA used.

Insertional mutagenesis of the slow-growing pathogen *M. tuberculosis* and of the BCG vaccine strain was of course the ultimate ambition of the mycobacterial geneticist. Such an effort was undertaken on a large scale in a subsequent study, which used a somewhat different approach (Kalpana *et al.* 1991). Here, a library of *M. smegmatis* genomic DNA was constructed in the *E. coli* vector pYUB36, and the *aph* gene conferring kanamycin resistance was moved from its genomic location in *E. coli ec^{2}270* into the recombinant

plasmids, which were then selected by their kanamycin hyperresistance phenotype. This library was then introduced into *M. smegmatis* by electroporation and kanamycin-resistant colonies were selected on rich medium. This was followed by a screen for auxotrophy. The auxotrophs of *M. smegmatis* thus obtained were then used for identification of the corresponding genes from *M. tuberculosis* and BCG by complementation. Thus the methionine auxotroph of *M. smegmatis* was used to obtain the homologous gene from BCG and *M. tuberculosis*. These genes cloned in *E. coli* vectors were in turn disrupted by introduction of the *aph* gene, and the plasmids linearized and introduced by electroporation into either BCG or *M. tuberculosis*. It was expected that homologous recombination resulting from a double crossover event of a linear DNA fragment containing the *aph*-inactivated methionine gene would yield Kan^r methionine auxotrophs of BCG or *M. tuberculosis*. However, this experiment threw up an unpleasant surprise; less than 1 in 200 of the Kan^r colonies turned out to be methionine auxotrophs! Analysis of genomic DNA from several Kan^r colonies confirmed that insertion of the disrupted copy of the methionine gene had occurred randomly by non-homologous recombination into the BCG chromosome. This propensity of the slow-growing mycobacteria for illegitimate recombination was confirmed by workers attempting to disrupt, by allelic exchange, the genes for urease (Reyrat *et al.* 1995) and mycocerosic acid synthase (Azad *et al.* 1996), where one of 38 and two of 50 selected colonies were allelic-exchange mutants, respectively, while the majority were products of illegitimate recombination.

This major setback to the progress of using genetic approaches and screens in the slow-growing pathogenic mycobacteria set off a fresh flurry of attempts to circumvent this problem of illegitimate recombination. One of the methods used to enhance the low rate of homologous recombination was the use of long linear recombination substrates obtained from cosmid clones with insert sizes about 40 kbp (Balasubramanian *et al.* 1996). The provision of long flanks drove up the proportion of homologous recombinants to a more respectable 6% for the *leuD* gene in two *M. tuberculosis* strains tested. Yet another more successful solution to this problem was the use of positive selection markers that would permit growth only of cells with double crossover events (Pelicic *et al.* 1996). In this latter method, the *sacB* gene from *B. subtilis*, which confers sensitivity to sucrose (Quandt and Hynes 1993), was used to eliminate all cells where single crossover had resulted in chromosomal integration of the entire plasmid along with the *sacB* gene following selection on kanamycin (figure 2). This two-step selection strategy gave rise to a high 26% of allelic-exchange mutants. Further honing of this approach involved use of a vector with a temperature-sensitive origin of replication

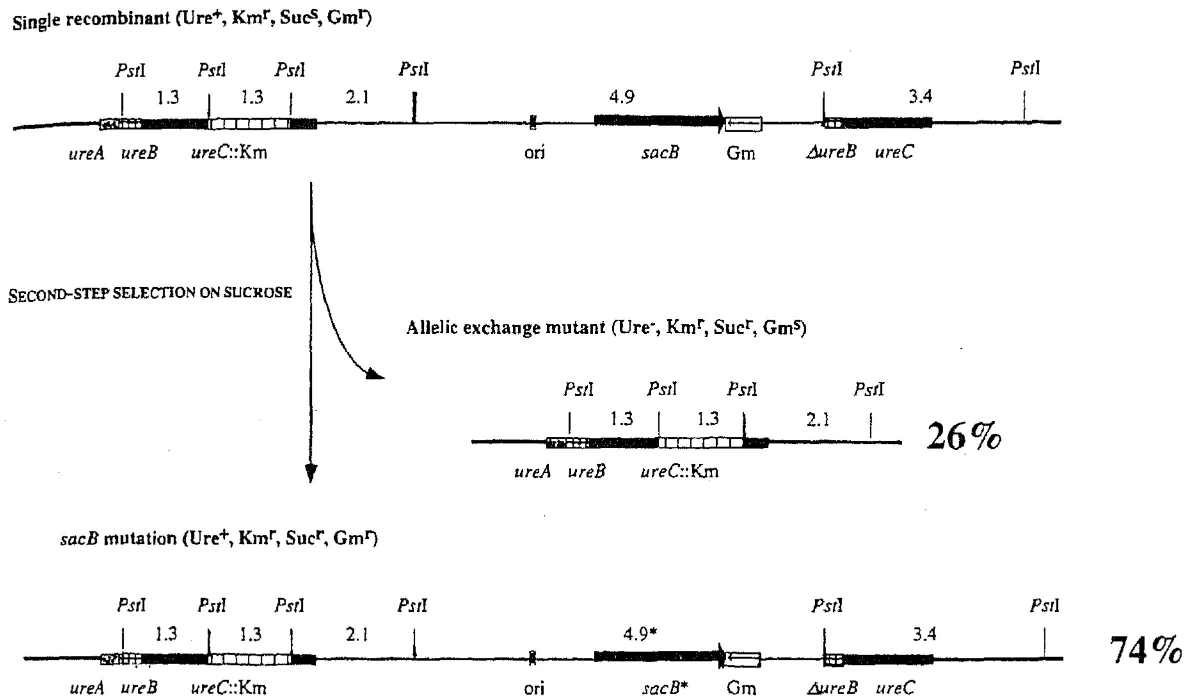

Figure 2. General strategy for allelic exchange mutagenesis using two-step selection on 2% sucrose. In the first step, a single recombinant is selected on kanamycin or gentamicin. In the second step, clones that have lost the *sacB* gene are positively selected on 2% sucrose. *sacB** indicates a mutant allele of the *sacB* gene. The results presented are for the *ureC* gene and the lengths of the hybridizing fragments are indicated in kbp. *ureA*, *ureB* and *ureC* are the three subunits of the mycobacterial urease. Km, Tn*903* gene encoding kanamycin resistance; Gm, *aacC1* gene encoding gentamicin resistance; ori, *E. coli* origin of replication. [Reprinted from *FEMS Microbiology Letters* (Pelicic V., Reyrat J.-M. and Gicquel B. 1996 Positive selection of allelic exchange mutants in *Mycobacterium bovis* BCG. **144**, 161–166), copyright Elsevier Science]

derived from a mutated pAL5000 (Pelicic *et al.* 1997) along with the kanamycin resistance marker and the *sacB* gene for positive selection of the desired double-recombination events. The initial selection was performed at 32°C in presence of kanamycin, following which the colonies obtained were subjected to 39°C in presence of sucrose. The latter step eliminated cells with free plasmids as well as those that had undergone a single recombination event resulting in the continued presence of the *sacB* gene. All (100%) of the colonies obtained using this strategy were demonstrated by Southern analysis of genomic DNA to be allelic-exchange mutants when the *purC* gene was disrupted. This heralded the dawn of a new era in mycobacterial genetics.

The transposon as a genetic tool

Transposons are yet another versatile genetic tool and mycobacterial geneticists have spared no efforts in identifying, characterizing and developing them for mutagenesis in mycobacteria. Two transposons of mycobacteria, Tn*611* (Martin *et al.* 1990) and IS*1096* (Cirillo *et al.* 1991), were identified and put to use in generating random mutations. However, the transposition frequencies were extremely low, of the order of one to 20 transposition events per microgram of DNA. This, and the low electroporation efficiencies obtained in mycobacteria, have resulted in as few as a hundred mutants per experiment (McAdam *et al.* 1995). As in all other approaches, the fast-growing *M. smegmatis* served as the testing ground for transposon mutagenesis in mycobacteria. The Tn*611* transposon carrying the kanamycin resistance marker was inserted into a temperature-sensitive *E. coli–Mycobacterium* shuttle vector (Guilhot *et al.* 1994). After introduction into *M. smegmatis* by electroporation, transformants were selected initially on kanamycin at 30°C, followed by incubation at the elevated temperature, when the transposon vector is lost, permitting growth only of those cells where the selection marker had transposed onto the chromosome. Altogether, 26,700 randomly selected thermoresistant clones included 78 auxotrophs, which were demonstrated to represent random transposition events. No transposition hotspots were detected in this study. The insertion sequence IS*1096* was also used to generate auxotrophic mutants of BCG (McAdam *et al.* 1995), using kanamycin selection on a nonreplicating plasmid. This study yielded very few transformants, including leucine and methionine auxotrophs of BCG. Clearly, replication of the inserted transposon-containing vector within mycobacteria for a period of time before the vector was forced out by selection enhanced the number of transposition mutants obtained. With the development of the *sacB*-containing positive selection vectors, which were also thermosensitive for maintenance and replication, the natural sequel was to introduce the transposon into these vectors and improve the transposition efficiencies. This was undertaken in *M. tuberculosis* (Pelicic *et al.* 1997), and $> 10^6$ transposition mutants were obtained. Southern analysis as well as cloning and sequence determination of several insertion sites confirmed the randomness of the transposition. Mention has to be made here that the vectors, developed at the Pasteur Institute in Paris, are freely available to investigators on request. A thermosensitive transposition vector was also adopted for use with the phasmids, generating conditionally replicating mycobacteriophages (Bardarov *et al.* 1997) based on D29 and TM4, which have been used to generate transposon-mutated libraries of *M. phlei*, BCG and *M. tuberculosis*. These vectors have now made it possible to undertake genetic analysis of individual mycobacterial genes, whose compilation from the genome sequence (Cole *et al.* 1998) has thrown open this organism for a plethora of studies hitherto not feasible.

References

Azad A. K., Sirakova T. D., Rogers L. M. and Kalattukudy P. E. 1996 Targeted replacement of the mycocerosic acid synthase gene in *Mycobacterium bovis* BCG produces a mutant that lacks mycosides. *Proc. Natl. Acad. Sci. USA* **93**, 4787–4792.

Balasubramanian V., Pavelka Jr M. S., Bardarov S., Martin J., Weisbrod T. R., McAdam R. A., Bloom B. R. and Jacobs Jr W. R. 1996 Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. *J. Bacteriol.* **178**, 273–279.

Bardarov S., Kriakov J., Carriere C., Yu. S., Vaamonde C., McAdam R. A., Bloom B. R., Hatfull G. F. and Jacobs Jr W. R. 1997 Conditionally replicating mycobacteriophages: A system for transposon delivery to *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **94**, 10961–10966.

Bloom B. R. and Fine P. E. M. 1994 The BCG experience: implications for future vaccines against tuberculosis. In *Tuberculosis: pathogenesis, protection and control* (ed. B. R. Bloom), pp. 531–557. American Society for Microbiology, Washington, DC.

Cirillo J. D., Barletta R. G., Bloom B. R. and Jacobs Jr W. R. 1991 A novel transposon trap for mycobacteria: isolation and characterization of IS1096. *J. Bacteriol.* **173**, 7772–7780.

Cole S. T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D. *et al.* 1998 Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

Guilhot C., Otal I., van Rompaey I., Martin C. and Gicquel B. 1994 Efficient transposition in mycobacteria: construction of *Mycobacterium smegmatis* insertional mutant libraries. *J. Bacteriol.* **176**, 535–539.

Husson R. N., James B. E. and Young R. A. 1990 Gene replacement and expression of foreign DNA in mycobacteria. *J. Bacteriol.* **172**, 519–524.

Jacobs Jr W. R., Tuckman M. and Bloom B. R. 1987 Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature* **327**, 532–534.

Jacobs Jr W. R., Barletta R., Udani R. A., Chan J., Kalkut G., Sarkis G., Hatfull G. F. and Bloom B. R. 1993 Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. *Science* **260**, 819–822.

Kalpana G. V., Bloom B. R. and Jacobs Jr W. R. 1991 Insertional mutagenesis and illegitimate recombination in mycobacteria. *Proc. Natl. Acad. Sci. USA* **88**, 5433–5437.

Labidi A., David H. L. and Roulland-Dussoix D. 1985 Restriction endonuclease mapping and cloning of *Mycobacterium fortuitum*

var. *fortuitum* plasmid pAL5000. *Ann. Microbiol. (Paris)* **136**, 209–215.

McAdam R. A., Weisbrod T. R., Martin J., Scuderi J. D., Brown A. M., Cirillo J. D., Bloom B. R. and Jacobs Jr W. R. 1995 *In vivo* growth characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. *Infect. Immun.* **63**, 1004–1012.

Martin C., Timm J., Rauzier J., Gomez-Luz R., Davies J. and Gicquel B. 1990 Transposition of an antibiotic resistance element in mycobacteria. *Nature* **345**, 739–743.

Pelicic V., Reyrat J.-M. and Gicquel B. 1996 Positive selection of allelic exchange mutants in *Mycobacterium bovis* BCG. *FEMS Microbiol. Lett.* **144**, 161–166.

Pelicic V., Jackson M., Reyrat J.-M., Jacobs Jr W. R., Gicquel B. and Guilhot C. 1997 Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **94**, 10955–10960.

Quandt J. and Hynes M. F. 1993 Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* **127**, 15–21.

Ranes M. G., Rauzier J., Lagranderie M., Gheorghiu M. and Gicquel B. 1990 Functional analysis of pAL5000, a plasmid from *Mycobacterium fortuitum*: construction of a "mini" *Mycobacterium–Escherichia coli* shuttle vector. *J. Bacteriol.* **172**, 2793–2797.

Rauzier J., Moniz-Pereira J. and Gicquel-Sanzey B. 1988 Complete nucleotide sequence of pAL5000, a plasmid from *Mycobacterium fortuitum*. *Gene* **71**, 315–321.

Reyrat J.-M., Berthet F.-X. and Giquel B. 1995 The urease locus of *Mycobacterium tuberculosis* and its utilization for the demonstration of allelic exchange in *Mycobacterium bovis* bacillus Calmette-Guerin. *Proc. Natl. Acad. Sci. USA* **92**, 8768–8772.

Snapper S. B., Lugosi L., Jekkel A., Melton R. E., Kieser T., Bloom B. R. and Jacobs Jr W. R. 1988 Lysogeny and transformation in mycobacteria: Stable expression of foreign genes. *Proc. Natl. Acad. Sci. USA* **85**, 6987–6991.

Sundar Raj C.V. and Ramakrishnan T. 1970 Transduction in *Mycobacterium smegmatis*. *Nature* **228**, 280–281.

Tokunaga T., Mizuguchi Y. and Suga K. 1973 Genetic recombination in mycobacteria. *J. Bacteriol.* **113**, 1104–1111.