A low G+C content genetic island in *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* with homologous genes in *Mycobacterium tuberculosis*

Mark Tizard,[†] Tim Bull, Douglas Millar,[‡] Tim Doran,[†] Helene Martin, Nazira Sumar, Jon Ford and John Hermon-Taylor

Author for correspondence: John Hermon-Taylor. Tel: +44 181 767 7631. Fax: +44 181 725 3594. e-mail: jhermon@sghms.ac.uk

Department of Surgery, St George's Hospital Medical School, Cranmer Terrace, Tooting, London SW17 0RE, UK

The technique of representation difference analysis PCR has been applied to find genes specific to *Mycobacterium avium* subsp. *paratuberculosis*. This generated a 671 bp fragment which was used to isolate a larger genetic element found in the enteric pathogens *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* but which was absent from the very closely related and relatively benign *M. avium* subsp. *avium*. This element, designated GS, is greater than 6.5 kbp in length and has a G+C content 9 mol% lower than other genes from this species. There is a previously uncharacterized insertion sequence associated with one end. The GS element encodes five ORFs in *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum*, all of which have counterparts encoded in *Mycobacterium tuberculosis*. Database searches revealed homologues for these ORFs in a number of bacterial species, predominantly Gram-negative organisms, including a number of enteric pathogens. These homologous genes encode functions related to LPS or extracellular polysaccharide biosynthesis. This element has a number of features in common with pathogenicity islands such as its low G+C content, an association with a putative insertion sequence and a grouping of genes of related function with a possible link to virulence. No direct link to pathogenicity has been shown but GS may belong to a group of related 'genetic islands' and represents the first such element to be identified in mycobacteria.

Keywords: mycobacteria, *Mycobacterium avium* subsp. *paratuberculosis*, polysaccharide biosynthesis, genetic island, pathogenicity island

INTRODUCTION

The *Mycobacterium avium* complex consists of three subspecies, *M. avium* subsp. *paratuberculosis*, *M. avium*

subsp. *silvaticum* and *M. avium* subsp. *avium*. *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* are enteric pathogens in immunocompetent hosts. Particular serotypes of *M. avium* subsp. *avium* cause disseminated infections in immunocompromised hosts such as patients with AIDS (Jacobson *et al.*, 1991), whilst various other *M. avium* subsp. *avium* serotypes have been associated with avian tuberculosis (Grange *et al.*, 1990). *M. avium* subsp. *paratuberculosis* can cause chronic enteritis in many animal species, including primates (Chiodini *et al.*, 1984; McClure *et al.*, 1987; Hines *et al.*, 1995), and *M. avium* subsp. *silvaticum* can cause chronic enteritis in calves (Matthews & Mc-

[†]**Present address:** CSIRO Division of Animal Health, Private Bag 24, Geelong, VIC 3216, Australia.

[‡]**Present address:** CSIRO Division of Biomolecular Engineering, PO Box 184, North Ryde, NSW 2113, Australia.

Abbreviation: RDA-PCR, representational difference analysis PCR.

The EMBL accession numbers for the sequence of the GS element are AJ223832 (*M. avium* subsp. *silvaticum*) and AJ223833 (*M. avium* subsp. *paratuberculosis*).

Diarmid, 1979). The condition in animals shows a broad range of histopathological characteristics (Buergelet *et al.*, 1978; Perez *et al.*, 1996; Clarke, 1997). In classical pluribacillary disease, acid-fast bacilli and macrophages are visible microscopically in the tissues but with little additional inflammatory cell infiltrate. In paucimicrobial disease, however, bacilli are rarely observed but there is a florid granulomatous inflammatory response (Carrigan & Seaman, 1990; Clarke & Little, 1996).

M. avium subsp. *paratuberculosis* and the wood pigeon strain *M. avium* subsp. *silvaticum* are very closely related organisms (Collins *et al.*, 1983; Saxegaard & Baess, 1988; Hurley *et al.*, 1989; Thorel *et al.*, 1990). DNA analysis of part of the rRNA operon shows only very minor differences between these organisms and other *M. avium* complex strains (Böddinghaus *et al.*, 1990; van der Giessen *et al.*, 1994). Strain differences within *M. avium* subsp. *paratuberculosis* (Levy-Frebault *et al.*, 1989; Collins *et al.*, 1990), however, particularly those recently identified between bovine and ovine isolates (Pavlik *et al.*, 1995; Bauerfeind *et al.*, 1996), suggest that these pathogens are adapting to their niche in individual hosts. Studies on four isolates of *M. avium* subsp. *paratuberculosis* from humans with chronic enteritis of the Crohn's disease type, so far suggest a common clonal origin (Francois *et al.*, 1997).

The principal distinctions between the genomes of *M. avium* complex strains are the presence of transposable insertion elements including IS1245 (Guerrero *et al.*, 1995) found in all three *M. avium* complex subspecies, IS900 found exclusively in all *M. avium* subsp. *paratuberculosis* isolates, IS901/2 (Green *et al.*, 1989; Kunze *et al.*, 1991; Moss *et al.*, 1992b), IS1110 (Hernandez Perez *et al.*, 1994) and IS1311 (Roiz *et al.*, 1995). Phylogenetic relationships of these elements suggest these organisms have been recipients of exogenous DNA in recent evolutionary times whilst some strains of *M. avium* subsp. *avium* lack all of these elements. This investigation was designed to explore further specific genetic differences between these enteric pathogens and *M. avium* subsp. *avium*, using representational difference analysis (Lisitsyn *et al.*, 1993).

METHODS

Cell and DNA extraction. *M. avium* subsp. *paratuberculosis* human Crohn's disease strains 0138 and 0139, bovine isolates 0022, 0025, 0204, 0206, 0207, 0209, 0210, 0211 and 0212, red deer isolate 0021 and isolate 0208 (ATCC 19698) were grown in Dubos broth (Difco) supplemented with 20% newborn calf serum (Gibco) and 2 mg mycobactin J l⁻¹ (Rhône Mérieux). *M. avium* subsp. *silvaticum* strains 0010, 0012 and 0013 (from wood pigeons) were grown in Dubos supplemented with 10% DMA (150 mM NaCl, 5% BSA, 7.5% dextrose) and 2 mg mycobactin J l⁻¹. *M. avium* subsp. *avium* strain 0018 (gift from F. Portaels, Institute of Tropical Medicine, Antwerp, Belgium; an AIDS isolate not related to *M. avium* subsp. *avium* strain 18; Chiodini, 1993) and *M. avium* subsp. *avium* strains 0033, 0034, 0037, 0038 and 0040 (blood culture isolates from AIDS patients) were grown in Dubos with 10% DMA

but without mycobactin J. Cells were harvested by centrifugation and DNA was extracted using the method described by Doran *et al.* (1994). Briefly, cells were incubated at 37 °C in 0.1% SDS, 1 mg proteinase K ml⁻¹, TNE buffer (50 mM Tris/HCl, pH 8.0; 150 mM NaCl; 10 mM EDTA) for 18 h followed by phenol/chloroform/isoamyl alcohol extraction. DNA for use as PCR template was diluted in TE_{0.1} buffer (10 mM Tris/HCl, pH 8.0; 0.1 mM EDTA).

Representational difference analysis PCR (RDA-PCR). The RDA-PCR technique (Lisitsyn *et al.*, 1993; Lisitsyn & Wigler, 1995) was adapted to subtract the genome of *M. avium* subsp. *avium* strain 0018 from that of *M. avium* subsp. *paratuberculosis* strain 0138. The method involved three rounds of amplification and subtractive hybridization. *M. avium* subsp. *paratuberculosis* strain 0138 DNA (1 µg) was digested with *Bam*HI (Boehringer), using manufacturer's recommended conditions, and ligated with the oligonucleotides *RBam*12 (5' GATCCTCGGTGA 3') and *RBam*24 (5' AGCACTCTCC-AGCCTCTCACCGAG 3'). The ligate (80 ng) was amplified with the *RBam*24 primer in a 50 µl reaction with *Taq* DNA polymerase (Promega) using manufacturer's recommended conditions, 4 mM MgCl₂ and 30 cycles of 93 °C for 1 min, 58 °C for 1 min, 72 °C for 3 min. This product represented the tester amplicon. This procedure was repeated using 1 µg *M. avium* subsp. *avium* strain 0018 DNA and the product generated represented the driver amplicon. Both the tester and driver amplicons were digested with *Bam*HI to remove the terminal adapter oligonucleotide sequences. The tester amplicon was ligated with *JBam*24 (5' ACCGACGTCGAC-TATCCATGAACG 3') and *JBam*12 (5' GATCCGTTTCATG 3'); 0.4 µg of this ligate was hybridized with 20 µg driver amplicon at 67 °C for 20 h in 1 M NaCl, 1 mM EDTA, 10 mM EPPS following denaturation at 94 °C for 5 min. The resulting mixture was digested with mung bean nuclease for 30 min and the nuclease was heat-inactivated. Then, 2 µg of this sample was amplified with the *JBam*24 primer using 20 cycles of the conditions described above. This *JBam* tester amplicon was digested with *Bam*HI and ligated with *NBam*24 (5' AGGCA-ACTGTGCTATCCGAGGGAG 3') and *NBam*12 (5' GATC-CCCTCG 3'); 0.4 µg of this ligate was hybridized with 20 µg driver amplicon as described above. Following nuclease treatment, 2 µg of this mixture was amplified for 15 cycles using the *NBam*24 primer and conditions described above.

Sequencing of the RDA fragments and PCR of bacterial DNAs.

The PCR product bands identified in the *NBam* amplicon were cloned into the pGEM-T vector (Promega). Three clones were generated and designated *Bam* RDA-PCR Product (BRP) 1–3. Sequence data were determined from the vector primers T7 and SP6 using fluorescent-dye-labelled nucleotide and automated sequence analysis on a PE ABI 373A instrument (performed by AltaBioscience, Birmingham University, UK). Sequence data were used to select primers to perform PCR specific for each fragment. DNA extracted from a number of mycobacteria, and bacteria of other genera, was amplified using the following primers and conditions. PCR for BRP1 was achieved with primers TD1 (5' GATGCCGTGAGGAG-GTAAAGCTGC 3') and TD2 (5' GATACGGCTCTTGA-ATCCTGCACG 3') using 1.5 mM MgCl₂ and 40 cycles of 93 °C for 1 min, 62 °C for 1 min, 72 °C for 1 min. PCR for BRP2 was performed with primers TD3 (5' CTCCCATTT-GCGCAGATACCGGAC 3') and TD4 (5' TACCGAGGC-CGCCGATTATCTGTG 3') using 2.0 mM MgCl₂ and 40 cycles of 93 °C for 1 min, 62 °C for 1 min, 72 °C for 1 min. PCR for BRP3 was performed with primers TD5 (5' TCAC-CGTGATGGTCGACTTGGTCC 3') and TD6 (5' ATGT-TCTTGGACAGCCGGTTGGTG 3') using 2.0 mM MgCl₂.

Fig. 1. RDA-PCR products generated using *M. avium* subsp. *paratuberculosis* strain 0138 DNA as tester and *M. avium* subsp. *avium* strain 0018 DNA as driver. Panels JBam, RBam and NBam represent the products of the three sequential subtraction steps. The three PCR products in the NBam panel are, from top to bottom, BRP1 (671 bp), BRP2 (573 bp) and BRP3 (514 bp). Lane L contains the size markers.

and 40 cycles of 93 °C for 1 min, 62 °C for 1 min, 72 °C for 1 min. In addition, reactions using primers specific for IS900 (Sanderson *et al.*, 1992) and IS902 (Moss *et al.*, 1992a) were performed. Genomic DNAs were diluted to a final concentration of 100 pg μl^{-1} and 5 μl was used as template for PCR. For each of the PCR reactions a negative control was used, consisting of a buffer blank that was processed in tandem with samples during DNA extraction.

Cloning the GS element. Genomic libraries were prepared from *M. avium* subsp. *paratuberculosis* 0138 and *M. avium* subsp. *silvaticum* 0012 (wood pigeon isolate) genomic DNA digested with *EcoRI* or *HindIII* and ligated with pre-cut and phosphatase-treated pUC18 (Pharmacia). These were plated, lifted on Hybond-N+ (Amersham) and probed with the BRP1 PCR product labelled with [^{32}P]dCTP (ICN; Amersham Redivue kit). Filters were hybridized in 0.1% (w/v) dried skimmed milk, 50% (v/v) formamide, 0.1% SDS, $6 \times \text{SSC}$ (890 mM NaCl, 90 mM trisodium citrate, pH 7.0) at 37 °C for 16 h. After screening with the BRP1 probe, the filters were re-probed with IS900 (for the *M. avium* subsp. *paratuberculosis* library) or IS902 (for the *M. avium* subsp. *silvaticum* library), to check the representation of the DNA within the libraries. Thirty-two thousand recombinants were screened for each of the *M. avium* subsp. *paratuberculosis* libraries (both *EcoRI* and *HindIII*) and 16000 recombinants were screened for the *M. avium* subsp. *silvaticum* library (*EcoRI*). A single clone hybridizing to the BRP1 probe was isolated from the *M. avium* subsp. *silvaticum* library, designated pGS16. As clones could not be obtained from the *M. avium* subsp. *paratuberculosis* libraries, sequence data were determined by PCR/cloning and sequencing from genomic DNA (see below).

Mapping and sequencing the GS element within pGS16. Plasmid DNA was prepared from the clone pGS16 and digested with a number of restriction enzymes including *HindIII*, *BamHI*, *XhoI*, *PstI*, *Sall* and *EcoRI* (Boehringer). These digests were analysed by agarose gel electrophoresis and Southern blotting. The blots were probed first with the radiolabelled BRP1 fragment. Following washing and exposure to Kodak X-Omat film, the blots were stripped (using 0.2 M NaOH, 0.1% SDS) and re-probed with various radiolabelled restriction enzyme fragments to produce an over-

lapping map of restriction sites. A separate blot was probed with radiolabelled DNA from *M. avium* subsp. *avium* strain 0018 to identify non-hybridizing fragments. Based on these mapping data, four overlapping subclones (pSGH27, pSGH33, pSGB21 and pSGX65) were prepared and plasmid DNA was isolated using QIA-miniprep (Qiagen). This DNA was sequenced fully on both strands using automated sequencing (as described above) by primer walking. An additional 1 kb of sequence 5' of the *EcoRI* junction in clone pGS16 was determined from *M. avium* subsp. *silvaticum* strain 0012 genomic DNA by use of inverse PCR following *BamHI* digestion and re-ligation, shown in Fig. 4(b). Sequence data for the GS element in *M. avium* subsp. *paratuberculosis* 0138 (Linda strain) were determined by PCR amplification of DNA segments from genomic DNA using the primers selected from the sequence determined from *M. avium* subsp. *silvaticum* strain 0012. PCR products were cloned into the pGEM.T vector (Promega) and two clones were sequenced for each PCR product (no base discrepancies were observed between clones). The distribution of the GS element in other species was investigated by Southern blotting. DNA from a range of mycobacterial species, *M. avium* subsp. *avium* serotypes, *Escherichia coli*, *Staphylococcus aureus* and *Pseudomonas aeruginosa* was prepared as previously described and digested with *PvuII*. This was blotted by standard protocols and probed at high stringency (0.1% SDS, $0.1 \times \text{SSC}$ at 35 °C) using a radiolabelled 282 bp GS region derived by PCR from *M. avium* subsp. *silvaticum* strain 0012 using the primers P/TGSBA1, 5' GCGGATCGATCACCTCTACG 3', and P/TGSBA2R, 5' GCCGAACATCTCCGACGAGG 3'.

Analysis and interpretation of sequence data. General editing and analysis of sequence data were performed using Gene-Runner v2.1 (Hastings Software). A final contig was compiled and assembled using GelAssemble in the UWGCG package courtesy of the MRC/Wellcome Trust HGMP facility (Hinxton, UK; www.hgmp.mrc.ac.uk). The predicted ORFs were searched against the GenBank/EMBL databases (release 53–12/97) using the BLAST and BLIXEM programs. G+C compositions for mycobacterial species were calculated from database entries and accessed from the Codon Usage Tabulated from GenBank (CUTG) website at the DNA Information and Stock Centre (www.dna.affrc.go.jp:10080).

RESULTS AND DISCUSSION

RDA-PCR isolation of a specific DNA fragment

The result of applying RDA-PCR using *M. avium* subsp. *paratuberculosis* as the tester and *M. avium* subsp. *avium* as the driver was a progression of simplifying PCR banding patterns which by the third round (NBam

amplicon) showed only three remaining products (Fig. 1). The results shown were representative of three independent experiments which yielded the same three bands: *Bam* RDA product 1 (BRP1) of 671 bp, BRP2 of 573 bp and BRP3 of 514 bp. Each product was cloned and shown by Southern blotting and probing to not cross-hybridize to the others. Nucleotide sequence data

Table 1. Distribution of BRP sequences

Distribution of the BRP DNAs. PCR analysis was performed on genomic DNA from a number of mycobacteria and other species to detect each of the BRP DNAs and IS900 and IS902. DNA was analysed from 11 isolates of *M. avium* subsp. *paratuberculosis*, three isolates of *M. avium* subsp. *silvaticum*, five isolates of *M. avium* subsp. *avium*, nine other mycobacteria and five other species. + + +, Strong clear PCR product; + +, weak PCR product; +, faint product or signal seen when probed with radiolabelled product; —, no signal even with radiolabelled probing. ND, Not done.

Species	BRP1	BRP2	BRP3	IS900	IS902
<i>M. avium</i> subsp. <i>paratuberculosis</i>					
0138	+++	+++	+++	+++	—
0025	+++	+++	+++	+++	—
0021	+++	+++	+++	+++	—
0022	+++	+++	+++	+++	—
0139	+++	+++	+++	+++	—
0209	+++	+++	+++	+++	—
0208	+++	+++	+++	+++	—
0211	+++	+++	+++	+++	—
0210	+++	+++	+++	+++	—
0212	+++	+++	+++	+++	—
0207	+++	+++	+++	+++	—
0204	+++	+++	+++	+++	—
0206	+++	+++	+++	+++	—
<i>M. avium</i> subsp. <i>silvaticum</i>					
0010	+++	—	+++	—	+++
0012	+++	—	+++	—	+++
0013	+++	—	+++	—	+++
<i>M. avium</i> subsp. <i>avium</i>					
0018	—	—	—	—	—
0033	+++	—	+++	+	—
0034	—	+++	+++	—	—
0037	—	+++	+++	—	—
0038	—	+++	+++	—	—
0040	—	+++	+++	—	—
<i>M. malmoeense</i>	—	++	++	ND	ND
<i>M. szulgai</i>	—	++	++	ND	ND
<i>M. gordonae</i>	—	++	++	ND	ND
<i>M. chelonae</i>	—	++	++	ND	ND
<i>M. fortuitum</i>	—	—	—	ND	ND
<i>M. bovis</i>	—	++	++	ND	ND
<i>M. tuberculosis</i>	—	++	—	ND	ND
<i>M. leprae</i>	—	—	++	ND	ND
<i>M. phlei</i>	—	—	++	ND	ND
<i>E. coli</i>	—	+	+	ND	ND
<i>Staphylococcus aureus</i>	—	+	+	ND	ND
<i>Nocardia</i> sp.	—	+	+	ND	ND
<i>Streptococcus</i> sp.	—	+	+	ND	ND
<i>Shigella</i> sp.	—	+	—	ND	ND
<i>Pseudomonas</i> sp.	—	+	—	ND	ND

Fig. 2. Southern blot of total DNA from several mycobacterial species and other bacteria cut with *PvuII* and hybridized at high stringency with a 282 bp probe derived from *gsbA* from within the GS element. MAP, *M. avium* subsp. *paratuberculosis*; MAA, *M. avium* subsp. *avium*; MAS, *M. avium* subsp. *silvaticum*.

were determined and the sequence was used to select PCR primers specific for each product (generating PCR products corresponding to 481 bp for BRP1, 407 bp for BRP2 and 315 bp for BRP3). These PCR reactions, together with PCR specific for IS900 and specific for IS902, with appropriate controls were applied to DNA from a number of bacteria (Table 1). This showed that application of the RDA-PCR technique had produced a DNA fragment, BRP1, present in *M. avium* subsp. *paratuberculosis* but not detected by high annealing temperature PCR in most *M. avium* subsp. *avium* isolates. This fragment was also present in isolates of *M. avium* subsp. *silvaticum*. The BRP3 fragment was present in *M. avium* subsp. *paratuberculosis*, but not detected by high annealing temperature PCR in *M. avium* subsp. *avium* strain 0018, and present in all other *M. avium* subsp. *avium* strains tested, but not detected by high annealing temperature PCR from the *M. avium* subsp. *silvaticum* strains. The power of the original RDA-PCR technique to identify subtle differences between highly homologous DNAs was demonstrated by the distribution of the BRP3 fragment, which was present in all subspecies and strains of the *M. avium* complex tested in this study with the sole exception of strain 0018 – the driver strain. To confirm the distribution of the GS element in mycobacteria, a 283 bp region within the gene *gsbA* derived from *M. avium* subsp. *silvaticum* was used as a probe in a standard Southern hybridization, at high stringency, using total DNA from a range of bacterial species digested with *PvuII*, including several serotypes of the *M. avium* complex and other closely related species of mycobacteria (Fig. 2). This indicated a single band that correlated in size to that predicted by the sequence data,

hybridizing only to *M. avium* subsp. *silvaticum* and *M. avium* subsp. *paratuberculosis* isolates.

PCR specific for BRP1 gave a signal with all *M. avium* subsp. *paratuberculosis* and all *M. avium* subsp. *silvaticum* strains and one *M. avium* subsp. *avium* strain (0033) but no signal was seen with any other bacterial DNA as template (Table 1). *M. avium* subsp. *avium* strain 0033 also gave a weak signal with IS900 PCR. Southern blotting of genomic DNA from this strain showed a single hybridizing band with an IS900 probe, not the multibanding pattern that would normally be expected with an IS of this family (Green *et al.*, 1989; Moss *et al.*, 1992b). This unusual strain of *M. avium* subsp. *avium* is under further investigation. In contrast, BRP2 and BRP3 PCR reactions gave signals with a variety of mycobacteria and other bacteria, both Gram-positive and Gram-negative. Products generated from other bacteria and mycobacteria were fainter, suggesting weaker hybridization of primers with these DNA templates (data not shown). However, the fact that any amplification was detected suggests quite a high degree of conservation of these sequences between genera, though no significant nucleotide sequence homologies were found when searching the databases with the BRP2 and BRP3 sequences. Since BRP2 and BRP3 were found across a broad range of bacteria they were not examined further.

Cloning and mapping the GS element in *M. avium* subsp. *silvaticum*

BRP1 was used to probe two genomic libraries from *M. avium* subsp. *paratuberculosis*. No clones were isolated from either an *EcoRI*- or a *HindIII*-derived library

Fig. 3. Analysis of pGS16 plasmid DNA. (a) Restriction digests of plasmid DNA: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sal*I. (b) Southern blot of (a) probed with radiolabelled *M. avium* subsp. *avium* strain 0018 genomic DNA after 3 h exposure. The same pattern was observed after 7 d exposure. Lane L contains the size markers.

despite multiple hybridizations when probed with IS900, indicating good representation in these libraries. Probing an *Eco*RI library from *M. avium* subsp. *silvaticum* yielded a single clone, pGS16. A restriction enzyme map was generated for this plasmid and insert using *Eco*RI, *Bam*HI, *Hind*III, *Pst*I and *Sal*I coupled with Southern blotting and probing, initially with radiolabelled BRP1

Fig. 4. (a) Map of the plasmid pGS16, showing the two copies of pUC18 (black shading) and the two independent genomic *Eco*RI fragments from *M. avium* subsp. *silvaticum* of 3 kbp and 17 kbp, respectively. (b) Restriction map of the 17 kbp *Eco*RI fragment (E 17) showing the location of the original BRP1 probe (black box), the unique region GS, the *Bam*HI inverse PCR fragment (BIP) and the position of the four fragments subcloned for sequence determination (black bars): *Hind*III 2.7 kbp (H 2.7), *Hind*III 3.3 kbp (H 3.3), *Bam*HI 2.1 kbp (B 2.1) and *Xho*I 6.5 kbp (X 6.5).

followed by probing with other isolated fragments (Fig. 3). This showed that pGS16 had an unusual structure, containing two genomic inserts, one of approximately 17 kbp and the other 3 kbp, and two copies of pUC18 (Fig. 4a). The 3.3 kbp and 2.7 kbp *Hind*III fragments from this clone did not hybridize to *M. avium* subsp. *avium* strain 0018 DNA (Fig. 3b), which suggests the region designated GS is not present in this subspecies. Four subclones were generated, pGSX65, pGSH33, pGSH27 and pGSB21 (Fig. 4b), and used as template to determine the nucleotide sequence for this region. Subsequent analysis of sequence data confirmed the assignment of restriction fragments, interpretation of which may have been confused by the presence of two copies of pUC18. The 2.1 kbp *Bam*HI fragment and the 6.9 kbp *Hind*III fragment (Fig. 3b) both showed stronger hybridization to the *M. avium* subsp. *avium* strain 0018 genomic DNA than other fragments flanking but not including GS. This is suggestive of the presence of a repetitive DNA element within these clones which is common to these two strains.

G+C content of the GS element and *M. tuberculosis* homologues

Analysis of the nucleotide sequence of the GS region in *M. avium* subsp. *silvaticum* and *M. avium* subsp. *paratuberculosis* (encompassing 8938 bp in *M. avium*

Fig. 5. Organization of the ORFs within the region sequenced showing the G + C composition of the region in *M. avium* subsp. *silvaticum* (a) and *M. avium* subsp. *paratuberculosis* (b). ORFs *gsa*–*gsd* span the unique region; the *ist* ORFs show sequence homology to insertion sequences. ORF6 was not found in *M. avium* subsp. *paratuberculosis*.

Fig. 6. GS ORF homologues in *M. tuberculosis* are found in two clusters. (a) Homologues of *gsa*–*gsc* are found in the region from 31972 to 35493 bp on cosmid Y277 (accession no. Z77724). (b) Homologues of *gsd* and *gsc* are found in the region 14306–15934 bp on cosmid Y349 (accession no. Z83018); this *gsc* homologue has a G + C composition of only 53.7 mol%.

subsp. *silvaticum*) showed five ORFs in a ‘unique’ region (Fig. 5), designated *gsa* (277 aa residues), *gsbA* (343 aa residues), *gsbB* (334 aa residues), *gsc* (240 aa residues) and *gsd* (266 aa residues). All five were found to have homologues in *M. tuberculosis* (Fig. 6). These were found in two groups: (a) ORFs homologous to *gsa*, *gsbA*, *gsbB* and *gsc* grouped together on the 42 kb *M. tuberculosis* genome fragment contained in cosmid Y277 (GenBank accession no. Z79701) in the region 31972 to 35493 bp; (b) ORFs homologous to *gsd* and *gsc* were adjacent on the 35 kb fragment of the

M. tuberculosis genome in cosmid Y349 (GenBank accession no. Z83018) in the region 14306 to 15934 bp. A further ORF designated ORF6 of 441 aa residues and not found in *M. avium* subsp. *paratuberculosis* was found to be homologous to one in *M. tuberculosis* on cosmid Y63 (GenBank accession no. Z96800) in the region 28870 to 30201 bp. It is interesting to note that despite the presence of these homologues no PCR signal was seen in *M. tuberculosis* when probing with BRP1, which spans the coding sequences for *gsbB* and *gsc*, which may have been due to the lower level of homology in the *gsc* region (Table 1). These *M. tuberculosis* genes have been sequenced but are as yet uncharacterized either translationally or functionally. The homologies with *gsbA* and *gsbB* were particularly high at 89% and 84.6%, respectively.

The overall G + C content of the GS region was 58.0 mol%, significantly lower than the mean G + C values for *M. avium* subsp. *paratuberculosis* (67.0 mol%). This suggests that *gsa*–*gsd* genes are a relatively recent acquisition for both *M. avium* subsp. *silvaticum* and *M. avium* subsp. *paratuberculosis*. Four of the gene sequences published for *M. avium* subsp. *paratuberculosis* had values notably below the mean. Two were the 16S and 23S rRNA sequences (56.8 and 56.1 mol%; accession nos M61680 and X74495). One was a partial sequence described as homologous to a bacteriophage integrase (62.0 mol%; GenBank accession no. L39071). The last was an uncharacterized sequence of 620 bp simply identified as ‘specific to *M. avium* subsp. *paratuberculosis*’ (fragment F57, 59 mol%; GenBank accession no. X70277). By comparison the mean G + C content for 48 coding sequences published for *M. avium* subsp. *avium* was 68.5 mol%. There was a further significant drop, to 50.5 mol%, in the G + C content of the GS element which occurs within the region encoding *gsc* (Fig. 5).

The GS element in *M. avium* subsp. *silvaticum* versus *M. avium* subsp. *paratuberculosis*

The nucleotide sequence of GS was determined for *M. avium* subsp. *silvaticum* from a genomic clone and for *M. avium* subsp. *paratuberculosis* from duplicate cloned PCR fragments with overlapping sections. These were 99.3% identical over a contiguous span of 4435 bp (Fig. 5). Within the contiguous span, 30 bp changes were found which caused 4 aa residue changes in *gsbA*, 1 aa residue change in *gsbB*, 2 aa residue changes in *gsc* and no changes in the translation of *gsd*. ORF6 was not identified in *M. avium* subsp. *paratuberculosis* (hence its different designation). The corresponding sequence in *M. avium* subsp. *paratuberculosis* showed a G + C content more consistent with the normal values for mycobacteria. This suggests that one end of the GS element may be situated at this divergence point (Fig. 5b).

In the subclone pGSB21, the G + C content increased again to a more characteristic level for mycobacteria (65.4 mol%). The two ORFs found in this region were designated *istA* (329 aa residues) and *istB* (264 aa

Table 2. BLAST search homologies for GS element ORFs

Summary of BLAST search result for GS ORFs against the translated EMBL nucleic acid database, release 53. The top four or five search results are shown for each ORF, including the species and where known the gene name and function. Identity is shown over a span of amino acids. EMBL accession numbers are indicated.

Species	Gene name and function	Accession no.	Span (aa)	Identity (%)
gsa				
<i>Mycobacterium tuberculosis</i>	Unknown	Z79701	187	63.6
<i>Helicobacter pylori</i>	Unknown	AE000532	143	38.5
<i>Mycobacterium tuberculosis</i>	Unknown	Z83018	152	32.9
<i>Escherichia coli</i>	<i>wcaE</i> ; colanic acid biosynthesis gene	AE000295	70	52.8
<i>Rhizobium meliloti</i>	<i>nodC</i> ; polysaccharide-polymerizing gene	U53327	76	50.0
gsbA				
<i>Mycobacterium tuberculosis</i>	Unknown	Z79701	338	89.0
<i>Yersinia enterocolitica</i>	<i>gmd</i> ; GDPmannose 4,6-dehydratase	U46859	311	58.5
<i>Vibrio cholerae</i>	<i>gmd/rfbD</i> ; oxidoreductase	Y07786	293	59.3
<i>Escherichia coli</i>	<i>gmd</i> ; GDPmannose 4,6-dehydratase	U38473	316	57.2
<i>Pseudomonas aeruginosa</i>	<i>gmd</i> ; GDPmannose 4,6-dehydratase	U18320	301	54.8
gsbB				
<i>Mycobacterium tuberculosis</i>	Unknown	Z79701	318	84.6
<i>Escherichia coli</i>	<i>wcaG</i> ; colanic acid biosynthesis	D90843	301	55.5
<i>Mycobacterium tuberculosis</i>	Unknown	U42580	307	47.2
<i>Yersinia enterocolitica</i>	<i>wbcJ</i> ; LPS biosynthesis	U46859	318	49.4
<i>Helicobacter pylori</i>	<i>nolK</i> ; nodulation protein	AE000526	261	39.8
gsc				
<i>Mycobacterium tuberculosis</i>	Unknown	Z79701	216	70.8
<i>Mycobacterium tuberculosis</i>	Unknown	Z83018	216	71.3
<i>Synechocystis</i> sp.	Unknown	D90913	131	41.9
<i>Vibrio cholerae</i>	<i>rfbT</i> , Ogawa serotype conversion (methylase?)	D11123	51	37.3
gsd				
<i>Mycobacterium tuberculosis</i>	Unknown	Z83018	216	70.3
<i>Helicobacter pylori</i>	Unknown	AE000532	159	31.4
<i>Haemophilus influenzae</i>	<i>lgtD</i> ; glycosyltransferase	U32768	113	30.9
<i>Lactococcus lactis</i> subsp. <i>cremoris</i>	<i>epsG</i>	U93364	89	40.4
<i>Shigella dysenteriae</i>	<i>rfpA</i>	S73325	89	32.5
ORF6				
<i>Xanthomonas campestris</i>	<i>ugd</i> ; UDPglucose dehydrogenase	X79772	201	53.2
<i>Bacillus subtilis</i>	<i>ugd</i> ; UDPglucose dehydrogenase	Z99122	208	54.8
<i>Pseudomonas aeruginosa</i>	<i>ugd</i> ; UDPglucose dehydrogenase	Z92952	199	54.3
<i>Mycobacterium tuberculosis</i>	<i>ugd</i> ; UDPglucose dehydrogenase homologue	Z96800	200	47.0

residues). Database searches revealed significant homology for the two ORFs, in particular *istB*, with the putative transposase of IS1415 from *Rhodococcus erythropolis* (38.8% over 219 aa residues) and a number of other IS sequences belonging to the IS21 family (Nagy *et al.*, 1997; Reimann *et al.*, 1989). Within this region, PCR performed on *M. avium* subsp. *paratuberculosis* using a number of primers derived from *M. avium* subsp. *silvaticum* sequence failed to yield products. This suggested that this sequence was significantly different or absent. It may have been a more recent acquisition or the copy of the IS in *M. avium* subsp. *paratuberculosis* may have excised. Further studies will be required to clarify this point.

Homologues of the GS genes in other bacteria

A number of genes encoded in the GS element show significant homologies to known genes (Table 2), the most striking being *gsbA* and ORF6. Homology to *gsbA* was found across a wide range of bacterial species with genes encoding GDP-D-mannose dehydratase (*gmd*; a gene involved in fucose biosynthesis). Examples of these included *rfbD* from *Vibrio cholerae* (59% identity over 293 aa residues), *gmd* from the *E. coli* *wca* gene cluster (57% over 316 aa residues), the *rfb*-associated orf13.7 from *Yersinia enterocolitica* (58.5% over 311 aa residues) and *algD* from *P. aeruginosa* (55% over 301 aa residues). In all these cases the next gene downstream,

usually by only 3–10 bp, was a homologue of *gsbB*. Whilst homologues to other genes in the GS cluster may have different arrangements, orientations or genomic locations, these two genes were always found in tandem. *gsa* and *gsd* had homologies with the glycosyltransferases that were associated with genes including *gmd* in *Escherichia*, *Yersinia* and *Pseudomonas*. These have been suggested to control the transfer and polymerization of the modified monosaccharide units generated by the homologues of *gsbA* and *gsbB* during assembly of LPS, colanic acid or extracellular polysaccharide (Bik *et al.*, 1996; Zhang *et al.*, 1996; Stevenson *et al.*, 1996; Currie *et al.*, 1995). None of these clusters showed homologues of *gsc*. Only *rfbT* of *V. cholerae* O1 showed significant homology to this gene. *rfbT* is a gene which has been shown to mediate the switch from the Ogawa to the Inaba epidemic serotype which encoded a cytoplasmic membrane protein thought to methylate sugar residues (Strocher *et al.*, 1992). The product of ORF6 showed homology with UDPglucose dehydrogenase from *Xanthomonas campestris* (53.2% over 201 aa residues), *Bacillus subtilis* (54.8% over 208 aa residues) and *P. aeruginosa* (54.3% over 199 aa residues).

Genetic islands and cell wall structures

The low G + C content of this cluster of genes indicates that they may have been acquired from a foreign source. The G + C content and association with a possible mobilization factor (in this case a putative IS sequence) are features of a class of genetic element, characterized by pathogenicity islands (Blum *et al.*, 1994). Examples of virulence islands, a subset of pathogenicity islands, have been found in both Gram-positive and Gram-negative bacteria ranging in size from 6 kbp in *Streptococcus pyogenes* to 190 kbp in uropathogenic *E. coli* (Hacker *et al.*, 1997). There is at present no experimental evidence to assign the GS element as a pathogenicity island; however, it has been shown to be present in two clearly characterized enteric pathogens and was absent from the majority of *M. avium* subsp. *avium* strains studied.

Interpretation of database homologies indicated that the GS element encodes functions related to extracellular polysaccharide synthesis or modification. The study of cell wall assembly is of current interest in the development of novel antibiotics to combat emergent multidrug-resistant *M. tuberculosis* (Kolattukudy *et al.*, 1997; Telenti *et al.*, 1997). Mycobacterial cell surface polysaccharides are known to have potent immunological effects and an implied role in pathogenicity (Brennan & Nikaido, 1995; Venisse *et al.*, 1995). There is the potential for genes involved in synthesis or modification of such polysaccharide to affect the virulence of mycobacteria. Determination of a serotype can be controlled by specific gene clusters and is related to cell wall structures, although in *M. avium* subsp. *avium*, no connection between serovars and pathogenicity has been shown. The *ser-2* locus has been cloned and characterized in *M. avium* subsp. *avium* and shown to encode functions that produce a dimethyl fucopyranosyl

rhamnopyranoside unit at the terminus of the glycopeptidolipid which is characteristic of serovar 2 and not found in other serovars (Belisle *et al.*, 1993). This locus was shown to encode functions such as synthesis of fucose and the appropriate glycosyltransferases and methylases (Mills *et al.*, 1994). It is possible that GS is related to the *ser-2* locus functionally or genetically, though at present the nucleotide sequence for *ser-2* is unavailable for comparison.

The identification of the GS element represents the first observation in mycobacteria of a 'genetic island'-like structure. Pathogenicity islands have been shown to include structures encoding metabolic functions, transport functions and metal-ion or antibiotic resistance functions (Hacker *et al.*, 1997). In this case the functions relate to modification of the mycobacterial extracellular polysaccharide component, possibly glycopeptidolipid. Further work is required to clarify the boundaries of the element in *M. avium* subsp. *silvaticum* and *M. avium* subsp. *paratuberculosis*, its specific functions and prevalence in other pathogens.

NOTE ADDED IN PROOF

Homologues to GS genes in *M. tuberculosis* are now annotated as follows: *gsa* = Rv1514, *gsbA* = *gmdA*; Rv1511, *gsbB* = *epiA*; Rv1512, *gscTB1* = Rv1513, *gscTB2* = Rv2956, *gsd* = Rv2957 (Cole *et al.*, 1998).

ACKNOWLEDGEMENTS

This work was supported by the Wellcome Trust, with additional support from the Colt Foundation and Dinwoodie Trust, to whom we wish to express our appreciation. Analysis of this data would not have been possible without the generous cooperation of the MRC HGMP computing facility, Hinxton, UK. Thanks to Dr H. S. Griffiths for discussion and critical appraisal of this manuscript.

REFERENCES

- Bauerfeind, R., Benazzi, S., Weiss, R., Schliesser, T., Willems, H. & Baljer, G. (1996). Molecular characterization of *Mycobacterium paratuberculosis* isolates from sheep, goats, and cattle by hybridization with a DNA probe to insertion element IS900. *J Clin Microbiol* 34, 1617–1621.
- Belisle, J. T., Klaczekiewicz, K., Brennan, P. J., Jacobs, W. R. & Inamine, J. (1993). Rough morphological variants of *Mycobacterium avium*: characterization of genomic deletions resulting in the loss of glycopeptidolipid expression. *J Biol Chem* 268, 10517–10523.
- Bik, E. M., Bunschoten, A. E., Chang, A. C. Y. & Mooi, F. R. (1996). Genetic organization and functional analysis of the *otn* DNA essential for cell wall polysaccharide synthesis in *V. cholerae* O139. *Mol Microbiol* 20, 799–811.
- Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschape, H. & Hacker, J. (1994). Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* 62, 606–614.
- Böddinghaus, B., Rogall, T., Flohr, T., Blocker, H. & Bottger, E. C. (1990). Detection and identification of mycobacteria by amplification of rRNA. *J Clin Microbiol* 28, 1751–1759.

- Brennan, P. J. & Nikaido, H. (1995). The envelope of mycobacteria. *Annu Rev Biochem* 64, 29–63.
- Buergelt, C. D., Hall, C., McEntee, K. & Duncan, J. R. (1978). Pathological evaluation of paratuberculosis in naturally infected cattle. *Vet Pathol* 15, 196–207.
- Carrigan, M. J. & Seaman, J. T. (1990). The pathology of Johne's disease in sheep. *Aust Vet J* 67, 47–50.
- Chiodini, R. J. (1993). Abolish *Mycobacterium paratuberculosis* strain 18. *J Clin Microbiol* 31, 1956–1957.
- Chiodini, R. J., Van Kruiningen, H. J. & Merkal, R. S. (1984). Ruminant paratuberculosis (Johne's disease): the current status and future prospects. *Cornell Vet* 74, 218–262.
- Clarke, C. J. (1997). The pathology and pathogenesis of paratuberculosis in ruminants and other species. *J Comp Pathol* 116, 217–261.
- Clarke, C. J. & Little, D. (1996). The pathology of ovine paratuberculosis: gross and histological changes in the intestine and other tissues. *J Comp Pathol* 114, 419–437.
- Cole, S. T., Brosch, R., Parkhill, J. & 39 other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.
- Collins, D. M., Gabric, D. M. & de Lisle, G. W. (1990). Identification of two groups of *Mycobacterium paratuberculosis* strains by restriction endonuclease analysis and DNA hybridization. *J Clin Microbiol* 28, 1591–1596.
- Collins, P., Matthews, P. R., McDiarmid, A. & Brown, A. (1983). The pathogenicity of *Mycobacterium avium* and related mycobacteria for experimental animals. *J Med Microbiol* 16, 27–35.
- Currie, H. L., Lightfoot, J. & Lam, J. S. (1995). Prevalence of *gca*, a gene involved in synthesis of A-band common antigen polysaccharide in *Pseudomonas aeruginosa*. *Clin Diagn Lab Immunol* 2, 554–562.
- Doran, T. J., Davies, J. K., Radford, A. J. & Hodgson, A. L. (1994). Putative functional domain within ORF2 on the *Mycobacterium* insertion sequences IS900 and IS902. *Immunol Cell Biol* 72, 427–434.
- Francois, B., Krishnamoorthy, R. & Elion, J. (1997). Comparative study of *Mycobacterium paratuberculosis* strains isolated from Crohn's disease and Johne's disease using restriction fragment length polymorphism and arbitrarily primed polymerase chain reaction. *Epidemiol Infect* 118, 227–233.
- van der Giessen, J. W., Haring, R. M. & van der Zeijst, B. A. (1994). Comparison of the 23S ribosomal RNA genes and the spacer region between the 16S and 23S rRNA genes of the closely related *Mycobacterium avium* and *Mycobacterium paratuberculosis* and the fast-growing *Mycobacterium phlei*. *Microbiology* 140, 1103–1108.
- Grange, J. M., Yates, M. D. & Boughton, E. (1990). The avian tubercule and its relatives. *J Appl Bacteriol* 68, 411–431.
- Green, E. P., Tizard, M. L., Moss, M. T., Thompson, J., Winterbourne, D. J., McFadden, J. J. & Hermon-Taylor, J. (1989). Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. *Nucleic Acids Res* 17, 9063–9073.
- Guerrero, C., Bernasconi, C., Burki, D., Bodmer, T. & Telenti, A. (1995). A novel insertion element from *Mycobacterium avium*, IS1245, is a specific marker for analysis of strain relatedness. *J Clin Microbiol* 33, 304–307.
- Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23, 1089–1097.
- Hernandez Perez, A., Fomukong, N. G., Hellyer, T., Brown, I. N. & Dale, J. W. (1994). Characterization of IS1110, a highly mobile genetic element from *Mycobacterium avium*. *Mol Microbiol* 12, 717–724.
- Hines, M. E., Kreeger, J. M. & Heron, A. J. (1995). Mycobacterial infections of animals: pathology and pathogenesis. *Lab Anim Sci* 45, 334–351.
- Hurley, S. S., Splitter, G. A. & Welch, R. A. (1989). Development of a diagnostic test for Johne's disease using a DNA hybridization probe. *J Clin Microbiol* 27, 1582–1587.
- Jacobson, M. A., Hopewell, P. C., Yajko, D. M. & 7 other authors (1991). Natural history and dissemination of *Mycobacterium avium* complex infection in AIDS. *J Infect Dis* 164, 994–998.
- Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. (1997). Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol Microbiol* 24, 263–270.
- Kunze, Z. M., Wall, S., Appelberg, R., Silva, M. T., Portaels, F. & McFadden, J. J. (1991). IS901, a new member of a widespread class of atypical insertion sequences, is associated with pathogenicity in *Mycobacterium avium*. *Mol Microbiol* 5, 2265–2272.
- Levy-Frebault, V. V., Thorel, M. F., Varnerot, A. & Gicquel, B. (1989). DNA polymorphism in *Mycobacterium paratuberculosis*, "wood pigeon mycobacteria," and related mycobacteria analyzed by field inversion gel electrophoresis. *J Clin Microbiol* 27, 2823–2826.
- Lisitsyn, N. & Wigler, M. (1995). Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol* 254, 291–304.
- Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259, 946–951.
- McClure, H. M., Chiodini, R. J., Anderson, D. C., Swenson, R. B., Thayer, W. R. & Coutu, J. A. (1987). *Mycobacterium paratuberculosis* infection in a colony of stump-tail macaques (*Macaca arctoides*). *J Infect Dis* 155, 1011–1019.
- Matthews, P. R. & McDiarmid, A. (1979). The production in bovine calves of a disease resembling paratuberculosis with a *Mycobacterium* sp. isolated from a woodpigeon (*Columba palumbus* L.). *Vet Rec* 104, 286.
- Mills, J. A., McNeil, M. R., Belisle, J. T., Jacobs, W. R. & Brennan, P. J. (1994). Loci of *Mycobacterium avium* *ser2* gene cluster and their functions. *J Bacteriol* 176, 4803–4808.
- Moss, M. T., Sanderson, J. D., Tizard, M. L., Hermon-Taylor, J., el-Zaatari, F. A., Markesich, D. C. & Graham, D. Y. (1992a). Polymerase chain reaction detection of *Mycobacterium paratuberculosis* and *Mycobacterium avium* subsp. *silvaticum* in long term cultures from Crohn's disease and control tissues. *Gut* 33, 1209–1213.
- Moss, M. T., Malik, Z. P., Tizard, M. L., Green, E. P., Sanderson, J. D. & Hermon-Taylor, J. (1992b). IS902, an insertion element of the chronic-enteritis-causing *Mycobacterium avium* subsp. *silvaticum*. *J Gen Microbiol* 138, 139–145.
- Nagy, I., Schoofs, G., Vanderleyden, J. & De Mot, R. (1997). Transposition of the IS21-related element IS1415 in *Rhodococcus erythropolis*. *J Bacteriol* 179, 4635–4638.
- Pavlik, I., Bejckova, L., Pavlas, M., Rozsypalova, Z. & Koskova, S. (1995). Characterization by restriction endonuclease analysis and DNA hybridization using IS900 of bovine, ovine, caprine and human dependent strains of *Mycobacterium paratuberculosis* isolated in various localities. *Vet Microbiol* 45, 311–318.
- Perez, V., Garcia Marin, J. F. & Badiola, J. J. (1996). Description and classification of different types of lesion associated with

natural paratuberculosis infection in sheep. *J Comp Pathol* 114, 107–122.

Reimann, C., Moore, R., Little, S., Savioz, A., Willetts, N. S. & Haas, D. (1989). Genetic structure, function and regulation of the transposable element IS21. *Mol Gen Genet* 215, 416–424.

Roiz, M. P., Palenque, E., Guerro, C. & Garcia, M. J. (1995). Use of restriction fragment length polymorphism as a genetic marker for typing *Mycobacterium avium* strains. *J Clin Microbiol* 33, 1389–1391.

Sanderson, J. D., Moss, M. T., Tizard, M. L. & Hermon-Taylor, J. (1992). *Mycobacterium paratuberculosis* DNA in Crohn's disease tissue. *Gut* 33, 890–896.

Saxegaard, F. & Baess, I. (1988). Relationship between *Mycobacterium avium*, *Mycobacterium paratuberculosis* and "wood pigeon mycobacteria". Determinations by DNA–DNA hybridization. *APMIS* 96, 37–42.

Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P. R. (1996). Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J Bacteriol* 178, 4885–4893.

Stroeher, U. H., Karageorgos, L. E., Morona, R. & Manning, P. (1992). Serotype conversion in *Vibrio cholerae* O1. *Proc Natl Acad Sci USA* 89, 2566–2570.

Telenti, A., Philipp, W. J., Sreevatsan, S., Bernasconi, C., Stockbauer, K. E., Wiele, B., Musser, J. M. & Jacobs, W. R. (1997). The *emb* operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nat Med* 3, 567–570.

Thorel, M. F., Krichevsky, M. & Levy-Frebault, V. V. (1990). Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. *Int J Syst Bacteriol* 40, 254–260.

Venisse, A., Fournie, J. J. & Puzo, G. (1995). Mannosylated lipoarabinomannan interacts with phagocytes. *Eur J Biochem* 231, 440–447.

Zhang, L., Toivanen, P. & Skurnik, M. (1996). The gene cluster directing O-antigen biosynthesis in *Yersinia enterocolitica* serotype O:8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase. *Microbiology* 142, 277–288.

Received 11 February 1998; revised 11 July 1998; accepted 3 August 1998.