Gene arrangement and organization in a ~ 76 kb fragment encompassing the *oriC* region of the chromosome of *Mycobacterium leprae*

Hafida Fsihi,¹ Edda De Rossi,^{1,2} Leiria Salazar,^{1,3} Rita Cantoni,² Monica Labò,² Giovanna Riccardi,² Howard E. Takiff,³ Karin Eiglmeier,¹ Staffan Bergh^{1,4} and Stewart T. Cole¹

Author for correspondence: Stewart T. Cole. Tel: +33 1 45 68 84 46. Fax: +33 1 40 61 35 83.
e-mail: stcole@pasteur.fr

- ¹ Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris cedex 15, France
- ² Dipartimento di Genetica e Microbiologia, Università di Pavia, 27100 Pavia, Italy
- ³ Laboratorio de Genética Molecular, Centro de Microbiología y Biología Celular, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas, Venezuela
- ⁴ Department of Biochemistry and Biotechnology, Royal Institute of Technology, S-100 44 Stockholm, Sweden

A continuous 75627 bp segment of the *Mycobacterium leprae* chromosome spanning the *oriC* region was sequenced. The gene order at this locus was similar to that found in the replication origin region of many other prokaryotes, particularly *Mycobacterium tuberculosis* and *Streptomyces coelicolor*. As in the case of several Gram-positive bacteria, essential genes involved in basic cellular functions, such as DNA or RNA metabolism (*dnaA*, *dnaB*, *dnaN*, *gyrB*, *gyrA*, *pcnB*, *recF*, *rnpA*, *ssb*), cell wall synthesis (*ponA*, *pbpA*) and probably cell division (*gidB*, *rodA*) were found. Strikingly, the *gidA* gene was absent from this part of the genome and there was no rRNA operon near *oriC*. The *gyrA* gene harbours an intein coding sequence indicating that protein splicing is required to produce the mature A subunit of DNA gyrase. Among the many other noteworthy features were ORFs encoding putative serine/threonine protein kinases and a protein phosphatase, three tRNA genes, one *M. leprae*-specific repetitive element and a *glnQ* pseudogene.

Keywords: *Mycobacterium leprae*, *oriC* region, protein kinases, tRNA genes, penicillin-binding proteins

INTRODUCTION

Mycobacterium leprae is an important human pathogen that infects macrophages and the Schwann cells of the peripheral nervous system causing irreversible nerve damage, loss of protective sensibility and, ultimately, the severe mutilations characteristic of leprosy. The leprosy bacillus is so well adapted to its intracellular habitat that it has evaded axenic cultivation (Wheeler, 1990). Studies of *M. leprae* were greatly facilitated by the development of animal models (Kirchheimer & Storrs, 1971; Shepard & Chang, 1962) and its generation time in infected mice has been determined to be 13 d, amongst the longest reported for bacteria (Shepard & McRae, 1965). Features such as dependence on some key nutrient, requirement for a

particular micro-environment with an unusual pH or redox potential, or an abnormal mode of replication of its chromosome would account for the failure of *M. leprae* to grow *in vitro*. The great challenge now is to apply modern techniques to understand the virulence mechanisms of *M. leprae*, and to resolve the problem of its long doubling time *in vivo* and its lack of growth *in vitro*, parameters that severely hamper biological analysis.

To facilitate molecular genetic analysis of *M. leprae* and obtain direct access to its genes, clones from a 10-fold redundant cosmid library were ordered into four contigs representing the ~ 2.8 Mb chromosome (Eiglmeier *et al.*, 1993) and used as starting material for a genome sequencing project (Fsihi & Cole, 1995; Honoré *et al.*, 1993). About 60% of the chromosome has been sequenced (Cole, 1994) and the vast amount of information generated is stored and disseminated via the mycobacterial database MycDB (Bergh & Cole, 1994).

To address the question of the slow generation time of *M. leprae* and to better understand its growth control mechanisms, we have sequenced and analysed a part of the

Abbreviations: RLEP, *Mycobacterium leprae*-specific repetitive element. The GenBank/EMBL accession numbers for the sequences of cosmids L222 and B1770 reported in this paper are L39923 and Z70722, respectively. The mycobacterial database (MycDB) accession number of the complete sequence is MD0058.

M. leprae chromosome around the *oriC* region, which is known in several other bacterial species to contain genes that are a clock for the cell cycle (Zyskind & Smith, 1992). We report in this paper the nucleotide sequence of a 75627 bp fragment of this region and compare the gene organization with that of its counterpart from other faster-growing and better-known organisms. It is crucial to understand the functions and the organization of key DNA replication genes as this may give fresh insight into the slow growth of *M. leprae* and lead to the identification of new therapeutic targets.

METHODS

DNA sequencing. DNA fragments obtained after digestion of cosmid L222 with the *Bam*HI and *Hind*III restriction enzymes were subcloned in pGEM-4Zf(+) (Promega) and sequenced on both strands with custom-made primers using [³⁵S]dATP (Amersham) and the Sequenase and T7 Polymerase kits (Amersham and Pharmacia, respectively). Cosmid B1770 was sequenced with 6·8-fold redundancy by the shotgun technique described by Fsihi & Cole (1995) using an automated DNA sequencer (A373, Applied Biosystems).

Informatics. Sequence data were transferred to a SPARC-SUN workstation, compiled and assembled using the Staden software package XBAP (Dear & Staden, 1991, 1992; Staden, 1987). DNA and protein sequences were compared with non-redundant databases at the National Center for Biotechnology Information, Washington, DC, USA using the BLAST programs (Altschul *et al.*, 1990, 1994). Further analysis was performed using version 8 of the GCG suite of programs (Genetics Computer Group, University of Wisconsin) and the PROSITE catalogue (Bairoch, 1992). Alignment of homologous proteins was accomplished with PILEUP and calculation of the isoelectric point of the proteins was determined with ISOELECTRIC.

Hybridization experiments. The intragenic *pknA* and *pknB* probes were prepared from appropriate M13 phage by PCR with 20-mer forward and reverse universal primers. To detect the presence of the *povE* gene, a 410 bp internal DNA fragment was purified, after digestion, with the GeneClean (Bio101) procedure. Radioactive labelling was done using [³²P]dCTP and a nick translation kit (Amersham). Hybridizations were carried out overnight at 42 °C (*M. leprae* genomic blots) or 30 °C (*Mycobacterium tuberculosis* cosmids). The blots were washed for 5 min in 2 × SSC (1 × SSC is 0·15 M sodium chloride, 0·015 M sodium citrate)/0·1 % SDS and then for 15 min in 0·2 × SSC/0·1 % SDS at 65 °C or room temperature as appropriate.

RESULTS AND DISCUSSION

Whilst updating the chromosomal map of *M. leprae* (Eiglmeier *et al.*, 1993), the *M. tuberculosis gyrB* and *dnaA* genes (Salazar *et al.*, 1996; Takiff *et al.*, 1994) were used as mapping probes to identify suitable clones carrying the origin of chromosomal replication as they represent predictive markers for *oriC* (Cole & Saint-Girons, 1994). To deduce its genomic organization, two cosmids that hybridized with *dnaA* were used to obtain the nucleotide sequence of this region. Cosmids L222 and B1770 carry *M. leprae* chromosomal DNA inserts of 40561 and 37819 bp, respectively, and overlap by 2972 bp. ORFs present in the resultant 75627 bp segment which has a G+C content of 55·7 mol%, typical of *M. leprae*, were identified on the basis of their codon usage and database

Table 1. Genes and predicted characteristics of deduced protein products in 75627 bp segment of *M. leprae* chromosome spanning the *oriC* region

ORF	Position (from-to)	Codon (first...last)	Molecular mass (kDa)	pI
<i>dnaB'</i>	778–	ATG...		ND
<i>rplI</i>	1785–1327	ATG...TAA	16·207	10·31
<i>ssb</i>	2617–2195	GTG...TGA	15·080	10·33
<i>rpsF</i>	3068–2778	ATG...TAG	10·861	6·53
<i>orf418</i>	4970–3714	GTG...TAG	45·721	9·70
<i>ponA</i>	7365–5305	GTG...TGA	72·725	6·34
<i>povE</i>	10067–11176	ATG...TAA	40·252	5·20
<i>pknB</i>	18362–17127	GTG...TAG	45·316	6·46
<i>orf251</i>	18555–19310	GTG...TGA	28·034	10·77
<i>orf286</i>	19307–20167	GTG...TGA	29·911	4·96
<i>orf512</i>	20164–21702	ATG...TGA	54·864	5·63
<i>mviB</i>	21699–22838	ATG...TAG	39·984	11·06
<i>pknC</i>	22885–25239	GTG...TAG	82·679	7·50
<i>trxD</i>	26893–28269	ATG...TAG	49·015	4·74
<i>cwlA</i>	28402–29019	ATG...TGA	22·219	6·15
<i>orf259</i>	30395–29616	GTG...TAA	28·613	5·12
<i>orf333</i>	31747–30746	ATG...TGA	36·327	6·68
<i>orf278</i>	32586–31750	ATG...TGA	30·466	5·24
<i>gidB</i>	33538–32780	GTG...TGA	27·521	6·63
<i>orf193</i>	34200–33619	ATG...TAG	21·435	4·27
<i>orf312</i>	35423–34485	GTG...TAA	35·675	11·03
<i>rnpA</i>	36124–35762	GTG...TGA	13·319	11·51
<i>rpmH</i>	36289–36146	GTG...TGA	5·689	–
<i>dnaA</i>	36897–38405	TTG...TGA	58·6	5·60
<i>dnaN</i>	38920–40119	ATG...TAA	41·915	4·70
<i>recF</i>	40120–41277	GTG...TGA	42·053	5·79
<i>orf199</i>	41274–41873	GTG...TGA	21·096	12·04
<i>gyrB</i>	42010–44103	GTG...TAG	76·5	6·09
<i>gyrA</i>	44156–47977	ATG...TGA	141·0*	5·98*
<i>ileT</i>	49031–49104	–	–	–
<i>alaT</i>	49241–49313	–	–	–
<i>ppiB</i>	52830–53378	ATG...TAG	19·355	5·85
<i>pabS</i>	55161–55859	ATG...TGA	25·076	6·34
<i>pknB</i>	57704–56757	ATG...TAG	34·117	5·40
<i>orf184</i>	58268–57714	GTG...TGA	19·983	11·47
<i>pknA</i>	59026–58265	ATG...TGA	27·366	9·07
<i>pbbA</i>	60053–59076	ATG...TAG	33·751	6·50
<i>rodA</i>	61893–60496	ATG...TGA	44·695	10·66
<i>orf228</i>	62576–61890	ATG...TGA	23·806	8·53
<i>ppp</i>	63420–62620	GTG...TAG	28·416	4·38
<i>orf176</i>	63947–63417	GTG...TGA	19·468	11·76
<i>orf357</i>	65396–64323	ATG...TGA	38·426	6·11
<i>leuV</i>	65722–65803	–	–	–
<i>orf202</i>	69928–70536	GTG...TAA	22·009	6·28
<i>leuS'</i>	73108–75626	GTG...	> 94·198	ND

ND, Not determined.

*Values correspond to the calculated molecular mass and pI of the GyrA precursor, which probably undergoes protein splicing to yield a mature DNA gyrase A subunit (94·313 kDa, pI 6·03) and an intein (46·834 kDa, pI 5·93) (Fsihi *et al.*, 1996).

Table 2. Similarity of the putative products of ORFs identified on *M. leprae* cosmids L222 and B1770

ORF	Length (aa)	Sequence similarity and/or PROSITE match	BLAST score (P-value)*
<i>dnaB'</i>	First 252	Replicative DNA helicase from: <i>B. subtilis</i> (sp/P37469) <i>E. coli</i> (pir/A03535)	5.6e-53 3.9e-40
<i>rplI</i>	152	50S ribosomal protein L9 from: <i>B. subtilis</i> (sp/P37437) <i>B. stearothermophilus</i> (gp/M57623)	2.9e-26 1.3e-25
<i>ssb</i>	140	Single-strand-DNA-binding protein from <i>B. subtilis</i> (sp/P37455)	1.8e-29
<i>rpsF</i>	96	30S ribosomal S6 from: <i>B. subtilis</i> (sp/P21468) <i>E. coli</i> (gp/U14003)	9.7e-20 1.7e-14
<i>L222-orf418</i>	418	Unknown PS00016: cell attachment sequence	
<i>ponA</i>	686	Penicillin-binding protein 1A from <i>E. coli</i> (sp/P02918) (PBP1A/PonA) Hypothetical 95.5 kDa protein (in locus involved in transformation) from <i>H. influenzae</i> (sp/P31776)	2.3e-60 1.4e-46
<i>povE</i>	369	Inositol-1-phosphate synthase from yeast (sp/P11986)	0.11
<i>pcnB</i>	411	PolyA polymerase from <i>B. subtilis</i> (gp/L38424) <i>gctA</i> and <i>gctB</i> gene products in <i>Acidaminococcus fermentans</i> (gp/X81440) tRNA nucleotidyl transferase (EC 2.7.7.25) from <i>E. coli</i> (sp/P06961) PcnB protein from <i>E. coli</i> (pir/S11887) PS00639: thiol (His) protease active site	8.7e-33 9.5e-28 2.3e-18 8.5e-14
<i>L222-orf251</i>	251	<i>mutT</i> gene product from <i>E. coli</i> (gp/M20791) MutT-like protein (ORF154) from <i>Streptomyces ambofaciens</i> (sp/P32091) PS00893: MutT domain signature	0.00061 0.0013
<i>L222-orf286</i>	286	Unknown	
<i>L222-orf512</i>	512	Unknown PS00342: microbody carboxy-terminal targeting signal	
<i>mviB</i>	379	Virulence factor MviN from <i>S. typhimurium</i> (sp/P37169) PS00029: leucine zipper pattern	2.6e-23
<i>pknC</i>	784	Protein-kinase-like sequence from <i>S. coelicolor</i> (gp/U19250) Eukaryotic-like threonine protein kinase Pkn2 (pir/S21533) from <i>M. xanthus</i> Serine/threonine kinase of <i>M. xanthus</i> (gp/M94857)	2.6e-06 7.1e-06 2.2e-05
<i>trxBA</i>	458	Thioredoxin reductase from: <i>S. clavuligerus</i> (gp/Z21946) <i>N. crassa</i> (gp/D45049) PS00639: thiol (His) protease active site PS00194: thioredoxin	2.2e-125 2.3e-125
<i>cwlA</i>	205	N-Acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) from: <i>Bacillus licheniformis</i> (sp/P36550) <i>B. subtilis</i> (sp/P24808) Autolysin from <i>Clostridium acetobutylicum</i> (sp/P34020)	2.1e-09 3.4e-05 0.00012
<i>L222-orf259</i>	259	Unknown Mob protein from <i>Bordetella bronchiseptica</i>	0.053
<i>L222-orf333</i>	333	Hypothetical 32.4 kDa protein in <i>gidB-uncI</i> intergenic region in <i>P. putida</i> (sp/P31857) Stage 0 sporulation protein J from <i>B. subtilis</i> (Spo0J) (sp/P26497) Hypothetical 32.8 kDa protein in <i>spo0J-gidB</i> intergenic region in <i>B. subtilis</i> (sp/P37524)	8.0e-78 9.6e-74

Table 2. (cont.)

ORF	Length (aa)	Sequence similarity and/or PROSITE match	BLAST score (P-value)*
<i>L222-orf278</i>	278	Hypothetical 28·9 kDa protein in <i>gidB-uncI</i> intergenic region in <i>P. putida</i> (sp/P31856) related to Soj of <i>B. subtilis</i>	4·3e-88
		u0247b (gp/U00021) from <i>M. leprae</i>	1·0e-56
		Nitrogenase iron reductase (EC 1.18.6.1) from <i>Clostridium pasteurianum</i> (sp/P22584)	1·4e-11
		Septum-site-determining protein MinD from <i>B. subtilis</i> (sp/Q01464)	3·5e-11
<i>gidB</i>	252	Glucose-inhibited division protein from: <i>E. coli</i> (sp/P17113)	2·3e-20
		<i>P. putida</i> (sp/P25757)	5·6e-15
		PS00002: glycosaminoglycan attachment site	
<i>L222-orf193</i>	193	Jag protein (SpoIIIJ-associated protein) from <i>B. subtilis</i> (sp/Q01620)	8·9e-13
<i>L222-orf312</i>	312	60 kDa inner-membrane protein from: <i>P. putida</i> (sp/P25754)	2·1e-19
		<i>E. coli</i> (sp/P25714)	2·8e-19
		Stage III sporulation protein J from <i>B. subtilis</i> (sp/Q01625)	4·7e-13
<i>rnpA</i>	120	Ribonuclease P protein component (EC 3.1.26.5) from: <i>Streptomyces coelicolor</i> (pir/D41870)	1·2e-19
		<i>B. subtilis</i> (sp/P21172)	2·3e-09
		PS00013: prokaryotic membrane lipoprotein lipid attachment site	
<i>rpmH</i>	47	50S ribosomal protein L34 from: <i>B. subtilis</i> (sp/P05647)	2·1e-20
		<i>P. putida</i> (sp/P16498)	7·8e-20
		PS00784: ribosomal protein L34 signature	
<i>dnaA</i>	521	Chromosomal replication initiator from: <i>S. coelicolor</i> (sp/P27902)	1·3e-185
		<i>M. luteus</i> (sp/P21173)	3·9e-168
		PS00017: ATP/GTP-binding site motif A (P-loop)	
<i>dnaN</i>	399	β -subunit of DNA polymerase III from: <i>S. coelicolor</i> (sp/P27903)	5·8e-130
		<i>M. luteus</i> (sp/P21174)	1·1e-92
		PS00004: cAMP- and cGMP-dependent protein kinase phosphorylation site	
<i>recF</i>	385	RecF protein from: <i>S. coelicolor</i> (sp/P36176)	1·6e-131
		<i>B. subtilis</i> (sp/P05651)	1·0e-46
		PS00618: RecF protein signature	
<i>B1770-orf199</i>	199	Hypothetical protein in the <i>recF-gyrB</i> intergenic region	
<i>gyrB</i>	697	β -subunit of DNA gyrase from: <i>M. tuberculosis</i> (sp/P41514)	0·0
		<i>M. smegmatis</i> (gp/X94224)	0·0
		<i>S. coelicolor</i> (sp/P35886)	2·1e-265
		PS00004: cAMP- and cGMP-dependent protein kinase phosphorylation site	
		PS00177: DNA topoisomerase II signature	
<i>gyrA</i>	1273†	A subunit of DNA gyrase from: <i>M. tuberculosis</i> (gp/L11919)	0·0
		<i>M. smegmatis</i> (gp/X94224)	0·0
		<i>S. coelicolor</i> (sp/P35885)	0·0
		PS00004: cAMP- and cGMP-dependent protein kinase phosphorylation site	
		P00881: protein splicing signature	
<i>ppiB</i>	182	ha1539 protein related to cyclophilin from <i>Homo sapiens</i> (gp/D38552)	2·5e-30
		Peptidylprolyl <i>cis-trans</i> -isomerase from <i>B. subtilis</i> (sp/P35137)	5·5e-24
		CypH protein from <i>M. leprae</i> (U00011)	1·6e-10
		PS00004: cAMP- and cGMP-dependent phosphorylation kinase site	

Table 2. (cont.)

ORF	Length (aa)	Sequence similarity and/or PROSITE match	BLAST score (P-value)*
<i>pabS</i>	232	Anthranilate synthase component II <i>p</i> -Aminobenzoate synthase PS00606: β -ketoacyl synthase active site	7.6e-73 1.1e-68
<i>pknB</i>	315	Ser/Thr-protein kinase Pkn1 (sp/P33973) from <i>M. xanthus</i> Eukaryotic-like protein threonine kinase Pkn2 (pir/S21533) from <i>M. xanthus</i> PS00107: protein_kinase_ST PS00108: protein_kinase_ATP	8.4e-55 5.3e-66
<i>B1770-orf184</i>	184	Proline-rich glycoprotein from <i>Sorghum bicolor</i>	3.7e-05
<i>pknA</i>	253	Ser/Thr protein kinase Pkn1 (sp/P33973) from <i>M. xanthus</i> Protein-kinase-like sequence from <i>S. coelicolor</i> (gp/U19250) PS00108: protein_kinase_ATP	1.3e-53 1.7e-46
<i>pbpA</i>	325	Penicillin-binding protein 2 from <i>E. coli</i> (sp/P08150) Penicillin-binding protein 2B from <i>B. subtilis</i> (sp/Q07868)	1.8e-21 2.1e-17
<i>rodA</i>	353	Stage V sporulation protein E from <i>B. subtilis</i> Cell division protein FtsW from <i>E. coli</i> Rod-shape-determining protein from <i>E. coli</i> PS00428: cell cycle proteins FtsW/RodA/SpoVE signature	3.2e-22 3.2e-16 0.0014
<i>B1770-orf228</i>	228	Mucin-human fragment (pir/A43932) Viral protein TPX from <i>Thermoproteus</i> phage 1 (pir/S06686)	9.9e-10 2.0e-08
<i>ppp</i>	266	Putative protein phosphatase Hypothetical protein of <i>Synechocystis</i> sp. (gp/D64002) Protein phosphatase 2C (gp/U39690)	6.1e-11 8.8e-08
<i>B1770-orf176</i>	176	Unknown	
<i>B1770-orf357</i>	357	Proline-rich protein from <i>Homo sapiens</i>	4.8e-13
<i>B1770-orf202</i>	202	<i>E. coli</i> K-12 genome (65–68 min region) (gp/U28377) Hypothetical protein H10304 of <i>H. influenzae</i> (sp/P43980)	1.5e-11 5.0e-08
<i>leuS'</i>	> 839	Leucyl-tRNA synthase (EC 6.1.1.4) from several species <i>B. subtilis</i> (sp/P36430) <i>E. coli</i> (sp/P07813)	5.7e-234 3.1e-123

*The P-value, calculated with the BLAST program, represents the probability that two proteins are related. The lower the P-value, the greater the degree of similarity.

†The expected length of the mature GyrA polypeptide is 853 aa, the remaining 420 aa corresponds to the *mle-gyrA* intein (Fsihi *et al.*, 1996).

search similarities (Tables 1, 2). The overall organization of the 42 ORFs identified (Table 1), representing a total coding capacity of 60.3%, as well as the position of the three tRNA genes, a 550 bp *M. leprae*-specific repetitive element (RLEP) (Woods & Cole, 1990) and the non-coding regions are shown in Fig. 1. Many of the genes appear to comprise independent monocistronic transcriptional units although there are at least six discernible operons (*orf251-pknC*, *rpmH-rnpA*, *dnaN-orf199*, *gyrB-gyrA*, *pbpA-pknB* and *orf176-rodA*). As observed in other bacteria, transcription of the majority of the genes (28 out of 45) is predicted to occur with the same polarity as replication, assuming that this is bidirectional and that *oriC* is situated downstream of *dnaA*. Similarities of predicted ORF products with database protein sequences and matches to PROSITE motifs (Table 2) led to the identification of many of the genes and their biological functions.

Organization of the *oriC* region

The regions around the *dnaA* gene are highly conserved in many bacteria such as *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida* (Yoshikawa & Ogasawara, 1991), and the *oriC* region of *M. leprae* resembles those of *M. tuberculosis* and *Mycobacterium smegmatis*, as well as that of *Streptomyces coelicolor* (Salazar *et al.*, 1996). At least six genes, *rnpA-rpmH-dnaA-dnaN-recF-gyrB* (Yoshikawa & Ogasawara, 1991), are conserved in that order in these bacterial species but there are differences in the presence of ORFs located between them (Salazar *et al.*, 1996). Other eubacteria display the same general pattern, but with transpositions or deletions of individual genes. In *M. smegmatis*, an ORF resembling the *gnd* gene of *E. coli* is located between *dnaN* and *recF* but is not present in the *oriC* regions of *M. leprae* or *M. tuberculosis* (Salazar *et al.*, 1996). The *recF* gene seems to be absent from certain

Fig. 1. Overall organization of the ORFs on cosmids L222 and B1770. Arrows indicate the direction of transcription of the genes relative to the physical map. Circles indicate the position of the tRNA genes. See Table 1 for coordinates.

intracellular bacteria and symbionts, for example *Borrelia burgdorferi*, *Mycoplasma capricolum*, *Spiroplasma citri*, and *Buchnera aphidicola* (Casjens & Huang, 1993; Fraser *et al.*, 1995; Lai & Baumann, 1992; Miyata *et al.*, 1993).

With the exception of *E. coli*, the regions flanking the *dnaA* gene function as autonomously replicating sequences and act as the chromosomal replication origin, *oriC*. They contain the DnaA boxes, the hallmark of bacterial chromosome origins (for a review, see Yoshikawa & Ogasawara, 1991) which have an asymmetric 9 bp consensus sequence, 5'-c/tTa/gTCCACA-3'. The *dnaA* gene of *M. leprae* (and *M. smegmatis* and *M. tuberculosis*) is preceded by three DnaA boxes and followed by a further seven (Salazar *et al.*, 1996). Cloning of the upstream and downstream regions of *dnaA* in a vector that is nonreplicative in mycobacteria yielded no transformants in *M. smegmatis* (Salazar *et al.*, 1996) suggesting that specificity may be involved in recognition of the cognate *oriC* sequence by DnaA protein. Biochemical experiments with DNA fragments containing these putative DnaA boxes and purified DnaA protein are underway to define the exact locations of the DnaA boxes of *M. leprae*.

In *M. leprae*, and other Gram-positive bacteria, the *gyrA* gene lies immediately downstream of *gyrB*, whereas in the *E. coli*, *Haemophilus influenzae* and *Mycoplasma genitalium* genomes it is elsewhere (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995). Many bacteria have an *rrn* gene, or a rRNA operon, closely linked to *gyrB* and together these can serve as a surrogate marker for *oriC* (Cole & Saint-Girons, 1994). The *oriC* region of *M. leprae* is devoid of *rrn* genes and its *rrn* operon must be at least 1 Mb from *oriC* as is the case in *M. tuberculosis* (Philipp *et al.*, 1996). This unusual arrangement may be related to the slow growth of these mycobacteria or represent a feature of actinomycetes since in *S. coelicolor*, the closest *rrn* operon, *rrnA*, is situated about 200 kb away from *gyrB* (Redenbach *et al.*, 1996).

On the 5' side, the *dnaA* gene is flanked by the *rpmH* and *rnpA* genes. The former encodes the ribosomal (r-) protein L34, highly homologous to the proteins of *S. coelicolor* (75%) and *Micrococcus luteus* (75%), while the latter codes for the protein subunit of RNase P. As noted by Morse & Schmid (1992), RnpA proteins are not well conserved between members of different genera, although the overall net basic charge and size are similar. Analysis of the 5 kb segment to the left of the *M. leprae* *rnpA* gene revealed an interesting gene organization. Four ORFs, 60K (*spoIIIJ* for *B. subtilis*), 50K, *gidA* and *gidB*, are found in the corresponding regions of *P. putida* and *B. subtilis*, respectively. Their precise functions are uncertain although the 50K protein is a putative GTPase and *gidB* may be involved in cell division (Ogasawara & Yoshikawa, 1992). Of these four ORFs, only the 60K homologue, *orf312* and *gidB* are present in the *oriC* region of *M. leprae* where the arrangement and sequences of *orf312* and *orf193* are homologous to the bicistronic operon *spoIIIJ-jag* in the corresponding region of *B. subtilis* (Errington *et al.*, 1992). The *gidA* gene is found near *oriC*, *dnaA* and *gyrB* in almost all bacteria studied although translocations have been reported in *E. coli*, *B. burgdorferi*, *M. genitalium* and *H. influenzae* (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995; Ogasawara & Yoshikawa, 1992; Old *et al.*, 1992). Extending the comparison between *B. subtilis*, *P. putida* and *M. leprae* in the same region, *orf278* and *orf333* of *M.*

Fig. 2. Comparison of DnaA proteins from four different Gram-positive bacteria. The domain structure of the DnaA protein and the ATP-binding site are shown.

leprae are analogous to *soj* (suppressor of *spo0J*) and *spo0J* of *B. subtilis* (Piggot & Coote, 1976), and to *orf263* and *orf290* of *P. putida*, respectively (Ogasawara & Yoshikawa, 1992). Null mutations in *spo0J* cause a defect in chromosome partitioning resulting in anucleate cells (Iretton *et al.*, 1994), and its product is similar to the ParB proteins that control partitioning of several bacterial plasmids. In addition, the product of the *soj* gene, located directly upstream of *spo0J*, is similar to partitioning proteins of the ParA classes (Ogasawara & Yoshikawa, 1992) although it does not appear to be required for chromosome partitioning in *B. subtilis*. It is suspected that the proteins of *M. leprae* and *P. putida* (Ogasawara & Yoshikawa, 1992), which are similar to Spo0J and to Soj, may also be involved in the process of chromosome partitioning.

Proteins involved in active partitioning of bacterial plasmids, and probably also of bacterial chromosomes, belong to a superfamily of ATPases with three distinct conserved amino acid sequence motifs (Koonin, 1993b). These motifs are all clearly present in the *M. leprae* ORF278 protein which is homologous to another protein of 314 amino acids and unknown function (39% identity), encoded by *M. leprae* cosmid L247 (gp/L11920), suggesting that they arose through gene duplication. Both of these putative mycobacterial proteins are distantly related (Table 2) to the cell-division inhibitor protein MinD (Koonin, 1993b), which, in conjunction with MinC, inhibits septation at all potential division sites in *B. subtilis* (de Boer *et al.*, 1989).

Genes and proteins involved in DNA replication and metabolism

The *M. leprae* *dnaA* gene product, the initiator protein for chromosome replication, contains the four domains (Fig. 2) differing in degrees of homology (Fujita *et al.*, 1992) described in DnaA proteins of other eubacteria. A short homologous amino-terminal section, domain I, is followed by domain II, which is different in sequence and in length. The next segment, domain III, is a large, highly homologous region which is separated from the well-conserved carboxy-terminal domain IV by a short variable stretch. Domain III contains the four motifs that are common to DnaA proteins and the NtrC transcription activator (Koonin, 1993a; Roth & Messer, 1995) and an ATP-binding site, found in many nucleotide-binding proteins: GXXGXGKT (Koonin, 1993a). Recently, it was demonstrated by *in vitro* binding studies that domain IV is responsible for DnaA box recognition (Koonin, 1993a; Roth & Messer, 1995). The functions of the other individual domains are unknown.

About 2 kb downstream from *M. leprae* *dnaA* (Fig. 1) are two ORFs that correspond to *dnaN* and *recF*. In *E. coli*, DnaN is the essential β -subunit of DNA polymerase (McHenry, 1988). Binding of this protein is the final step in formation of the holoenzyme and is believed to anchor the core polymerase to the primed template and increase its processivity. Computer analysis showed that mycobacterial DnaN protein is 52 and 45% identical to the DnaN proteins of *S. coelicolor* and *M. luteus*, respectively,

Fig. 3. Alignment of RecF proteins from six bacteria. Conserved domains are boxed. As in the report by Alonso & Fisher (1995), greek letters denote those regions conserved in all proteins, whereas roman letters indicate the domains conserved only in RecF proteins of Gram-positive bacteria. The amino- (A) and carboxy- (B) terminal nucleotide-binding domains are also shown.

with the identical residues evenly distributed along the polypeptide chain of 399 amino acids.

The termination codon for *dnaN* is immediately followed by the start codon for the *recF* gene whose product is involved in recombination, DNA repair and induction of the SOS response. Searches for protein patterns in the PROSITE database (Bairoch, 1992) detected the RecF signature and the nucleotide-binding motif characteristic for such proteins (Table 2). Amino acid sequence comparison of the six RecF proteins from *S. coelicolor*, *Staphylococcus aureus*, *B. subtilis*, *M. leprae*, *M. tuberculosis* and *M. smegmatis* (Fig. 3) revealed the presence of eight conserved domains specific for these proteins (α - θ) and six further regions conserved only among RecF proteins of Gram-positive origin (a-f).

Once the strands of the DNA duplex are separated by DnaA, the DnaB helicase is recruited to unwind the helix (Baker *et al.*, 1986). A truncated gene, *dnaB'*, encoding the first 252 aa of a protein highly similar to the *B. subtilis* and *E. coli* DnaB helicases was found at the upper end of cosmid L222 (Fig. 1). A second crucial role is performed by DnaB; the presence of the bound protein on the template DNA strand empowers the DnaG primase to synthesize multiple RNA primers at many different sites on the template strand (Baker & Wickner, 1992; Marians, 1992). About 2.0 kb upstream of *dnaB'* (Fig. 1), is the *ssb* gene encoding a 140 aa single-strand binding (SSB) protein that is 42% identical to the *B. subtilis* SSB protein. In *E. coli*, SSB protein not only plays a central and varied

role in DNA replication but also participates in DNA repair and recombination (Meyer & Laine, 1990).

DNA gyrase is a type II topoisomerase that catalyses the ATP-dependent negative supercoiling of covalently closed circular DNA (for a review see Reece & Maxwell, 1991). It is also the primary target of quinolones. Bacterial DNA gyrase is a tetramer consisting of two well-conserved A and B subunits that are encoded by the *gyrA* and *gyrB* genes, respectively. The *M. leprae* GyrA and GyrB proteins are highly similar to other DNA gyrases from several species but especially with the *M. tuberculosis* enzyme (Table 2). Detailed analysis of the deduced primary sequence of GyrA protein revealed the existence of a 420 aa intein (Perler *et al.*, 1994), that may correspond to a homing endonuclease, immediately after the catalytic residue Tyr-130 indicating that protein splicing is required to liberate the A subunit of DNA gyrase (Fsihi *et al.*, 1996).

Genes involved in cell wall biosynthesis and cell division

Penicillin-binding proteins (PBPs) are bacterial enzymes which are involved in peptidoglycan biosynthesis, and some of them are targets of β -lactam antibiotics. Together with β -lactamases, they form a superfamily of penicillin-interacting serine D,D-peptidases (Ghuysen, 1991). On the basis of protein sequence similarities, PBPs have been grouped into three classes: the class A and B high-

molecular-mass PBPs, and the low-molecular-mass PBPs (Ghuysen, 1991). Transglycosylase and transpeptidase activities have been clearly demonstrated for *E. coli* class A high-molecular-mass PBP1A and 1B (Ghuysen, 1991). Class B high-molecular-mass PBPs possess transpeptidase activity, and the low-molecular-mass PBPs generally have D,D-carboxypeptidase activity (Ghuysen, 1991).

PBPs belonging to two of these classes appear to be encoded by the *oriC* region of the *M. leprae* chromosome. The putative product of the *M. leprae* *ponA* gene is predicted to comprise 686 aa with a molecular mass of 72.8 kDa and to belong to class A as defined by Ghuysen (1991). It is most closely related to PBP1A of *E. coli* (Table 2). Within the putative penicillin-binding domain, three appropriately spaced active site motifs can be identified: an SXXK box with the active site Ser-356, an SXN box and the K(H)T(S)G triad (Ghuysen, 1991; Spratt & Cromie, 1988). The *M. leprae* PBP also contains a hydrophobic domain, which most likely acts as a membrane anchor, at the amino-terminus, and a highly conserved 10 aa sequence found in the amino-terminal regions of the class A high-molecular-mass PBPs (Popham & Setlow, 1994). This is the second PBP belonging to class A from *M. leprae* to be discovered as, in a recent study, the characterization of PBP1, which resembles both PBP1A and 1B of *E. coli*, was described (Basu *et al.*, 1996).

Mycobacteria are in general resistant to β -lactam antibiotics; biochemical and genetic studies in other micro-organisms have shown that such resistance involves either enzymic inactivation of the drug or development of PBPs with reduced affinity for the antibiotic (Spratt, 1994). It is known that the high level of intrinsic resistance to β -lactams shown by *Mycobacterium chelonae* and *Mycobacterium fortuitum* is multifactorial and results from a combination of low permeability to these drugs, reduced binding to PBPs and the capacity to produce β -lactamases (Fattorini *et al.*, 1992; Jarlier *et al.*, 1991). It is not clear whether *M. leprae* produces β -lactamases, although its recombinant PBP1 has been shown to exhibit remarkably low affinity for penicillins and cephalosporins (Basu *et al.*, 1996).

The *pbpA* gene (Fig. 1) is predicted to encode a 325 aa residue protein that is similar to the carboxy-terminal domains of the *E. coli* PBP2 and to the *B. subtilis* PBP2B proteins (Table 2). Both of these PBPs belong to class A and, comprising > 600 aa, are considerably longer than the *M. leprae* protein. It thus seems likely that either the mycobacterial *pbpA* product is a D,D-carboxypeptidase or that the gene has been truncated. Preceding *pbpA* is a functionally related gene, *rodA*, that may be involved in the cell cycle as it resembles the *ftsW/rodA* genes of *E. coli* and the *spoVE* gene from *B. subtilis*, respectively. Moreover, the hydropathy profile indicated that, like the related cytoplasmic membrane proteins FtsW, RodA and SpoVE (Ikeda *et al.*, 1989), the *M. leprae* protein is mainly hydrophobic with only a few short hydrophilic domains (data not shown).

In *E. coli*, the *pbpA* and *rodA* genes constitute a single transcriptional unit at about 15 min on the chromosome

in the *leuS-lip* region (Spratt *et al.*, 1980) and the RodA protein is required for the expression of the enzymic activity of the PBP2 protein (Matsuzawa *et al.*, 1989). Fig. 1 shows that in *M. leprae* the chromosomal organization of *pbpA* and *rodA* is different since these two genes are separated by 517 bp. Another gene located near *oriC* that may be involved in cell wall metabolism is *cwlA* as it shows weak similarity to genes from *Bacillus* spp. encoding the cell wall hydrolases (Table 2), that hydrolyse the bond between N-acetylmuramoyl and L-amino acid residues in cell wall glycopeptides.

Translation: tRNA genes, leucyl-tRNA synthetase and poly(A) polymerase

In addition to *rpmH*, two other genes encoding r-proteins were identified on cosmid L222, *rplI* and *rpsF*, separated by about 1.3 kb. The former codes for 50S r-protein L9 while the latter encodes the 30S r-protein S6 (Table 2). It seems likely that, as in *E. coli*, the ribosomal genes are scattered throughout the chromosome and not clustered mainly in one region of the genome as in *B. subtilis*.

Three tRNA genes, *ileT*, *alaT* and *leuV*, were identified on cosmid B1770 (Fig. 1), consistent with the mapping analysis (Eiglmeyer *et al.*, 1993). The *M. leprae* tRNA^{Ala} (TGC), tRNA^{Ile}(GAT) and tRNA^{Leu}(CAG) are 73, 74 and 83 bp in length, respectively, and the highly conserved A at position 37 is present in the Ala and Ile tRNAs. Six other tRNA genes have already been sequenced and, like the r-protein genes, the tRNA genes in *M. leprae* appear to be scattered throughout the genome, probably constituting independent transcriptional units. This is unlike the situation in several eubacteria where tRNA genes are clustered and often closely associated with rRNA genes (Green & Vold, 1993) but reminiscent of the dispersed arrangement seen in *Streptomyces* spp. (Sedlmeier *et al.*, 1994).

The attachment of an amino acid to its cognate transfer RNA molecule is a highly specific reaction catalysed by aminoacyl-transfer RNA synthetases. The truncated gene for the leucyl-tRNA synthetase, *leuS'*, was identified on the lower extremity of cosmid B1770 (Fig. 1). The leucyl-tRNA synthetases studied so far belong to a subgroup of type I synthetase known as the XUX family according to the classification established by Eriani *et al.* (1990). The *M. leprae* LeuS' protein contains all of the motifs characteristic of this family. The HIGH box starts with amino acid 101 whereas the second part of the nucleotide-binding site is KIGKS instead of KMSKS found in many organisms.

In recent years it has become apparent that translational control in prokaryotes is also imposed at the level of transcript stability through the addition of multiple adenosine residues to the 3' end of mRNA, leading to accelerated degradation by 3' exonucleolytic digestion (for a recent review, see Sachs & Whale, 1993). The *E. coli* *pcnB* gene encodes a poly(A) polymerase, which catalyses the template-independent sequential addition of AMP to the 3'-terminal hydroxyl groups of RNAs, and its

Fig. 4. Alignment of the primary structure of three putative serine/threonine protein kinases from *M. leprae* (Pkn2 is encoded by a gene on cosmid B1551). The eleven domains identified in eukaryotic serine/threonine protein kinases are shown and positions of key amino acid residues indicated by an asterisk.

properties are similar to those reported for the mammalian enzyme involved in the polyadenylation of mRNA (Cao & Sarkar, 1992). The amino acid sequence of the poly(A) polymerase is similar to that of the *E. coli* tRNA nucleotidyl transferase, the product of the *cca* gene, which attaches -CCA to the 3' termini of tRNAs (Cudny *et al.*, 1986) but lacks the two nucleotide-binding motifs that occur in the Cca protein sequence (Zhu *et al.*, 1986). *M. leprae* also appears to be able to polyadenylate mRNA as it contains a *pcnB* gene that encodes a protein lacking nucleotide-binding motifs, which resembles the poly(A) polymerases of *B. subtilis* and *E. coli* (Tables 1, 2).

Genes for eukaryotic-like protein kinases and protein phosphatases

Three of the proteins encoded by genes in the *oriC* region of *M. leprae* (*pknA*, *pknB* and *pknC*) are similar to both the eukaryotic-like serine/threonine protein kinases Pkn1 and Pkn2 of *Myxococcus xanthus* (Muñoz-Dorado *et al.*, 1991), to protein kinases from *S. coelicolor* (Muñoz-Dorado *et al.*, 1991; Udo *et al.*, 1995; Urabe & Ogawara, 1995) and to another putative protein kinase, Pkn2 (on cosmid B1551) of *M. leprae* (Table 2). The genes *pknA* and *pknB* are contained in the same transcriptional unit whereas *pknC* is situated in another operon (Fig. 1). There is 38.4% identity in a 271 aa overlap between Pkn1 and Pkn2 of *M. xanthus*, and 46.3% identity in a 253 aa overlap between PknA and PknB. Sequences homologous to *pknA* and *pknB* (data not shown) are also present in the *oriC* region of the *M. tuberculosis* chromosome (Philipp *et al.*, 1996).

The catalytic domains of the eukaryotic serine/threonine protein kinases consist of 11 subdomains, some of which contain highly conserved amino acid residues (Hanks & Quinn, 1991). As shown in Fig. 4, these characteristic subdomains are present in all of the mycobacterial proteins, except PknC. Preliminary experiments in which the *pknB* gene was overexpressed in *E. coli* using a T7 *in vitro* expression system, and incubated with radiolabelled γ -ATP, suggest that PknB can be autophosphorylated. Given the close similarity of PknB to PknA and Pkn2, it is reasonable to expect that this too will have protein kinase activity although it remains to be determined if PknC is really a serine/threonine kinase. Interestingly, situated close to *pknA* and *pknB* on the chromosome of *M. leprae*, is *ppp* (Fig. 1, Table 2) whose product is related to a variety of serine/tyrosine protein phosphatases from yeast and higher eukaryotes as well as to products of some unidentified prokaryotic ORFs (Fig. 5). As the first conserved segment includes the Mg^{2+} -dependent protein phosphatase signature (Bairoch, 1992) and extended similarities are apparent to known eukaryotic protein phosphatases, it is reasonable to suspect that the *ppp* product may also be endowed with this activity.

Cyclophilin-related proteins

The 182 aa protein encoded by *ppiB* shows significant homology with cyclophilin-related proteins from eukaryotes, and peptidylprolyl *cis-trans*-isomerases (PPIases) from prokaryotes, including *M. leprae* itself, that are essential for isomerization of proline peptide (Xaa-Pro) bonds in oligopeptides and for protein folding during

<i>M. leprae</i>	29	RLALADGMGGHAAG	43	94	MGTTLTAILFAGNRLGLVHIGDSRGYLLRDGELTQITKDDTFVQTLVDEGRITAE	150
<i>M. genitalium</i>		LLIVVCDGLGGYKGG			MASTLVLGILTKSDIYIFWIGDSRAYLLFENQAKLVTKDHNLYNQLVA...MNADE	
<i>C. elegans</i>		SVLAVFDGHHGHECS			STAVCAIDMDQKLMALAWLGDSPGYVMSNIEFRQLTRGHSPSDEAREARRVEEAGG	
Mouse		SFFAVYDGHAGSRVA			SGSTAVGVMSPTTHMYF INCGDSRAVLCRNGQVCFSTQDHPKPCNPVEKERIQNAGG	
<i>A. thaliana</i>		HFYGVFDGHGCSHVA			VGSTAVVSVVTPEKIIVSNCGDSRAVLCRNGVAIPLSVDHKPDRPDELIRIQQAGG	
<i>S. cerevisiae</i>		.YFAVFDGHAGIQAS			VSDSDMDLAQHQRKLYTANVGDSRIVLFRNGNSIRLTVDHKASDTLEMQRVEQAGG	
Rat		SFFAVYDGHAGSRVA			SGSTAVGVMSPTTHMYF INCGDSRAVLCRNGQVCFSTQDHPKPCNPVEKERIQNAGG	
Mouse		SFFAVYDGHAGSRVA			SGSTAVGVMSPTTHMYF INCGDSRAVLCRNGQVCFSTQDHPKPCNPVEKERIQNAGG	
<i>S. pombe</i>		.FVAVYDGHAGIQAS			KNRTRRVLYTANAGDARIVLCRDGKAIRLSYDHKGS DANESRRVTQLGG	
<i>A. thaliana</i>		HFFGVYDGHGGSQVA			VGSTSVVAVVFP SHIFVANCGDSRAVLCRGKTALPLSVDHKPDREDEAARIEAAGG	
Consensus		-FFAVYDGH-G----			---T-----Y--N-GDSR-VL-R-G-----T-DHK-----E--RI--AGG	

<i>M. leprae</i>	179	AGDRYLLCSDGLSDPVSDETILEAL	203	220	LRGGGPDNVTVVVADV	235
<i>M. genitalium</i>		QQEFLLLCSDGLYNFVEKELFFEII			IENASNDNITAAVLNL	
<i>C. elegans</i>		SDYLVLACDGLSDVFNERNLDYQLV			IEAGSADNVSVVIGFL	
Mouse		EDEFVVLACDGLWDVMSNEELCEFV			LHKGSRDNMSVVLVCF	
<i>A. thaliana</i>		EDECLILASDGLWDVVPNETACGVA			LARQSSDNVSVVVVDL	
<i>S. cerevisiae</i>		EDKFLILACDGLWDVIDDQDACELI			LENGTTDNVTVMVVFL	
Rat		EDEFVVLACDGLWDVMSNEELCEFV			LHKGSRDNMSIVLVCF	
Mouse		EDEFVVLACDGLWDVMSNEELCEFV			LHKGSRDNMSVVLVCF	
<i>S. pombe</i>		HDEFFIIACDGLWDVSDQEAADFV			LKRLSTDNITCIVVNL	
<i>A. thaliana</i>		EDDCLILASDGLWDVMTDEEACEMA			IQRGSKDNISVVVVDL	
Consensus		-DEF-ILA-DG-WDV---EE-CEFV			L--GS-DNV-VV-V-L	

Fig. 5. Alignment of the amino acid sequences of selected domains from the putative protein phosphatase encoded by *ppp* of *M. leprae* with those of known or suspected serine/threonine protein phosphatases of eukaryotic origin. The numbering refers to the *M. leprae* protein and the first domain shown (aa 29–43) corresponds to the protein phosphatase 2C signature defined by Bairoch (1992). The sources of the other sequences were: *M. genitalium*, s18707; *Caenorhabditis elegans*, u29515; two isoenzymes from *Mus musculus*, s39781, p36993; two isoenzymes from *Arabidopsis thaliana*, s55457, x77116; *Saccharomyces cerevisiae*, p35182; *Rattus norvegicus*, p35815; *Schizosaccharomyces pombe*, p40371.

protein synthesis (Handschrumer *et al.*, 1984). Cyclophilins are a family of proteins that bind the immunosuppressive drug cyclosporin A (Hacker & Fischer, 1993). In *Legionella pneumophila*, an intracellular parasite that, like *M. leprae*, multiplies in macrophages, the Mip (macrophage infectivity potentiator) protein is an essential virulence factor that has been shown to exhibit PPIase activity (Fisher *et al.*, 1992). A *mip*-like gene has also been reported in another intracellular bacterium, *Chlamydia trachomatis* (Lundemose *et al.*, 1991). The *ppiB* gene probably encodes the cytoplasmic form of PPIase (since it lacks a signal peptide) and its product is similar (38.3% identity in 94 aa overlap) to that of *cypH* on *M. leprae* cosmid B1177 (accession number: gp/U00011).

Fused genes: thioredoxin and thioredoxin reductase

The *M. leprae* thioredoxin and thioredoxin reductase have been studied by Wieles *et al.* (1995) who showed that both the substrate and the enzyme are encoded by a single gene, *trxD*. The thioredoxin system operates via redox-active disulphides and provides electrons for a wide variety of different metabolic processes in prokaryotic and eukaryotic cells. Thioredoxin is also a general protein disulphide reductase involved in redox regulation. The *trxD* gene of *M. leprae* encodes a protein of 458 aa, with a predicted molecular mass of 49 kDa, and its amino-terminal part contains the three functional domains that occur in all flavoprotein disulfide reductases (Russel & Model, 1988). The carboxy-terminal part of the protein has the characteristic cysteine dithiol redox-active peptide -WCGPC- of thioredoxin. In bacteria, the Trx and Trp proteins previously identified were encoded by two distinct genes

(*trxA* and *trxB*) located either in separate sites or in a cluster. However, in *M. leprae*, a single, fused gene *trxD*, situated ~9 kb from *dnaA* (Fig. 1), codes for both proteins (Wieles *et al.*, 1995). In *S. coelicolor* and *M. tuberculosis*, the *trx* gene(s) are also located near *oriC* (Philipp *et al.*, 1996; Redenbach *et al.*, 1996).

Duplicated genes

Curiously, although the function of the *povE* gene on cosmid L222 (Table 1) could not be established by database searches, a second copy was found in the nucleotide sequence of the *M. leprae* cosmid B1620 (accession number U00015) located elsewhere on the contig map (Eiglmeier *et al.*, 1993). The two genes are 100% identical; the possibility that either cosmid was chimeric could be excluded, as the region of sequence identity between the two clones was confined to *povE*, beginning 18 bp upstream of the ATG codon and ending 58 bp after the TAA codon. The upstream and downstream sequences were unrelated.

To determine how many copies of this gene were present in the *M. leprae* genome, a Southern blot of DNA digested with four different restriction enzymes was hybridized with an intergenic *povE* probe (Fig. 6). Two different hybridizing fragments were detected in two digestions (*EcoRI* and *ClaI*, lanes 2 and 4) and their sizes were consistent with those predicted by the DNA sequence of cosmids L222 and B1620, respectively. In the *BamHI* and *SaI* digests (lanes 1 and 3), only one hybridizing fragment was identified, but this is consistent with the sequences of both cosmids. From these data it seems that only two

Fig. 6. Southern blot analysis of *M. leprae* DNA showing that two copies of the *povE* gene can be detected by hybridization. The probe was a 450 bp DNA fragment internal to *povE*. Lanes 1–4 contain genomic DNA digested with *Bam*HI, *Eco*RI, *Sal*I and *Cla*I, respectively.

copies of *povE* are present on the *M. leprae* genome. A copy of *povE* was also identified in the *oriC* region of *M. tuberculosis* (Philipp *et al.*, 1996).

Pseudogenes

High scoring matches (e^{-20} – e^{-48}) were obtained when BLASTX analysis of the nucleotide sequence from positions 14816 to 15531 was performed and extensive similarities observed with several ABC amino acid transporters, particularly the glutamine, glutamate and histidine transport proteins. Surprisingly, on translation, no complete ORF was found as four stop codons and two frameshifts were identified, thus indicating that the sequence represented a pseudogene but, nevertheless, the homology with the corresponding regions of amino acid permease genes was clear. Like the other three transport proteins, the ‘product’ of the *glnQ* pseudogene would be membrane-associated, with ‘Q-linkers’ (LSGGQQQ; Argos,

1990) joining domains, but with a mutated ATP-binding site (Fig. 7).

Other genes of known and unknown function

Several of the remaining ORFs showed homology to genes described in other micro-organisms although their functions were often unknown. The known genes include *pabS*, which probably encodes a component of *p*-amino-benzoate synthase, and *mviB* coding for a leucine-rich protein (61/379 residues) that shows 26.9% identity to putative virulence factor genes of *S. typhimurium* and *H. influenzae* (Table 2). The products of some ORFs contain motifs found in adhesins and the MutT repair protein, while others were devoid of any known features.

Implications for cell division and genome evolution

A first step towards understanding the molecular mechanisms that underlie chromosome replication in *M. leprae* is reported in this paper with the description of the *oriC* region containing key replication genes. No gross abnormalities in the nature or organization of this locus were found which might explain the long generation time and functional studies will be required. Extensive similarities in the general organization of the *oriC* region of the chromosomes of the related members of the Actinomycetales family, *M. tuberculosis* (Philipp *et al.*, 1996) and *S. coelicolor* (Redenbach *et al.*, 1996), were detected.

The apparent concentration of genes involved in replication (*dnaA*, *dnaB*, *dnaN*, etc.), partitioning (*orf263*, *orf278*) and cell wall synthesis (*ponA*, *pbpA*, *cwlA*) near *oriC* may be indicative of a link between replication of the chromosome and cell division. Given the presence of three genes whose products are, or resemble, serine/threonine protein kinases and another that may encode a protein phosphatase, it is tempting to suggest that protein phosphorylation may play an important regulatory role in *M. leprae*, as the eukaryotic-like protein kinases have been shown to be involved in the timing of the developmental processes of *M. xanthus* (Muñoz-Dorado *et al.*, 1991; Udo *et al.*, 1995). Protein kinase cascades play major roles in

	1	ATP-binding site		50		100				
<i>M. leprae</i>	MKNKDRNNL	YGT.TVLCVS	DIEYPTNSTV	VTGPAV*SK	STL...NLL	YENDNSNILL	DGRSVLN..Y	DPN*LGQRIG	MLLQHPNLLL	RRTVLDSTVL
<i>B. stearothermophilus</i>	MIYFHVQNKY	YGDFHVLKDI	NLTIHQGEVV	VILGPSGSGK	STLYRCLNRL	ETISSGELIV	DNVKVNDKHI	DINQLRRNIG	MVFQHPNLYP	HMTVLQNITL
<i>C. glutamicum</i>	MIKMTGVQKY	FGDFHALTDI	DLEIPRGQVV	VVLGPSGSGK	STLCTINRL	ETIEEGTIEI	DGKVLPEEGK	GLANLRADVG	MVFQSFNLFH	HLTIKDNVTL
<i>E. coli</i>	MITLKNVSKW	YGHFQVLTDC	STEVKKGEVV	VVGGPSGSGK	STLKIYTNGL	EPVQQGEITV	DGIVVNDKKT	DLAKLSRVG	MVFQHFELFP	HLSIIENLTL
	101	Q-linker		150						200
<i>M. leprae</i>	AQHKLLRLSV	APYRSWVLT	QLYRVSLKNK	ADSRPGTISR	TQQRVVAIAR	ALAMTPQVMF	LYEATSALDP	*TVKGIMKPI	SHLSNDGV.M	IVVTQEIGFA
<i>B. stearothermophilus</i>	APMKVLRIFE	KE.AKETAMY	YLEKVGIPDK	ANAYPSHLSG	GQQQRVAIAR	GLAMKPKIML	FDEPTSALDP	ETIGEVLDMV	KQLAKEGNTM	VVVTHEMGFA
<i>C. glutamicum</i>	APIKVRKMKK	SE.AEKLAMS	LLERVGIANQ	ADKYPAGLSG	GQQQRVAIAC	ALAMNPKIML	FDEPTSALDP	EMVNEVLDMV	ASLAKEGNTM	VCVTHEMGFA
<i>E. coli</i>	AQVKVLKRDK	AP.AREKALK	LLERVGLSAH	ANKFPAGLSG	GQQQRVAIAR	ALCMDPIAML	FDEPTSALDP	EMINEVLDMV	VELANEGNTM	MVVTHEMGFA
	201			243						
<i>M. leprae</i>	RSASDTVVFI	YHDKVVERSK	PEQIFETAQT	KRLQRSLSQV	L..					
<i>B. stearothermophilus</i>	REVADRIVFM	DQGRILEEAP	PEEFSNPKE	ERAKVFLSRI	LNH					
<i>C. glutamicum</i>	RKAADRVLFM	ADGLIVEDTE	PDSFFTNPKS	DRAKDFLGKI	LAH					
<i>E. coli</i>	RKVANRVIFM	DEKGIVEDSP	KDAFFDDPKS	DRAKDFLAKI	LH.					

Fig. 7. Alignment of the primary sequence of the *glnQ* pseudogene with those involved in ATP-dependent glutamate or glutamine transport in other bacteria. Sources were: *Bacillus stearotheophilus*, M61017; *Corynebacterium glutamicum*, X81191; and *E. coli*, U10981. The positions of the ATP-binding site and the Q-linker are shown. Arrowheads denote in-frame stop codons and * indicates where frame-shifting was required.

intracellular signal transduction, regulating various cellular functions in both prokaryotes and eukaryotes (Hanks & Quinn, 1991) as well as in control of the cell cycle of eukaryotes (for a review, see Morgan, 1995).

Another interesting finding to emerge from the present study is further evidence for genome plasticity in *M. leprae*, a bacterium that is generally considered to be genetically inert (Krawiec & Riley, 1990; Williams *et al.*, 1990). Not only is it clear that gene fusion has occurred in the case of the thioredoxin genes but also that gene duplication has taken place, as two copies of *povE* were detected. Gene amplification occurs relatively frequently in bacteria and has been considered as the first step in divergent evolution. For many cellular functions of *E. coli* there are two genes, as if the genetic programme calls for backup systems. In at least some cases the redundant information is regulated differently and is used for different purposes in the cell. This appears to be the case with *povE* as the sequences flanking the two copies are quite distinct. It is also intriguing that the copy located near *oriC* is situated in the middle of a long apparently non-coding region (Fig. 1) although a *povE* homologue has been detected in the *oriC* region of *M. tuberculosis*. The only other example of a gene present in two identical copies on the *M. leprae* chromosome is that of the *polA* gene, encoding DNA polymerase I, which is flanked by two RLEPs, thus resembling a transposable element. A second copy of this gene was found at another locus in the genome, but the two copies were not present in equal amounts and could not be found in all isolates of *M. leprae* (Fsihi & Cole, 1995).

Labadan & Riley (1995) compared 1264 *E. coli* protein sequences and estimated that 38–45 % of them have arisen by gene duplication, followed by sequence divergence leading to different degrees of change in function. In parallel, Brenner *et al.* (1995) have examined the complete set of sequences from the much smaller *H. influenzae* genome and found that at least 30 % of the proteins have arisen by processes that involve gene duplication. In secondary metabolic pathways of *Streptomyces* spp., gene duplication is also a prominent feature. Examples include the genes responsible for the biosynthesis of polyketides such as actinorhodin, daunorubicin and 6-deoxyerythronolide B. In *M. leprae*, possible examples of redundant genes are the two *groEL* genes, the second being arranged in an operon with *groES* (Rinke de Wit *et al.*, 1992), and multiple genes of the fibronectin-binding antigen 85 complex (Thole *et al.*, 1992).

Pseudogenes are typically observed in eukaryotes and presumably represent duplicated genes that are no longer active because they have drifted in sequence. In prokaryotes, pseudogenes were observed in *Bordetella* spp., and it was suggested that the genus has evolved relatively recently and that the species are still adapting to a new host (Willems *et al.*, 1992). In addition to the *hisP* pseudogene, at least two other pseudogenes exist in the *M. leprae* genome. These are carried by cosmids B1496 (accession number gi/466868) and B1549 (EMBL/U00014) and correspond to reading frames with many stop codons and frameshifts that exhibit homology with

the glucose-6-phosphate dehydrogenase and the 1,4- α -glucan branching enzyme, respectively. It is not clear whether these three examples, which are all related to intermediary metabolism, represent failed gene duplication events or genes that no longer served a purpose in *M. leprae* and hence accumulated mutations.

ACKNOWLEDGEMENTS

We are grateful to O. Lefebvre for encouragement and help with the figures. This work received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, and additional funds from the Association Française Raoul Follereau, the Groupement d'Intérêt Public – Groupement de Recherches et d'Études sur les Génomes (GIP-GREG), the STD3 programme of the European Community (grant TS3*CT93*0243) and the Institut Pasteur.

REFERENCES

- Alonso, J. C. & Fisher, L. M. (1995). Nucleotide sequence of the *recF* gene cluster from *Staphylococcus aureus* and complementation analysis in *Bacillus subtilis* *recF* mutants. *Mol Gen Genet* **246**, 680–686.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). A basic local alignment search tool. *J Mol Biol* **215**, 403–410.
- Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994). Issues in searching molecular sequence databases. *Nat Genet* **6**, 119–129.
- Argos, P. (1990). An investigation of oligopeptides linking domains in protein tertiary structure and possible candidates for general gene fusion. *J Mol Biol* **221**, 943–948.
- Bairoch, A. (1992). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res* **20**, 2013–2018.
- Baker, T. A. & Wickner, S. H. (1992). Genetics and enzymology of DNA replication in *Escherichia coli*. *Annu Rev Genet* **26**, 447–477.
- Baker, T. A., Sekimizu, K., Funell, B. E. & Kornberg, A. (1986). Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. *Cell* **45**, 53–64.
- Basu, J., Mahapatra, S., Kundu, M., Mukhopadhyay, S., Nguyen-Distèche, M., Joris, B., van Beeumen, J., Cole, S. T., Chakrabarti, P. & Ghuyssen, J.-M. (1996). Identification and overexpression in *Escherichia coli* of a *Mycobacterium leprae* gene, *pon1*, encoding a high-molecular-mass class A penicillin-binding protein PBP1. *J Bacteriol* **178**, 1707–1711.
- Bergh, S. & Cole, S. T. (1994). MycDB – an integrated mycobacterial database. *Mol Microbiol* **12**, 517–534.
- de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. (1989). A division inhibitor and a topological specificity factor encoded by the minicell locus determine proper placement of the division septum in *E. coli*. *Cell* **56**, 641–649.
- Brenner, S. E., Hubbard, T., Murzin, A. & Chothia, C. (1995). Gene duplication in *H. influenzae*. *Nature* **378**, 140.
- Cao, G. & Sarkar, N. (1992). Identification of the gene for an *Escherichia coli* poly(A) polymerase. *Proc Natl Acad Sci USA* **89**, 10380–10384.
- Casjens, D. P. & Huang, W. M. (1993). Linear chromosomal physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol Microbiol* **8**, 967–980.
- Cole, S. T. (1994). The genome of *Mycobacterium leprae*. *Int J Lepr* **62**, 122–125.

- Cole, S. T. & Saint-Girons, I. (1994).** Bacterial genomics. *FEMS Microbiol Rev* **14**, 139–160.
- Cudny, H., Lupski, J. R., Godson, G. N. & Deutscher, M. P. (1986).** Cloning, sequencing and species relatedness of the *Escherichia coli* *cca* gene encoding the enzyme tRNA nucleotidyltransferase. *J Biol Chem* **261**, 6444–6449.
- Dear, S. & Staden, R. (1991).** A sequence assembly and editing package for efficient management of large projects. *Nucleic Acids Res* **19**, 6481–6483.
- Dear, S. & Staden, R. (1992).** A standard file format for data from DNA sequencing instruments. *DNA Seq* **3**, 107–110.
- Eiglmeier, K., Honoré, N., Woods, S. A., Caudron, B. & Cole, S. T. (1993).** Use of an ordered cosmid library to deduce the genomic organisation of *Mycobacterium leprae*. *Mol Microbiol* **7**, 197–206.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Morras, D. (1990).** Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206.
- Errington, J., Appleby, L., Daniel, R. A., Goodfellow, H., Partridge, S. R. & Yudkin, M. D. (1992).** Structure and function of the *spoIIIJ* gene of *Bacillus subtilis*: a vegetatively expressed gene that is essential for σ^G activity at an intermediate stage of sporulation. *J Gen Microbiol* **138**, 2609–2618.
- Fattorini, L., Orefici, G., Jin, S. H., Scardaci, G., Amicosante, G., Franceschini, N. & Chopra, I. (1992).** Resistance to β -lactams in *Mycobacterium fortuitum*. *Antimicrob Agents Chemother* **36**, 1068–1078.
- Fisher, G., Bang, H., Ludwig, B., Mann, K. & Hacker, J. (1992).** Mip protein of *Legionella pneumophila* exhibits peptidyl-prolyl-cis/trans isomerase (PPIase) activity. *Mol Microbiol* **6**, 1375–1383.
- Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995).** Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Fraser, C. M., Gocayne, J. D., White, O. & 26 other authors (1995).** The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
- Fsihi, H. & Cole, S. T. (1995).** The *M. leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel *polA* locus associated with genomic variability. *Mol Microbiol* **16**, 909–919.
- Fsihi, H., Vincent, V. & Cole, S. T. (1996).** Homing events in the *gyrA* gene of some mycobacteria. *Proc Natl Acad Sci USA* **93**, 3410–3415.
- Fujita, M. Q., Yoshikawa, H. & Ogasawara, N. (1992).** Structure of the *dnaA* and DnaA-box region in the *Mycoplasma capricolum* chromosome: conservation and variations in the course of evolution. *Gene* **110**, 17–23.
- Ghuysen, J. M. (1991).** Serine β -lactamases and penicillin-binding proteins. *Annu Rev Microbiol* **45**, 37–67.
- Green, C. J. & Vold, B. S. (1993).** tRNA, tRNA processing, and aminoacyl-tRNA synthetases. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology and Molecular Genetics*, pp. 683–698. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.
- Hacker, J. & Fischer, G. (1993).** Immunophilins: structure–function relationship and possible role in microbial pathogenicity. *Mol Microbiol* **10**, 445–456.
- Handschumacher, R., Harding, M., Rice, J. & Drugge, R. (1984).** Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* **226**, 544–547.
- Hanks, S. K. & Quinn, A. M. (1991).** Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol* **200**, 38–61.
- Honoré, N., Bergh, S., Chanteau, S & 15 other authors (1993).** Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif–Str regions. *Mol Microbiol* **7**, 207–214.
- Ikeda, M., Sato, T., Wachi, M., Jung, H. K., Ishino, F., Kobayashi, Y. & Matsushashi, M. (1989).** Structural similarity among *Escherichia coli* FtsW and RodA proteins and *Bacillus subtilis* SpoVE, which function in cell division, cell elongation, and spore formation, respectively. *J Bacteriol* **171**, 6375–6378.
- Ireton, K., Gunther, I. N. W. & Grossman, A. D. (1994).** *spoIJ* is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. *J Bacteriol* **176**, 5320–5329.
- Jarlier, V., Gutmann, L. & Nikaido, H. (1991).** Interplay of cell wall barrier and β -lactamase activity determines high resistance to β -lactam antibiotics in *Mycobacterium chelonae*. *Antimicrob Agents Chemother* **35**, 1937–1939.
- Kirchheimer, W. K. & Storrs, E. E. (1971).** Attempts to establish the armadillo (*Dasyus novemcinctus* Linn.) as a model for the study of leprosy. I. Report of lepromatoid leprosy in an experimentally infected armadillo. *Int J Lepr* **39**, 693–702.
- Koonin, E. V. (1993a).** A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res* **21**, 2541–2547.
- Koonin, E. V. (1993b).** A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motifs. *J Mol Biol* **229**, 1165–1174.
- Krawiec, S. & Riley, M. (1990).** Organization of the bacterial chromosome. *Microbiol Rev* **54**, 502–539.
- Labedan, B. & Riley, M. (1995).** Widespread protein sequence similarities: origins of *Escherichia coli* genes. *J Bacteriol* **177**, 1585–1588.
- Lai, C. Y. & Baumann, P. (1992).** Genetic analysis of an aphid endosymbiont DNA fragment homologous to the *rnpA-rpmH-dnaA-gyrB* region of eubacteria. *Gene* **113**, 175–181.
- Lundemose, A. G., Birkelund, S., Fey, S. J., Larsen, P. M. & Christiansen, G. (1991).** *Chlamydia trachomatis* contains a protein similar to the *Legionella pneumophila* mip gene product. *Mol Microbiol* **5**, 109–115.
- McHenry, C. S. (1988).** DNA polymerase III holoenzyme of *Escherichia coli*. *Annu Rev Biochem* **57**, 519–550.
- Marians, K. J. (1992).** Prokaryotic DNA replication. *Annu Rev Biochem* **61**, 673–719.
- Matsuzawa, H., Asoh, S., Kunai, K., Muraiso, K., Takasuga, A. & Ohta, T. (1989).** Nucleotide sequence of the *rodA* gene, responsible for the rod shape of *Escherichia coli*: *rodA* and the *pbpA* gene, encoding penicillin-binding protein 2, constitute the *rodA* operon. *J Bacteriol* **171**, 558–560.
- Meyer, R. R. & Laine, P. S. (1990).** The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev* **54**, 342–380.
- Miyata, M., Sano, K.-I., Okada, R. & Fukumura, T. (1993).** Mapping of replication initiation site in *Mycoplasma capricolum* genome by two-dimensional gel-electrophoresis analysis. *Nucleic Acids Res* **21**, 4816–4823.
- Morgan, D. O. (1995).** Principles of CDK regulation. *Nature* **374**, 131–134.
- Morse, D. P. & Schmidt, F. J. (1992).** Sequences encoding the

- protein and RNA component of ribonuclease P from *Streptomyces bikiniensis* var. *zoozonensis*. *Gene* 117, 61–66.
- Muñoz-Dorado, J., Inouye, S. & Inouye, M. (1991). A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium. *Cell* 67, 995–1006.
- Ogasawara, N. & Yoshikawa, H. (1992). Genes and their organization in the replication origin region of the bacterial chromosome. *Mol Microbiol* 6, 629–634.
- Old, I. G., Saint-Girons, I. & Davidson, B. E. (1992). Unique genetic arrangement in the *dnaA* region of the *Borrelia burgdorferi* linear chromosome: nucleotide sequence of the *dnaA* gene. *FEMS Microbiol Lett* 99, 245–250.
- Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N., Noren, C. J., Thorner, J. & Belfort, M. (1994). Protein splicing elements: inteins and exteins; a definition of terms and recommended nomenclature. *Nucleic Acids Res* 22, 1125–1127.
- Philipp, W. J., Poulet, S., Eiglmeier, K., Pascopella, L., Subramanian, B., Heym, B., Bergh, S., Bloom, B. R., Jacobs, W. R., Jr & Cole, S. T. (1996). An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc Natl Acad Sci USA* 93, 3132–3137.
- Piggot, P. J. & Coote, J. C. (1976). Genetic aspects of bacterial endospore formation. *Bacteriol Rev* 40, 908–962.
- Popham, D. L. & Setlow, P. (1994). Cloning, nucleotide sequence, mutagenesis, and mapping of the *Bacillus subtilis* *phpD* gene, which codes for penicillin-binding protein 4. *J Bacteriol* 176, 7197–7205.
- Redenbach, M., Kieser, H. M., Denapate, D., Eichner, A., Cullum, J., Kinashi, H. & Hopwood, D. A. (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. *Mol Microbiol* 21, 77–96.
- Reece, R. J. & Maxwell, A. (1991). DNA gyrase: structure and function. *CRC Crit Rev Biochem Mol Biol* 26, 335–375.
- Rinke de Wit, T. F., Bekelie, S., Osland, A., Miko, T. L., Hermans, P. W. M., van Soolingen, D., Drijfhout, J.-W., Schoningh, R., Janson, A. A. M. & Thole, J. E. R. (1992). Mycobacteria contain two *groEL* genes: the second *Mycobacterium leprae* *groEL* gene is arranged in an operon with *groES*. *Mol Microbiol* 6, 1995–2007.
- Roth, A. & Messer, W. (1995). The DNA binding domain of the initiator protein DnaA. *EMBO J* 14, 2106–2111.
- Russel, M. & Model, P. (1988). Sequence of thioredoxin reductase from *Escherichia coli*. Relation to other flavoprotein disulphide oxidoreductase. *J Biol Chem* 263, 9015–9019.
- Sachs, A. B. & Whale, E. (1993). Poly(A) tail metabolism and function in eukaryotes. *J Biol Chem* 268, 22955–22958.
- Salazar, L., Fsihi, H., de Rossi, E., Riccardi, G., Rios, C., Cole, S. T. & Takiff, H. (1996). Organization of the origins of replication of the chromosomes of *Mycobacterium smegmatis*, *M. leprae* and *M. tuberculosis* and isolation of a functional origin from *M. smegmatis*. *Mol Microbiol* 20, 283–293.
- Sedlmeier, R., Werner, T., Kieser, H. M., Hopwood, D. A. & Schmieger, H. (1994). tRNA genes of *Streptomyces lividans*: new sequences and comparison of structure and organization with those of other bacteria. *J Bacteriol* 176, 5550–5553.
- Shepard, C. C. & McRae, D. H. (1965). *Mycobacterium leprae* in mice: minimal infectious dose, relationship between staining quality and infectivity and effect of cortisone. *J Bacteriol* 89, 365–372.
- Shepard, C. C. & Chang, Y. T. (1962). Effect of several anti-leprosy drugs on multiplication of human leprosy bacilli in footpads of mice. *Proc Soc Exp Biol Med* 109, 636–638.
- Spratt, B. G. (1994). Resistance to antibiotics mediated by target alterations. *Science* 264, 388–393.
- Spratt, B. G. & Cromie, K. D. (1988). Penicillin-binding proteins of gram-negative bacteria. *Rev Infect Dis* 10, 699–711.
- Spratt, B. G., Boyd, A. & Stoker, N. (1980). Defective and plaque-forming lambda transducing phage carrying penicillin-binding protein-cell shape genes: genetic and physical mapping and identification of gene products from *lip-dacA-rodA-phpA-leuS* region of the *Escherichia coli* chromosome. *J Bacteriol* 143, 569–581.
- Staden, R. (1987). Computer handling of sequence projects. In *Nucleic Acid and Protein Sequence Analysis: a Practical Approach*, pp. 173–217. Edited by M. J. Bishop & C. J. Rawlings. Oxford, IRL Press.
- Takiff, H. E., Salazar, L., Guerrero, C., Philipp, W., Huang, W. M., Kreisworth, B., Cole, S. T., Jacobs, W. R. & Telenti, A. (1994). Cloning and nucleotide sequence of *Mycobacterium tuberculosis* *gyrA* and *gyrB* genes, and characterization of quinolone resistance mutations. *Antimicrob Agents Chemother* 38, 773–780.
- Thole, J. E. R., Schöningh, R., Janson, A. A. M., Garbe, T., Cornelisse, Y. E., Clark-Curtiss, J. E., Kolk, A. H. J., Ottenhoff, T. H. M., De Vries, R. R. P. & Abou-Zeid, C. (1992). Molecular and immunological analysis of a fibronectin-binding protein antigen secreted by *Mycobacterium leprae*. *Mol Microbiol* 6, 153–163.
- Udo, H., Munoz-Dorado, J., Inouye, M. & Inouye, S. (1995). *Mycococcus xanthus*, a Gram-negative bacterium, contains a transmembrane protein serine/threonine kinase that blocks the secretion of β -lactamases by phosphorylation. *Genes Dev* 9, 972–983.
- Urabe, H. & Ogawara, H. (1995). Cloning, sequencing and expression of serine/threonine kinase-encoding genes from *Streptomyces coelicolor* A3(2). *Gene* 153, 99–104.
- Wheeler, P. R. (1990). Recent research into the physiology of *Mycobacterium leprae*. *Adv Microb Physiol* 31, 71.
- Wiele, B., van Soolingen, D., Holmgren, A., Offringa, R., Ottenhoff, T. & Thole, J. (1995). Unique gene organization of thioredoxin and thioredoxin reductase in *Mycobacterium leprae*. *Mol Microbiol* 16, 921–929.
- Willems, R. J. L., van der Heide, H. G. J. & Mooi, F. R. (1992). Characterization of a *Bordetella pertussis* fimbrial gene cluster which is located directly downstream of the filamentous haemagglutinin gene. *Mol Microbiol* 18, 2661–2671.
- Williams, D. L., Gillis, T. P. & Portaels, F. (1990). Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol* 4, 1653–1659.
- Woods, S. A. & Cole, S. T. (1990). A family of dispersed repeats in *Mycobacterium leprae*. *Mol Microbiol* 4, 1745–1751.
- Yoshikawa, H. & Ogasawara, N. (1991). Structure and function of DnaA and the DnaA-box in eubacteria: evolutionary relationships of bacterial replication origins. *Mol Microbiol* 5, 2589–2597.
- Zhu, L., Cudny, H. & Deutscher, M. P. (1986). A mutation in *Escherichia coli* tRNA nucleotidyl transferase that affects only AMP incorporation is in a sequence often associated with nucleotide binding proteins. *J Biol Chem* 261, 14875–14877.
- Zyskind, J. W. & Smith, D. W. (1992). DNA replication, the bacterial cell cycle, and cell growth. *Cell* 69, 5–8.

Received 29 April 1996; revised 16 July 1996; accepted 17 July 1996.