4420–4449 *Nucleic Acids Research, 1996, Vol. 24, No. 22*

Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*

Ralf Himmelreich, Helmut Hilbert[+], Helga Plagens, Elsbeth Pirkl, Bi-Chen Li[§] and Richard Herrmann*

Zentrum für Molekulare Biologie Heidelberg, Mikrobiologie, Universität Heidelberg, 69120 Heidelberg, Germany

Received August 22, 1996; Revised and Accepted October 10, 1996

DDBJ/EMBL/GenBank accession no. U00089

ABSTRACT

The entire genome of the bacterium *Mycoplasma pneumoniae* M129 has been sequenced. It has a size of 816 394 base pairs with an average G+C content of 40.0 mol%. We predict 677 open reading frames (ORFs) and 39 genes coding for various RNA species. Of the predicted ORFs, 75.9% showed significant similarity to genes/proteins of other organisms while only 9.9% did not reveal any significant similarity to gene sequences in databases. This permitted us tentatively to assign a functional classification to a large number of ORFs and to deduce the biochemical and physiological properties of this bacterium. The reduction of the genome size of *M.pneumoniae* during its reductive evolution from ancestral bacteria can be explained by the loss of complete anabolic (e.g. no amino acid synthesis) and metabolic pathways. Therefore, *M.pneumoniae* depends in nature on an obligate parasitic lifestyle which requires the provision of exogenous essential metabolites. All the major classes of cellular processes and metabolic pathways are briefly described. For a number of activities/functions present in *M.pneumoniae* according to experimental evidence, the corresponding genes could not be identified by similarity search. For instance we failed to identify genes/proteins involved in motility, chemotaxis and management of oxidative stress.

INTRODUCTION

The bacterium *Mycoplasma pneumoniae* has a genome size of ~800 kb and completely lacks a cell wall. The bacterium is surrounded by a cytoplasmic membrane only, which contains cholesterol as an indispensable component. *Mycoplasma pneumoniae* is a human pathogen, causing 'atypical pneumonia' (1) usually in older children and young adults. As a surface parasite, it attaches to the host's respiratory epithelium by means of a differentiated terminal structure termed attachment organelle or tip structure. For a long time, research activities mainly focused on pathogenicity-related topics such as studies on cytadherence (2), vaccination and diagnosis (3). *Mycoplasma pneumoniae* was not considered as an organism suitable for basic studies partly because of its fastidious growth requirements and partly because of the lack of established standard genetic tools like conjugation or transformation with self-replicating vectors (4). These disadvantages can be compensated now to a large extent by the methods of molecular biology.

Morowitz pointed out in 1984, that mycoplasmas would be suitable candidates for defining the genetic constitution of a minimal self-replicating cell (5). The advantage of these bacteria for such studies (6,7), mainly due to their small genome size, was so obvious that several initiatives were started to sequence five different mycoplasma genomes: *Mycoplasma genitalium* (8,9), *M.pneumoniae* (10), *Mycoplasma capricolum* (11), *Mycoplasma mycoides* (12) and a species from the related genus *Ureaplasma*, *Ureaplasma urealyticum* (13). So far, only the complete sequence of the *M.genitalium* genome has been published (9) which, with 580 070 bp, is the smallest bacterial genome known so far. In the genus *Mycoplasma*, *M.pneumoniae* and *M.genitalium* are the closest related species. We report in this publication the complete nucleotide sequence of the genome of *M.pneumoniae*, which thus provides information on a second small bacterial genome. All *M.pneumoniae* genes which had been already sequenced were reanalyzed except for the P1 operon (14). Our sequencing strategy, early results and a detailed description of *M.pneumoniae* as an experimental system have been recently published (10).

MATERIALS AND METHODS

Mycoplasma strain

The strain *Mycoplasma pneumoniae* M129 (ATTC 29342) in the 18th broth passage was used to construct an ordered cosmid library containing the complete genome (15). This cosmid library was the basis for the DNA sequence analysis. We selected this specific bacterial strain because it has been used in cytadherence and pathogenicity studies (2,16,17). The strain in the 20th broth passage was still infectious in hamsters (H. Brunner, unpublished data).

DNA sequencing

Using the enzymatic dideoxy chain-termination method (18), the sequence data for this study were exclusively generated on a fluorescent-based sequence-gel reader (Model 373A, Applied Biosystems). Sequencing strategies and methods were as described in Hilbert *et al.* (10).

*To whom correspondence should be addressed. Tel: +49 6221 54 68 27; Fax: +49 6221 54 58 93; Email: r.herrmann@mail.zmbh.uni-heidelberg.de

Present addresses: [+]QIAGEN GmbH, 40724 Hilden, Germany and [§]Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

Computer assisted analysis

Sequence assembly, map drawing and multiple alignments were done with the *Lasergene* program package (DNA STAR).

Other analyses were performed with the *HUSAR* (Heidelberg Unix Sequence Analysis Resources) program package release 4.0 at the German Cancer Research Center, Heidelberg, Germany. This package is based on the *GCG* program package version Unix-8.1 of the Genetics Computer Group, Wisconsin. For searching the DNA and protein databases [*SWISS-PROT* (19) and *PIR* (20)] the *FASTA* (21) and *BLAST* (22) programs (*BLASTX*, *BLASTN* and *BLASTP*) were used. Conserved motifs in proteins and peptides were identified by using the program *PROSITE* (23). Open reading frames (ORFs) were calculated by the program *FRAMES* allowing AUG (or GUG, UUG) as start codons using the Mycoplasma translation table where UGA codes for tryptophan (24). The G+C content was calculated by the program *WINDOW*. Codon usage was performed with the program *CODONFREQUENCY*.

The programs *TopPred 1.1.1* (Manuel G. Carlos, Ecole Normale Superieure, Laboratoire de Genetique Moleculaire, Paris, France) and PSORT (25) (http://psort.nibb.ac.jp/) were used for the prediction of transmembrane domains and the membrane topology of proteins.

Each ORF analysis is accessible as a *File Maker Pro* (Claris) database which can be accessed at our world wide web (www) site (http://zmbh.uni-heidelberg.de/M_pneumoniae). It contains, besides genome and cosmid position of each ORF/gene, data about expression, availibility of antibodies, comments, literature, prosite patterns, amino acid composition and database search homology scores. All the annotations in this paper were done on the basis of the highest score values.

Accession number

The complete *M.pneumoniae* sequence has been annotated in GenBank (NCBI) with the accession number U00089.

RESULTS AND DISCUSSION

The strategy and methodology for sequencing the complete genome has been described by us recently (10). A total of 2 415 202 nucleotides primary sequence data were provided by 6385 sequencing reactions. Each strand of the genome was completely sequenced at least once. The direct sequencing approach, combining primer walking with a limited shotgun strategy based on a complete cosmid and plasmid library considerably facilitated the assembly of the individual sequences to the entire genome sequence. The average redundancy of the sequencing was 2.95 (calculated for both strands). This very low redundancy was achieved by the use of 5095 oligonucleotides.

The complete *M.pneumoniae* genome has a size of 816 394 bp and a G+C content of 40.0 mol%. Altogether 677 open reading frames (ORFs) and 39 genes coding for various RNA species were predicted. All ORFs were sorted into categories according to their proposed functions (Tables 1 and 2; Fig. 1). Only 333 ORFs (49.2%) were functionally assigned, based on significant sequence similarities to genes or proteins from other organisms with known functions (e.g. ribosomal proteins) or at least known categories of function (e.g. proteins involved in cytadherence). Significant similarities to proteins without known function from other bacteria, mostly *M.genitalium*, were shown for 181 proposed ORFs (26.7%). We also included in this group those *M.pneumoniae* proteins which were identified in protein extracts of *M.pneumoniae* by monospecific antibodies or by the N-terminal amino acid sequences of enriched proteins (26,27). The group of ORFs without significant similarity or without indication for their *in vivo* expression comprised 109 members (16.1%); 42 of them carry characteristic motifs, which are not sufficient for defining a function. Examples of such motifs are the leucine zipper (29 cases; refered to all predicted ORFs), the typical prokaryotic lipoprotein sequence pattern (46 cases) or ATP- and GTP-binding sites (73 cases). In addition all predicted gene products were analyzed by programs for structure predictions, e.g. coiled/coiled structures (29 cases) or transmembrane segments (275 cases). The latter result supports the analysis of cell fractionation experiments which indicate that the membrane fraction contains ~50% of the total proteins estimated by SDS-PAGE. About 8% of the genome is composed of repetitive DNA elements RepMP1, RepMP2/3, RepMP4 and RepMP5, while only 67 of all predicted ORFs (9.9%) code for a product without any similarity to a known RNA or protein.

Finally, 58 gene families were defined comprising 298 proteins with at least two but frequently with more paralogs; these are proteins with similarities within the same species (see www pages).

The proposed ORFs are not equally distributed over the genome. A lower coding density coincides with regions of lower or higher G+C content than the average. There are regions with a G+C content of up to 56 mol%. These regions code almost exclusively for the gene P1 and gene ORF6 of the P1 operon, the repetitive DNA sequences RepMP4, RepMP2/3, RepMP5 and tRNAs (for details see www pages).

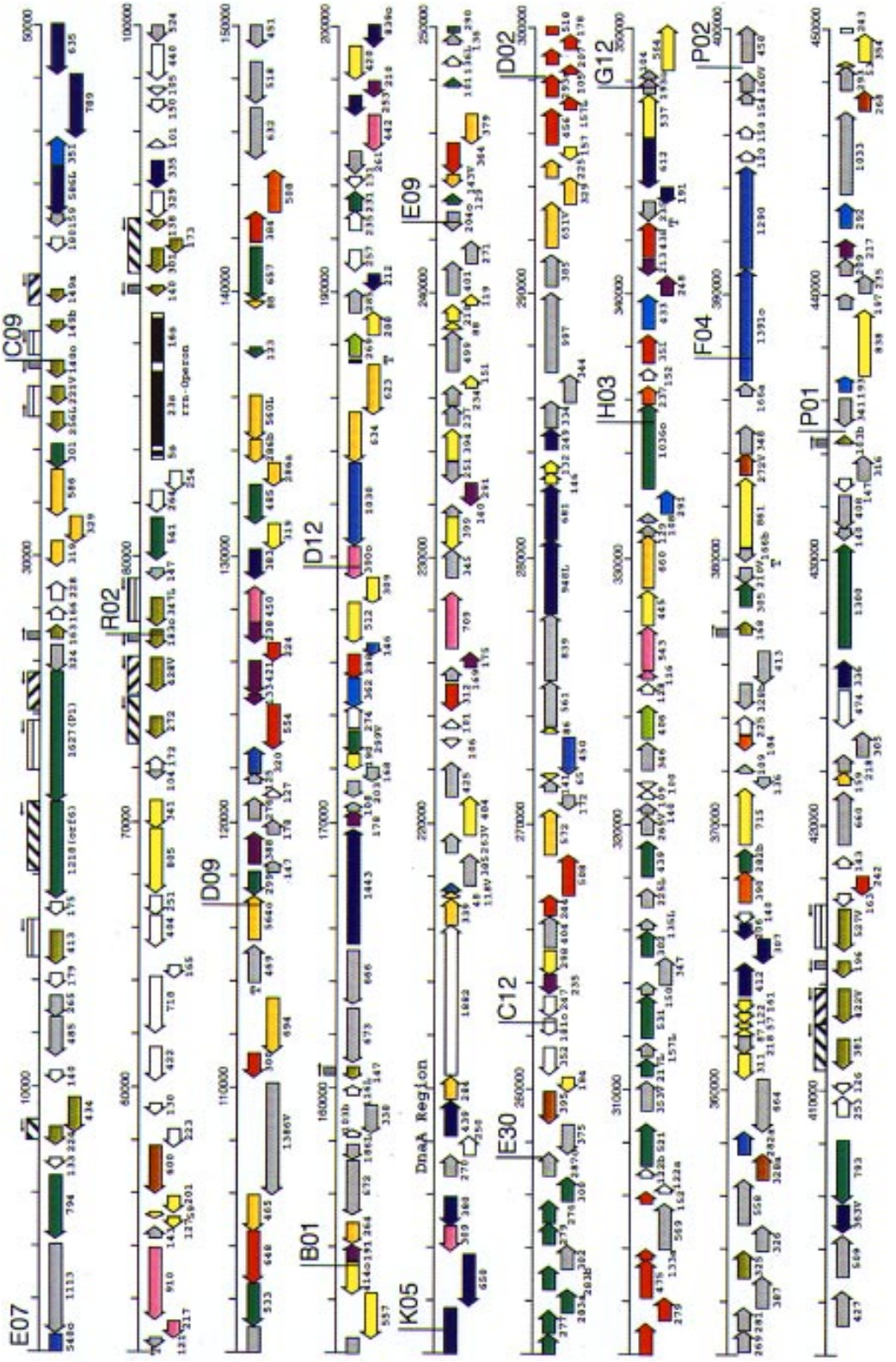
The P1 protein, the main adhesin, is essential for adherence of *M.pneumoniae* to its host cell (28) and the ORF6 gene product which is only found as a cleavage product, namely a 40 and 90 kDa protein, instead of the expected 130 kDa protein, is involved in an as yet unknown manner in cytadherence (14). Gene P1 contains a copy each of RepMP2/3 and RepMP4 and gene ORF6 one of RepMP5 (29). In addition, several copies of each of these repetitive DNA sequences can easily be recognized by their relative high G+C content (Fig. 2).

At the other extreme is the proposed origin of replication around nucleotide position 205 000 (pcosMPK05, dnaA region), with a G+C content of only 26 mol% (10).

Other regions with a low G+C content do not show a similar obvious coding pattern, but proposed ORFs coding for lipoproteins or the hsd modification/restriction system are frequently located in these regions.

The total length of all coding regions is 724 174 bp. The average coding density of 88.7% was calculated for the *M.pneumoniae* genome which gives an average gene size of 1011 bp. Similar

Figure 1. (Following two pages) The gene map of the complete *M.pneumoniae* genome. The arrows indicate the position and the size of the predicted ORFs. The colour refers to the functional category in which the ORFs are sorted. The complete name of an ORF can be deduced by the cosmid name above the horizontal scale-line and the number below the arrows (e.g. the ORF name of the first complete arrow in this figure is E07_orf1113). Rectangles above the scale-line indicate the size and the position of different repetitive DNA sequences (see also Table 4).



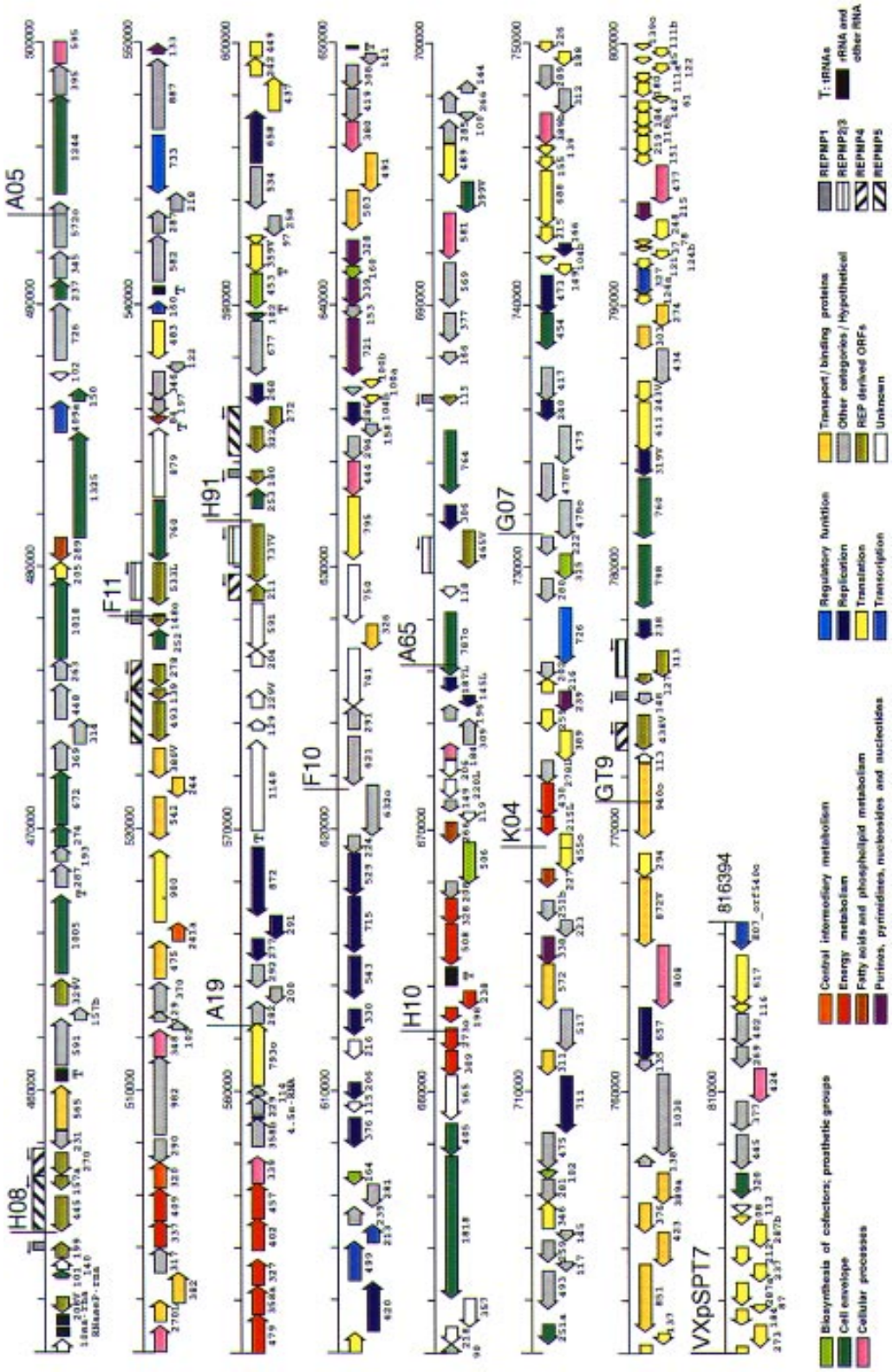


Table 1. Predicted functions and classification of all *M.pneumoniae* ORFs

• **Biosynthesis of cofactors, prosthetic groups and carrier - Folic acid [5]**		
F10_orf160	[a]MG228	dihydrofolate reductase (dhfr); LACLA
H10_orf506	MG213	dihyrofolate reductase (dyr) homolog protein; ENTFC
D12_orf269	MG013	5,10-methylene-tetrahydrofolate dehydrogenase (mtd1); HAEIN
D02_orf606	MG394	serine hydroxymethyltransferase (glyA); ACTAC
H91_orf164	MG245	5-formyl tetrahydrofolate cyclo-ligase (HI0858) homolog; HAEIN
• **Biosynthesis of cofactors, prosthetic groups and carrier - Heme and porphyrin [1]**		
H91_orf453	MG259	possible protoporphyrinogen oxidase (hemK); ECOLI
• **Biosynthesis of cofactors, prosthetic groups and carrier - Thioredoxin [2]**		
A65_orf102	MG124	thioredoxin (trx); YEAST
K04_orf315	MG102	thioredoxin reductase (trxB); EUBAC
• **Cell envelope - Membranes, lipoproteins and porines [42]**		
A05_orf1244	MG307	putative lipoprotein, MG307 homolog, MYCGE
A05_orf252	MG440	putative lipoprotein, MG440 homolog, MYCGE
A65_orf251a	MG440	putative lipoprotein, MG440 homolog, MYCGE
A65_orf787o	MG260	putative lipoprotein, MG260 homolog, MYCGE
A65_orf794	MG260 (MG185)	putative lipoprotein, MG260 homolog, MYCGE
D02_orf217L	MG395 (MG068)	putative lipoprotein, MG395 homolog, MYCGE
D02_orf302	MG068 (MG395)	putative lipoprotein, MG068 homolog, MYCGE
D02_orf439	MG068 (MG395)	putative lipoprotein, MG068 homolog, MYCGE
D02_orf521	MG395 (MG068)	putative lipoprotein, MG395 homolog, MYCGE
D02_orf531	MG395 (MG068)	putative lipoprotein, MG395 homolog, MYCGE
D09_orf123	-	putative lipoprotein
D09_orf485	MG045	putative lipoprotein, MG045 homolog, MYCGE
D09_orf657	MG040	putative lipoprotein, MG040 homolog, MYCGE
D12_orf231	-	putative lipoprotein
E07_orf301	MG186	putative lipoprotein, MG186 homolog, MYCGE
E07_orf794	MG260 (MG185)	putative lipoprotein, MG260 homolog, MYCGE
E09_orf101	marginal MG440	putative lipoprotein
E09_orf129	-	putative lipoprotein
E09_orf276	MG440	putative lipoprotein, MG440 homolog, MYCGE
E09_orf277	MG440	putative lipoprotein, MG440 homolog, MYCGE
E09_orf279	MG439	putative lipoprotein, MG439 homolog, MYCGE
E09_orf283a	MG439	putative lipoprotein, MG439 homolog, MYCGE
E09_orf283b	MG439	putative lipoprotein, MG439 homolog, MYCGE
E09_orf290	MG439	putative lipoprotein, MG439 homolog, MYCGE
E09_orf300	MG439	putative lipoprotein, MG439 homolog, MYCGE
F11_orf760	MG260 (MG185)	putative lipoprotein, MG260 homolog, MYCGE
G07_orf454	MG095	putative lipoprotein, MG095 homolog, MYCGE
G12_orf305	MG348	putative lipoprotein, MG348 homolog, MYCGE
GT9_orf760	MG185	putative lipoprotein, MG185 homolog, MYCGE
GT9_orf798	MG260	putative lipoprotein, MG260 homolog, MYCGE
H08_orf1005	MG321	putative lipoprotein, MG321 homolog, MYCGE
H08_orf1325	MG309	putative lipoprotein, MG309 homolog, MYCGE
H08_orf150	MG307	putative lipoprotein, MG307 homolog, MYCGE
H08_orf237	MG307	putative lipoprotein, MG307 homolog, MYCGE
H91_orf102	MG260	putative lipoprotein, MG260 homolog, MYCGE
H91_orf253	-	putative lipoprotein
P01_orf101	-	putative lipoprotein
P02_orf1300	MG338	putative lipoprotein, MG338 homolog, MYCGE
P02_orf793	MG260	putative lipoprotein, MG260 homolog, MYCGE
R02_orf533	MG067	putative lipoprotein, MG067 homolog, MYCGE
R02_orf541	MG260	putative lipoprotein, MG260 homolog, MYCGE
VXpSPT7_orf320	MG149	putative lipoprotein, MG149 homolog, MYCGE
• **Cell envelope - Surface structures and cytadherence [8]**		
E07_orf1627	MG191 (MG192)	adhesin P1 (orf5, P1 operon); MYCPN
E07_orf1218	MG192 (MG191)	hypothetical 130K protein (orf6; P1 operon); MYCPN
H08_orf274	MG318	30K adhesin-related protein; MYCPN
H08_orf1018	MG312	cytadherence accessory protein (hmw1); MYCPN
F10_orf1818	MG218	cytadherence accessory protein (hmw2); MYCPN
H08_orf672	MG317	cytadherence accessory protein (hmw3); MYCPN
D02_orf1036o	MG386	protein P200; MYCPN
F10_orf405	MG217	protein P65; MYCPN
• **Cell envelope - Surfaces polysaccharides, lipopolysaccharides and antigens [4]**		
A65_orf399V	MG137	YefE protein homolog; ECOLI
B01_orf299V	MG025	TrsB protein; YEREN
D09_orf299	MG060	hypothetical protein YWDF homolog; BACSU
G12_orf282b	MG356	LicA protein homolog; HAEIN
• **Cellular processes - Cell division [2]**		
F10_orf380	MG224	cell division protein (ftsZ); BACSU
K05_orf709	MG457	cell division protein (ftsH); BACSU
• **Cellular processes - Cell killing [1]**		
VXpSPT7_orf424	MG146	hemolysin (hlyC) homolog protein; HAEIN
• **Cellular processes - Chaperones [7]**		
A05_orf595	MG305	heat shock protein DnaK, ERYRH

Table 1. *Continued*

C09_orf217	MG201	heat shock protein GrpE, HAEIN
D02_orf116	MG393	heat shock protein GroES; BACSU
D02_orf543	MG392	heat shock protein GroEL; BACSU
D12_orf390o	MG019	heat shock protein DnaJ; BACSU
C09_orf910	MG200	DnaJ homolog protein, MYCCA
K05_orf309	MG002	DnaJ homolog protein; YEAST
• **Cellular processes - Detoxification [1]**		
D12_orf442	MG008	possible thiophene and furan oxidation protein (tdhF); BACSU
• **Cellular processes - Protein and peptide secretion [9]**		
A05_orf348	MG297	cell division protein (ftsY); ECOLI
D09_orf450	MG048	signal recognition particle protein (ffh); MYCMY
G07_orf808	MG072	preprotein translocase (secA); BACSU
GT9_orf477	MG170	preprotein translocase secY subunit; MYCCA
A65_orf581	MG138	GTP-binding membrane protein (lepA); HAEIN
F10_orf444	MG238	trigger factor (tig); HAEIN
H10_orf184	MG210	prolipoprotein signal peptidase (lsp); STACA
G07_orf389b	MG086	prolipoprotein diacylglyceryl transferase (lgt); ECOLI
F11_orf339	MG270	lipoate protein ligase (lplA); ECOLI
• **Central intermediary metabolism - Other [5]**		
A05_orf241a	MG293	glycerophosphoryl diester phosphodiesterase (glpQ); BACSU
A05_orf320	MG299	phosphotransacetylase (pta); BACSU
D09_orf508	MG038	glycerol kinase (glpK), HAEIN
G12_orf390	MG357	acetate kinase (ackA); BACSU
H03_orf237	MG385	glycerophosphoryl diester phosphodiesterase (glpQ); STAAU
• **Central intermediary metabolism - Phosphorous compounds [1]**		
G12_orf184	MG351	inorganic pyrophosphatase (ppa); THEAC
• **Energy metabolism - Aerobic [3]**		
K05_orf312	MG460	L-lactate dehydrogenase (ldh); MYCHY
D09_orf384	MG039	aerobic glycerol-3-phospate dehydrogenase (glpD); ECOLI
F11_orf479	MG275	NADH oxidase (nox); ENTFA
• **Energy metabolism - Amino acids and amines [5]**		
F10_orf309	-	carbamate kinase (EC 2.7.2.2) (arcC); PSEAE
H03_orf438	-	arginine deiminase (arcA); PSEPU
H10_orf198	-	arginine deiminase (arcA); MYCCA
H10_orf238	-	arginine deiminase (arcA); MYCCA
H10_orf273o	-	ornithine carbamoyl transferase (otc1); ECOLI
• **Energy metabolism - Anaerobic [1]**		
H03_orf351	-	NADP-dependent alcohol dehydrogenase (adh); THEBR
• **Energy metabolism - ATP-proton motive force interconversion [9]**		
C12_orf293o	MG405	ATP synthase A chain (atpB); MYCGA
D02_orf207	MG403	ATP synthase B chain (atpF); MYCGA
D02_orf105	MG404	ATP synthase C chain (atpE); MYCGA
C12_orf157L	MG406	ATP synthase protein I (atpI); MYCGA
D02_orf518	MG401	ATP synthase alpha chain (atpA); MYCGA
D02_orf475	MG399	ATP synthase beta chain (atpD); MYCGA
D02_orf279	MG400	ATP synthase gamma chain (atpG); MYCGA
D02_orf178	MG402	ATP synthase delta chain (atpH); MYCGA
D02_orf133a	MG398	ATP synthase epsilon chain (atpC); MYCGA
• **Energy metabolism - Glycolysis [10]**		
A05_orf337	MG301	glyceraldehyde-3-phosphate dehydrogenase(gap), CLOPA
A05_orf409	MG300	phosphoglycerate kinase (pgk); THEMA
B01_orf288	MG023	fructose-bisphosphate aldolase (tsr); BACSU
C12_orf244	MG431	triosephosphate isomerase (tim); ECOLI
C12_orf456	MG407	enolase (eno) (EC 4.2.1.11); PLAFA
C12_orf508	MG430	phosphoglycerate mutase (pgm); BACSU
H10_orf328	MG215	6-phosphofructokinase (pfk); ECOLI
H10_orf508	MG216	pyruvate kinase (pyk); LACLA
K04_orf430	MG111	phosphoglucose isomerase B (pgiB); BACST
R02_orf300	MG063	1-phosphofructokinase (fruK); HAEIN
• **Energy metabolism - Pentose Phosphate pathway [2]**		
P02_orf242	-	L-ribulose-5-phosphate 4-epimerase (araD); ECOLI
R02_orf648	MG066	transketolase 1 (TK 1; tktB); RHOSH
• **Energy metabolism - Pyruvate DHase [4]**		
F11_orf327	MG273	pyruvate dehydrogenase E1-beta subunit (pdhB); ACHLA
F11_orf358a	MG274	pyruvate dehydrogenase E1-alpha subunit (pdhA); ACHLA
F11_orf402	MG272	dihydrolipoamide acetyltransferase component (E2) (pdhC); ACHLA
F11_orf457	MG271	dihydrolipoamide dehydrogenase (pdhD); BACST
• **Energy metabolism - Sugars [5]**		
D02_orf152	MG396	galactose-6-phosphate isomerase subunit (LacA); STRMU
D09_orf224	MG050	deoxyribose-phosphate aldolase (deoC); MYCPN
D09_orf554	MG053	phosphomannomutase (cpsG); MYCPI
E09_orf364	-	mannitol-1-phosphate 5-dehyrogenase (EC 1.1.1.17)(mtlD); STRMU

Table 1. *Continued*

• **Fatty acid and phospholipid metabolism [9]**		
A65_orf227	MG114	phosphatidylglycerophosphate synthase (pgsA); HAEIN
C09_orf600	-	carnitine palmitoyltransferase II precursor(cpt2); HUMAN
E30_orf395	MG437	CDP-diglyceride synthetase (cdsA); HAEIN
F11_orf84	MG287	(acyl carrier protein; STRGA)
G12_orf272V	MG344	triacylglycerol lipase (lip) 3; MYCMY
G12_orf328a	MG368	fatty acid/phospholipid synthesis protein (plsX); ECOLI
H08_orf289	MG310	triacylglycerol lipase (lip) 3; Mycoplasma sp
H10_orf266	MG212	1-acyl-sn-glycerol-3-phosphate acyltransferase (plsB); YEAST
P01_orf268	MG327	triacylglycerol lipase (lip) 2; MYCMY
• **Purines, pyrimidines, nucleosides and nucleotides - 2'-Deoyribonucleotide metabolism [3]**		
F10_orf328	MG227	thymidylate synthase (thyA); STAAU
F10_orf339	MG229	ribonucleotide reductase 2 (nrdF); SALTY
F10_orf721	MG231	ribonucleoside-diphosphate reductase (nrdE); SALTY
• **Purines, pyrimidines, nucleosides and nucleotides - Nucleotide and nucleoside interconversions [2]**		
C12_orf235	MG434	uridylate kinase (pyrH); ECOLI
H03_orf213	MG382	uridine kinase (udk); HAEIN
• **Purines, pyrimidines, nucleosides and nucleotides - Purine ribonucleotide biosynthesis [3]**		
D09_orf388	MG058	phosphoribosylpyrophosphate synthetase (prs); SYNP
GT9_orf215	MG171	adenylate kinase (adk); BACST
K04_orf239	MG107	5'guanylate kinase (gmk); HAEIN
• **Purines, pyrimidines, nucleosides and nucleotides - Salvage of nucleosides and nucleotides [9]**		
B01_orf178	MG030	uracil phosphoribosyltransferase (upp); STRSL
B01_orf191	MG034	thymidine kinase (tdk); BACSU
D09_orf133	MG052	cytidine deaminase (cdd); MYCPI
D09_orf238	MG049	purine-nucleoside phosphorylase (deoD); ECOLI
D09_orf421	MG051	thymidine phosphorylase (deoA); MYCPI
F11_orf133	MG276	adenine phosphoribosyltransferase (apt); HAEIN
K05_orf175	MG458	hypoxanthine-guanine phosphoribosyltransferase (HPT); LACLA
P01_orf217	MG330	cytidylate kinase (cmk); BACSU
D12_orf210	MG006	thymidylate kinase (CDC8) homolog, MYCGE
• **Purines, pyrimidines, nucleosides and nucleotides - Sugar-nucleotide biosynthesis and conversions [2]**		
A65_orf338	MG118	UDP-glucose 4-epimerase (galE); STRTR
K05_orf291	MG453	UDP-glucose pyrophosphorylase (gtaB); BACSU
• **Pyridine nucleotide synthesis [1]**		
H03_orf248	MG383	probable NH(3)-dependent NAD(+) synthetase (outB); BACSU
• **Regulatory function [8]**		
B01_orf362	MG024	hypothetical protein (yyaF) homolog; BACSU
C09_orf351	MG205	protein hrcA homolog, BACSU
D02_orf291	MG387	GTP-binding protein era homolog; STRMU
F11_orf733	MG278 (MG376)	stringent response protein SpoT; ECOLI
H03_orf433	MG384	GTP-binding protein (obg); BACSU
K04_orf726	MG104	virulence associated protein homolog (vacB); HAEIN
P01_orf193	MG335	hypothetical protein YihA (era like) homolog; ECOLI
P01_orf292	MG329	hypothetical protein HI0136 (era like) homolog; HAEIN
• **Replication - DNA replication, restriction, modification, recombination and repair [46]**		
A65_orf711	MG122	DNA topoisomerase I (topA); BACSU
A19_orf291	MG262	DNA polymerase I (polI, 5'-3' exonuclease) homolog; STRPN
A19_orf872	MG261	DNA polymerase III alpha subunit (dnaE); HAEIN
B01_orf1443	MG031	DNA polymerase III (dnaE) alpha chain (3'-5' exonuclease); BACSU
K05_orf380	MG001	DNA polymerase III beta subunit (dnaN); STAAU
D12_orf253	MG007	DNA polymerase III subunit delta' (holB); ECOLI
C12_orf681	MG420(C-Term:MG419)	DNA polymerase III subunit gamma and tau (dnaX); ECOLI
G07_orf473	MG094	replicative DNA helicase (dnaC); BACSU
H91_orf620	MG250	DNA primase (dnaG); BACSU
D12_orf212	MG010	DNA primase motif (dnaG); CLOAB
H91_orf658	MG254	DNA ligase (lig); ECOLI
G07_orf166	MG091	single-stranded DNA binding protein (ssb); HAEIN
K05_orf439	MG469	chromosomal replication initiator protein (dnaA); MYCCA
P02_orf336	MG339	recombination protein (recA); STAAU
C09_orf635	MG203	topoisomerase IV subunit B (parE), BACSU
C09_orf789	MG204	topoisomerase IV subunit A (parC), BACSU
K05_orf650	MG003	DNA gyrase subunit B (gyrB); MYCPN
K05_orf839o	MG004	DNA gyrase subunit A (gyrA); STAAU
G12_orf206	MG358	Holliday junction DNA helicase (ruvA); ECOLI
G12_orf307	MG359	Holliday junction DNA helicase (ruvB); HAEIN
H91_orf715	MG244	DNA helicase II (mutB1); HAEIN
H91_orf529	MG244	DNA helicase pcrA homolog; STAAU
F10_orf286	MG235	endonuclease IV (nfo); ECOLI
C12_orf948L	MG421	excinuclease ABC subunit A (uvrA); ECOLI
G07_orf657	MG073	excinuclease ABC subunit B (uvrB); ECOLI
C09_orf586L	MG206	excinuclease ABC subunit C (uvrC), BACSU
G12_orf412	MG360	UV protection protein (mucB); ECOLI
A19_orf277	MG(M2)	formamidopyrimidine-DNA glycosylase (fpg); BACFI
A65_orf306	-	PrrB homolog protein, ECOLI

Table 1. *Continued*

D09_orf383	MG047	S-adenosylmethionine synthetase 2 (metX); ECOLI
G07_orf240	MG097	uracil DNA glycosylase (ung); ECOLI
C12_orf249	-	restriction-modification enzyme subunit S1B (hsdS); MYCPU
GT9_orf238	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
GT9_orf319V	MG184	adenine-specific methyltransferase EcoRI (mte1); ECOLI
H03_orf191	MG380	glucose inhibited division protein (gidB); ECOLI
H03_orf612	MG379	glucose inhibited division protein (gidA); ECOLI
H10_orf145L	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
H10_orf187V	-	HsdS1B protein homolog; MYCPU
H91_orf206	-	Type 1 restriction enzyme (hsdR) homolog; ECOLI
H91_orf268	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
H91_orf330	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
H91_orf376	-	Type 1 restriction enzyme (hsdR) homolog; ECOLI
H91_orf543	-	type I restriction enzyme (hsdM); ECOLI
P02_orf363V	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
R02_orf335	-	type I restriction enzyme ecokI specificity protein (hsdS) homolog; HAEIN
E30_orf375	MG438	MG438 homolog, MYCGE
• **Transcription - Degradation of RNA [2]**		
G12_orf282a	MG367	ribonuclease III (rnc); ECOLI
K05_orf118V	MG465	RNaseP C5 chain (rnpA); MYCCA
• **Transcription - RNA synthesis, modification and DNA transcription [11]**		
GT9_orf327	MG177	RNA polymerase alpha core subunit (rpoA); BACSU
G12_orf1391o	MG341	RNA polymerase beta subunit (rpoB); BACSU
P04_orf1290	MG340	DNA-directed RNA polymerase beta' chain (rpoC); THEMA
B01_orf146	MG022	DNA-directed RNA polymerase delta subunit (rpoE); BACSU
H91_orf499	MG249	RNA polymerase sigma-A factor (sigA); BACSU
F11_orf160	MG282	transcription elongation factor (greA); RICPR
D09_orf320	MG054	transcription antitermination factor (nusG); BACSU
E07_orf540o	MG141	N-utilization substance protein A homolog (nusA); BACSU
C12_orf450	MG425	ATP-dependent RNA helicase (deaD); HAEIN
H08_orf409	MG308	ATP-dependent RNA helicase (deaD); ECOLI
D12_orf1030	MG018	hypothetical helicase Yb95 homolog; YEAST
• **Translation - Amino acyl tRNA synthetases and tRNA modification [24]**		
A05_orf900	MG292	alanyl-tRNA synthetase (alaS); ECOLI
H03_orf537	MG378	arginyl-tRNA synthetase (argS); BRELA
K04_orf455o	MG113	asparaginyl-tRNA synthetase (asnS); ECOLI
D09_orf557	MG036	aspartyl-tRNA synthetase (aspS); THEAQ
H91_orf437	MG253	cysteinyl-tRNA synthetase (cysS); BACSU
K05_orf484	MG462	glutamyl-tRNA synthetase (gltX); BACST
H91_orf449	MG251	glycyl-tRNA synthetase (grs1); YEAST
B01_orf414o	MG035	histidyl-tRNA synthetase (hisS); STREQ
G12_orf861	MG345	isoleucine-tRNA ligase (ileS); STAAU
F11_orf793o	MG266	leucyl-tRNA synthetase (leuS); BACSU
A65_orf489	MG136	lysyl-tRNA synthetase (lysS); BACSU
G12_orf311	MG365	methionyl-tRNA formyltransferase (fmt); ECOLI
B01_orf512	MG021	methionyl-tRNA synthetase (metS); BACST
G07_orf188	MG083	peptidyl-tRNA hydrolase homolog (pth); HAEIN
C09_orf341	MG194	phenylalanyl-tRNA synthetase alpha-subunit (pheS); BACSU
C09_orf805	MG195	phenylalanyl-tRNA synthetase beta chain (pheT); BACSU
GT9_orf243V	MG182	pseudouridylate synthase I (hisT); ECOLI
F11_orf483	MG283	putative prolyl-tRNA synthetase (YHI0; proS); YEAST
D12_orf420	MG005	seryl-tRNA synthetase (serS); BACSU
G12_orf564	MG375	threonyl-tRNA synthetase (thrSv); BACSU
K05_orf210	MG445	tRNA (guanine-N1)-methyltransferase (trmD); HUMAN
A65_orf346	MG126	tryptophanyl-tRNA synthetase (trpS); HAEIN
K05_orf399	MG455	tyrosyl tRNA synthetase (tyrS); BACCA
P01_orf838	MG334	valyl-tRNA synthetase (valS); BACST
• **Translation - Degradation of proteins, peptides and glycopeptides [8]**		
B01_orf309	MG020	proline iminopeptidase (pip); NEIGO
D02_orf445	MG391	nonspecified aminopeptidase; MYCSA
D09_orf319	MG046	o-sialoglycoprotein endopeptidase (gcp); PASHA
F10_orf795	MG239	ATP-dependent protease (lon); BACSU
G12_orf715	MG355	ATP-dependent protease binding subunit (clpB) homolog; HAEIN
GT9_orf611	MG183	oligoendopeptidase F (pepF); LACLA
H03_orf193o	MG377	MG377 homolog (put. zinc protease), MYCGE
P01_orf354	MG324	X-Pro dipeptidase (pepX); LACDE
• **Translation - Protein modification and translation factors [15]**		
GT9_orf78	MG173	initiation factor 1 (infA); BACSU
VXpSPT7_orf617	MG142	protein synthesis initiation factor 2 (infB); BACST
C09_orf201	MG196	translation initiation factor IF3 (infC); MYCPE
G07_orf688	MG089	elongation factor G (fus); THEAQ
B01_orf190	MG026	elongation factor P (efp) homolog; HAEIN
C12_orf298	MG433	elongation factor Ts (tsf); SPICI
K05_orf394	MG451	elongation factor TU (tuf); MYCGE
H91_orf359V	MG258	peptide chain release factor 1 (RF1; prfA);BACSU
E30_orf184	MG435	ribosome releasing factor (frr); HAEIN
GT9_orf248	MG172	methionine amino peptidase (map); BACSU
K04_orf216	MG106	polypeptide deformylase (def); HAEIN
K04_orf259	MG108	protein phoshatase 2C homolog; YEAST

Table 1. *Continued*

K04_orf389	MG109	probable protein serine/threonine kinase; CAEEL
K05_orf151	MG448	pilB homolog (fragment); HAEIN
C12_orf157	MG408	peptide methionine sulfoxide reductase (pmsR), ECOLI
• **Translation - Ribosomal proteins: synthesis and modification [53]**		
G07_orf226	MG082	ribosomal protein L1 (rpL1); BACST
VXpSPT7_orf287a	MG154	ribosomal protein L2 (rpL2); MYCCA
VXpSPT7_orf287b	MG151	ribosomal protein L3 (rpL3); MYCCA
VXpSPT7_orf212	MG152	ribosomal protein L4 (rpL4); MYCCA
GT9_orf180b	MG163	ribosomal protein L5 (rpL5); HAEIN
GT9_orf184	MG166	ribosomal protein L6 (rpL6); MYCCA
G12_orf122	MG362	ribosomal protein L7/L12 ('A' type) (rpL7/L12); MICLU
G07_orf149	MG093	ribosomal protein L9 (rpL9); BACST
G12_orf161	MG361	ribosomal protein L10 (rpL10); THEMA
G07_orf137	MG081	ribosomal protein L11 (RPL11); THEMA
C12_orf146	MG418	ribosomal protein L13 (rpL13); ECOLI
GT9_orf122	MG161	ribosomal protein L14 (rpL14); BACST
GT9_orf151	MG169	ribosomal protein L15 (rpL15); MYCCA
VXpSPT7_orf139o	MG158	ribosomal protein L16 (rpL16); MYCCA
GT9_orf124a	MG178	ribosomal protein L17 (rpL17); BACSU
GT9_orf116b	MG167	ribosomal protein L18 (rpL18); BACST
K05_orf119	MG444	ribosomal protein L19 (rpL19); BACST
C09_orf127	MG198	ribosomal protein L20 (rpL20); MYCFE
F10_orf100b	MG232	ribosomal protein L21 (rpL21); BACSU
VXpSPT7_orf184	MG156	ribosomal protein L22 (rpL22); HAEIN
VXpSPT7_orf237	MG153	ribosomal protein L23 (rpL23); THEMA
GT9_orf111a	MG162	ribosomal protein L24 (rpL24); BACST
F10_orf104	MG234	ribosomal protein L27 (rpL27); BACSU
C12_orf65	MG426	ribosomal protein L28 (rpL28); BACSU
GT9_orf111b	MG159	ribosomal protein L29 (rpL29); THEMA
H91_orf97	MG257	ribosomal protein L31 (rpL31); ECOLI
G12_orf57	MG363	ribosomal protein L32 (rpL32); HAEIN
P01_orf53	MG325	ribosomal protein L33 (rpL33); BACST
K05_orf48	MG466	ribosomal protein L34 (rpL34); PROMI
C09_orf59	MG197	ribosomal protein L35 (rpL35); BACST
GT9_orf37	MG174	ribosomal protein L36 (rpL36); CHLTR
G07_orf294	MG070	ribosomal protein S2 (rpS2); SPIPL
VXpSPT7_orf273	MG157	ribosomal protein S3 (rpS3); MYCCA
H08_orf205	MG311	ribosomal protein S4 (rpS4); BACSU
GT9_orf219	MG168	ribosomal protein S5 (rpS5); BACSU
G07_orf215	MG090	ribosomal protein S6 (rpS6); ECOLI
G07_orf155	MG088	ribosomal protein S7 (rpS7); BACST
GT9_orf142	MG165	ribosomal protein S8 (rpS8); MYCCA
C12_orf132	MG417	ribosomal protein S9 (rpS9); BACST
VXpSPT7_orf108	MG150	ribosomal protein S10 (rpS10); THEMA
GT9_orf121	MG176	ribosomal protein S11 (rpS11); BACST
G07_orf139	MG087	ribosomal protein S12 (rpS12); BACST
GT9_orf124b	MG175	ribosomal protein S13 (rpS13); BACSU
GT9_orf61	MG164	ribosomal protein S14 (rpS14); MYCCA
C12_orf86	MG424	ribosomal protein S15 (BS18); BACST
K05_orf88	MG446	ribosomal protein S16 (BS17); BACSU
GT9_orf85	MG160	ribosomal protein S17 (rpS17); MYCCA
G07_orf104b	MG092	ribosomal protein S18 (rpS18); ECOLI
VXpSPT7_orf87	MG155	ribosomal protein S19 (rpS19); MYCBO
G12_orf87	MG(M3)	ribosomal protein S20 (rpsT); ECOLI
D12_orf288	MG012	ribosomal protein S6 modification protein (rimK); ECOLI
H91_orf242a	MG252	hypothetical protein YacO (rRNA methylase) homolog; BACSU
VXpSPT7_orf116	MG143	ribosome binding factor A homolog (rbfA); ECOLI
• **Transport and binding proteins - ABC transport [34]**		
A05_orf382	MG303	abc transport ATP-binding protein (artP); ECOLI
D09_orf286a	MG044	spermidine/putrescine transport system permease (potI); ECOLI
D09_orf286b	MG043	spermidine/putrescine transport system permease (potB); HAEIN
D09_orf560L	MG042	spermidine/putrescine transport ATP-binding prot (potA); ECOLI
F10_orf491	MG225	hypothetical protein (gi: 710640) homolog (put. amino acid permease); CLOPE
F10_orf503	MG226	general amino acid permease GAP1 homolog; YEAST
G07_orf376	MG078	oligopeptide transport system permease protein (amiD); STRPN
G07_orf389a	MG077	oligopeptide transport system permease protein (oppB); BACSU
G07_orf423	MG079	oligopeptide transport ATP-binding protein (oppD); BACSU
G07_orf851	MG080	oligopeptide transport ATP-binding protein (oppF); BACSU
GT9_orf303	MG180	histidine transport ATP-binding protein (hisP); ECOLI
R02_orf465	MG065	glutamine transport ATP-binding protein (glnQ); ECOLI
C12_orf225	MG409	phosphate transport system regulatory protein (phoU); ECOLI
C12_orf329	MG410	phosphate transport ATP-binding protein (pstB); ECOLI
C12_orf651V	MG411	phosphate transport system permease protein (pstA); ECOLI
GT9_orf274	MG179	sulfate transport ATP-binding protein (cysA); SYNP
K05_orf284	MG065 (MG467)	sulfate transport ATP-binding protein (cysA); SYNP
A65_orf311	MG121	high affinity ribose transport protein (rbsC); HAEIN
A65_orf572	MG119	hypothetical ABC transporter (yjcW) homolog; ECOLI
E07_orf319	MG189	sn-glycerol-3-phosphate transport system permease protein (ugpE); ECOLI
E07_orf329	MG188	sn-glycerol-3-phosphate transport system permease protein (ugpA); ECOLI
E07_orf586	MG187	sn-glycerol-3-phosphate transport system permease protein (ugpC); ECOLI
A05_orf270L	MG304	abc transport ATP-binding protein (cbiO), SALTY
G07_orf872V	MG071	MG(2+) transport ATPase, P-type 1 (mgtA); ECOLI

Table 1. *Continued*

A05_orf244	MG290	ATP-binding protein P29; MYCHR
A05_orf380V	MG289	high affinity transport system protein P37; MYCHR
A05_orf542	MG291	transport system permease protein P69; MYCHR
D02_orf660	MG390	lactococcin transport ATP-binding protein (lcnDR3); LACLA
D12_orf623	MG014	transport ATP-binding protein (pmd1); SCHPO
D12_orf634	MG015	transport ATP-binding protein (msbA); HAEIN
F10_orf326	MG179	bcrA homolog protein; BACLI
F10_orf750	-	putative ABC transport permease
H08_orf565	MG322	Na(+) translocating ATPase subunit J (ntpJ); ENTHR
K05_orf339	MG467	devA protein homolog; ANASP
• **Transport and binding proteins - PTS transport [7]**		
E09_orf143V	-	PTS system mannitol-specific component IIA (EIIA-MTL)(mtlF); STRMU
E09_orf379	-	PTS system mannitol-specific component IIA (EIIA-MTL)(mtlA); STACA
R02_orf694	MG062	fructose-permease IIBC component (fruA); ECOLI
GT9_orf940o	MG069	PTS system, glucose-specific IIABC component (EIIABC-GLC); BACSU
D09_orf88	MG041	phosphocarrier protein HPr (ptsH); MYCCA
P02_orf159	-	hypothetical phosphotransferase protein YjfU homolog; ECOLI
C12_orf572	MG429	PEP-dependent HPr protein kinase phosphoryltransferase (Enzyme I) (ptsI); STRSL
• **Transport and binding proteins - Other transport systems [3]**		
B01_orf264	MG033	glycerol uptake facilitator (glpF); BACSU
R02_orf564o	MG061	hexosephosphate transport protein (uhpT); SALTY
A05_orf475	MG294	MG294 homolog(put. permease), MYCGE
• **Other categories - Adaptations and atypical conditions [3]**		
K05_orf140	MG454	osmotically inducible protein (osmC); ECOLI
K05_orf270	MG470	soj homolog protein; BACSU
K05_orf263V	MG463	S-adenosylmethionine-6-N',N'-adenosyl (rRNA) dimethytransferase (ksgA); ECOLI
• **Other categories - Other [188]**		
A05_orf102	-	hypothetical 13.2 KD protein homolog (ylxM); BACSU
A05_orf129	MG296	MG296 homolog, MYCGE
A05_orf290	(MG125)	hypothetical protein (YidA) homolog; ECOLI
A05_orf317	MG302	MG302 homolog, MYCGE
A05_orf370	MG295	hypothetical protein (HI0174); HAEIN
A05_orf395	MG306	MG306 homolog, MYCGE
A05_orf982	MG298	P115 protein homolog (SGC3); MYCHR
A19_orf200	MG264	hypothetical protein (HI0890) homolog; HAEIN
A19_orf282	MG265	hypothetical protein (YidA) homolog; ECOLI
A19_orf292	MG263	hypothetical protein (YidA) homolog; ECOLI
A65_orf100	MG134	hypothetical protein YaaK homolog; BACSU
A65_orf117	MG129	MG129 homolog, MYCGE
A65_orf144	MG132	hypothetical protein Hit1 homolog; YEAST
A65_orf145	MG127	hypothetical protein Ygl1 homolog; STRVR
A65_orf166	MG260 (MG185)	MG260 homolog, MYCGE
A65_orf223	MG117	MG117 homolog, MYCGE
A65_orf251b	MG116	MG116 homolog, MYCGE
A65_orf259	MG128	hypothetical protein HI0072 homolog; HAEIN
A65_orf266	MG133	MG133 homolog, MYCGE
A65_orf281	MG125	hypothetical protein (gi: 973220) homolog; ECOLI
A65_orf285	MG135	MG135 homolog, MYCGE
A65_orf377	MG260 (MG185)	MG260 homolog, MYCGE
A65_orf475	MG123	MG123 homolog, MYCGE
A65_orf493	MG130	hypothetical protein Ysr1 homolog; MYCMY
A65_orf517	MG120	MG120 homolog, MYCGE
A65_orf569	MG139	MG139 homolog, MYCGE
B01_orf108	MG029	hypothetical protein (gi: 606093) homolog; ECOLI
B01_orf168	MG027	MG027 homolog, MYCGE
B01_orf186L	MG032	MG032 homolog, MYCGE
B01_orf203	MG028	MG028 homolog, MYCGE
B01_orf338	MG032	MG032 homolog, MYCGE
B01_orf666	MG032	MG032 homolog, MYCGE
B01_orf672	MG032	MG032 homolog, MYCGE
B01_orf673	MG032	MG032 homolog, MYCGE
C09_orf104	MG191	(MG191 homolog, MYCGE)
C09_orf121	MG202	MG202 homolog, MYCGE
C09_orf143b	MG199	MG199 homolog, MYCGE
C09_orf159	MG207	MG207 homolog, MYCGE
C12_orf141	MG427	MG427 homolog, MYCGE
C12_orf172	MG428	MG428 homolog, MYCGE
C12_orf334	MG413 (MG414)	MG413 homolog, MYCGE
C12_orf344	MG415	MG415 homolog, MYCGE
C12_orf385	MG412	MG412 homolog, MYCGE
C12_orf404	MG432	hypothetical protein (yflB) homolog; SPICI
C12_orf561	MG423	MG423 homolog, MYCGE
C12_orf839	MG422	MG422 homolog, MYCGE
C12_orf997	MG414	MG414 homolog, MYCGE
D02_orf108	MG388	MG388 homolog, MYCGE
D02_orf129	MG389	MG389 homolog, MYCGE
D02_orf135L	MG067 (MG395, MG068)	MG067 homolog, MYCGE
D02_orf140	MG395 (MG068)	MG395 homolog, MYCGE

Table 1. *Continued*

D02_orf150	MG068 (MG395)	MG068 homolog, MYCGE
D02_orf157L	MG395 (MG068)	MG395 homolog, MYCGE
D02_orf225L	MG068 (MG067, MG395)	MG068 homolog, MYCGE
D02_orf265V	MG068 (MG395, MG067)	MG068 homolog, MYCGE
D02_orf346	MG068 (MG395)	MG068 homolog, MYCGE
D02_orf347	MG067 (MG395, MG068)	MG067 homolog, MYCGE
D02_orf353V	MG068 (MG395)	MG068 homolog, MYCGE
D02_orf569	MG397	MG397 homolog, MYCGE
D09_orf125	MG055	MG055 homolog, MYCGE
D09_orf147	MG059	hypothetical protein A43259 homolog; ENTHR
D09_orf178	MG057	hypothetical protein YabF homolog; BACSU
D09_orf276	MG056	hypothetical protein YabC homolog; BACSU
D09_orf451	MG037	pre-B cell enhancing factor homolog (pbeF); HUMAN
D09_orf518	MG096	MG096 homolog, MYCGE
D09_orf632	MG288 (MG096)	MG288 homolog, MYCGE
D12_orf261	MG009	hypothetical protein yabD homolog; BACSU
D12_orf285	MG011	MG011 homolog, MYCGE
E07_orf1113	MG140	MG140 homolog, MYCGE
E07_orf265	MG260 (MG185)	MG260 homolog, MYCGE
E07_orf324	MG190	hypothetical 28K protein (orf4, P1 operon); MYCPN
E07_orf485	MG260 (MG185)	MG260 homolog, MYCGE
E09_orf136	MG441	MG441 homolog, MYCGE
E09_orf204o	-	protein P30, MYCPN
E09_orf287o	MG439	MG439 homolog, MYCGE
E09_orf302	MG440	MG440 homolog, MYCGE
F04_orf154	MG288 (MG096)	MG288 homolog, MYCGE
F04_orf260V	MG288	MG288 homolog, MYCGE
F10_orf100a	MG233	hypothetical protein YsxB homolog; BACSU
F10_orf141b	MG221	hypothetical protein YabB homolog; ECOLI
F10_orf153	MG230	MG230 homolog, MYCGE
F10_orf158	MG236	MG236 homolog, MYCGE
F10_orf291	MG240	MG240 homolog, MYCGE
F10_orf294	MG237	MG237 homolog, MYCGE
F10_orf308	MG222	hypothetical protein YabC homolog; ECOLI
F10_orf419	MG223	MG223 homolog, MYCGE
F10_orf621	MG241	MG241 homolog, MYCGE
F10_orf632o	MG242	MG242 homolog, MYCGE
F10_orf90	MG220	MG220 homolog, MYCGE
F11_orf114	MG267	MG267 homolog, MYCGE
F11_orf122a	MG284	MG284 homolog, MYCGE
F11_orf197	MG286	MG286 homolog, MYCGE
F11_orf218	MG279	MG279 homolog, MYCGE
F11_orf229	MG268	hypothetical protein YaaF homolog; BACSU
F11_orf287	MG280	MG280 homolog, MYCGE
F11_orf346	MG285	MG285 homolog, MYCGE
F11_orf358b	MG269	MG269 homolog, MYCGE
F11_orf582	MG281	MG281 homolog, MYCGE
F11_orf887	MG277	MG277 homolog, MYCGE
G07_orf1030	MG075	protein P100; MYCPN
G07_orf135	MG074	MG074 homolog, MYCGE
G07_orf138	MG076	MG076 homolog, MYCGE
G07_orf289	MG084	hypothetical protein (yacA) homolog; BACSU
G07_orf312	MG085	MG085 homolog, MYCGE
G07_orf417	MG288 (MG096)	MG288 homolog, MYCGE
G07_orf478o	MG100	PET112 protein homolog; YEAST
G07_orf478V	MG099	amidase homolog (S47454); YEAST
G07_orf479	MG098	MG098 homolog, MYCGE
G12_orf104	MG376	MG376 homolog, MYCGE
G12_orf109	MG353	MG353 homolog, MYCGE
G12_orf136	MG354	MG354 homolog, MYCGE
G12_orf166a	MG342	MG342 homolog, MYCGE
G12_orf166b	MG346	hypothetical protein Ygl3 homolog; BACST
G12_orf210V	MG347	hypothetical protein HI0340 homolog; HAEIN
G12_orf218	MG364	MG364 homolog, MYCGE
G12_orf269	MG374	MG374 homolog, MYCGE
G12_orf281	MG373	MG373 homolog, MYCGE
G12_orf325	MG371	hypothetical 28K protein (P1 operon) homolog; MYCPN
G12_orf326	MG370	hypothetical protein (HI0176) homolog; HAEIN
G12_orf328b	MG350	MG350 homolog, MYCGE
G12_orf348	MG343	MG343 homolog, MYCGE
G12_orf387	MG372	MG372 homolog, MYCGE
G12_orf413	MG349	MG349 homolog, MYCGE
G12_orf558	MG369	MG369 homolog, MYCGE
G12_orf664	MG366	MG366 homolog, MYCGE
GT9_orf148	MG260	MG260 homolog, MYCGE
GT9_orf434	MG181	MG181 homolog, MYCGE
H03_orf235	MG381	MG381 homolog, MYCGE
H08_orf157b	MG321	MG321 homolog, MYCGE
H08_orf193	MG319	MG319 homolog, MYCGE
H08_orf231	MG323	hypothetical protein YZAC homolog; BACSU
H08_orf263	MG313	MG313 homolog, MYCGE
H08_orf287	MG320	(cytochrome C oxidase polypeptide I (CtaD); BACSU)
H08_orf314	MG315	MG315 homolog, MYCGE
H08_orf345	MG307	MG307 homolog, MYCGE

Table 1. *Continued*

H08_orf369	MG316	(competence locus E (comE3); BACSU)
H08_orf448	MG314	MG314 homolog, MYCGE
H08_orf572o	MG307	MG307 homolog, MYCGE
H08_orf591	MG321	MG321 homolog, MYCGE
H08_orf726	MG307	MG307 homolog, MYCGE
H10_orf149	MG211	MG211 homolog, MYCGE
H10_orf196	MG208	MG208 homolog, MYCGE
H10_orf208	MG214	hypothetical protein P35155 homolog; BACSU
H10_orf309	MG209	hypothetical protein YceC homolog; ECOLI
H91_orf213	MG248	MG248 homolog; MYCGE
H91_orf224	MG243	MG243 homolog, MYCGE
H91_orf239	MG247	hypothetical protein YgiH homolog; ECOLI
H91_orf258	MG256	MG256 homolog, MYCGE
H91_orf281	MG246	MG246 homolog, MYCGE
H91_orf534	MG255	MG255 homolog, MYCGE
H91_orf677	MG260	MG260 homolog, MYCGE
K04_orf202	MG105	MG105 homolog, MYCGE
K04_orf222	MG101	MG101 homolog, MYCGE
K04_orf278L	MG110	hypothetical protein YjeQ homolog; ECOLI
K04_orf280	MG103	MG103 homolog, MYCGE
K05_orf169	MG459	hypothetical protein HI0671 homolog; HAEIN
K05_orf234	MG449	MG449 homolog, MYCGE
K05_orf237	MG450	degV homolog protein; BACSU
K05_orf251	MG452	MG452 homolog, MYCGE
K05_orf271	MG442	MG442 homolog, MYCGE
K05_orf345	MG456	MG456 homolog, MYCGE
K05_orf385	MG464	hypothetical protein 1 (S42122); MYCCA
K05_orf401	MG443	hypothetical protein (P27712); SPICI
K05_orf425	MG461	MG461 homolog, MYCGE
K05_orf499	MG447	MG447 homolog, MYCGE
P01_orf1033	MG328	MG328 homolog, MYCGE
P01_orf197	MG333	hypothetical protefn HI1366 homolog; HAEIN
P01_orf209	MG331	MG331 homolog, MYCGE
P01_orf235	MG332	hypothetical protein HI0315 homolog; HAEIN
P01_orf293	MG326	degV homolog protein; BACSU
P01_orf341	marginal MG025	hypothetical protein YibD homolog; ECOLI
P02_orf140	MG337	MG337 homolog, MYCGE
P02_orf218	-	hypothetical protein YjfV homolog; ECOLI
P02_orf305	-	hypothetical protein YjfW homolog; ECOLI
P02_orf316	MG338	MG338 homolog, MYCGE
P02_orf408	MG336	nitrogen fixation protein (nifS); HAEIN
P02_orf427	MG288 (MG096)	MG288 homolog, MYCGE
P02_orf458	MG096 (MG288)	MG096 homolog, MYCGE
P02_orf509	MG288 (MG096)	MG288 homolog, MYCGE
P02_orf660	-	hypothetical protein YjfS homolog; ECOLI
R02_orf1386V	MG064	MG064 homolog, MYCGE
R02_orf147	MG260	MG260 homolog, MYCGE
R02_orf469	MG061	MG061 homolog, MYCGE
R02_orf524	MG068 (MG067)	MG068 homolog, MYCGE
VXpSPT7_orf269	MG145	hypothetical protein (YaaC) homolog; PSEFL
VXpSPT7_orf377	MG147	MG147 homolog, MYCGE
VXpSPT7_orf402	MG144	MG144 homolog, MYCGE
VXpSPT7_orf445	MG148	MG148 homolog, MYCGE
• **no classification so far [86]**		
A19_orf1140	-	-
A19_orf129	-	-
A19_orf204	-	-
A19_orf229V	-	-
A19_orf591	-	-
A65_orf115	-	-
A65_orf118	-	-
B01_orf103b	-	-
B01_orf116L	-	-
B01_orf147	-	-
b01_orf182l	-	-
B01_orf274	-	-
C09_orf130b	-	-
C09_orf140o	-	-
C09_orf165	-	-
C09_orf172	-	-
C09_orf223	-	-
C09_orf251	-	-
C09_orf404	-	-
C09_orf422	-	-
C09_orf718	-	-
C12_orf181o	-	-
C12_orf247	-	-
D02_orf100	-	-
D02_orf109	-	-
D02_orf122a	-	-
D02_orf122b	-	-
D02_orf128	-	-
D09_orf127a	-	-

Table 1. *Continued*

D12_orf131	-	-
D12_orf235	-	-
D12_orf257	-	-
E07_orf133	-	-
E07_orf140	-	-
E07_orf163	-	-
E07_orf166	-	-
E07_orf175	-	-
E07_orf179	-	-
E07_orf228	-	-
E09_orf136L	marginal MG440	-
E30_orf352	-	-
F04_orf120	-	-
F04_orf150	-	-
F10_orf218	-	-
F10_orf357	marginal MG011	-
F10_orf565	-	-
F10_orf741	-	-
F11_orf148o	-	-
F11_orf879	-	-
G12_orf140b	-	-
G12_orf168	-	-
G12_orf225	-	-
GT9_orf113	-	-
H03_orf152	-	-
H08_orf102	-	-
H10_orf119	-	-
H10_orf206	-	-
H10_orf220L	-	-
H91_orf115	-	-
H91_orf180	-	-
H91_orf216	-	-
K05_orf101a	-	-
K05_orf106	-	-
K05_orf1882	marginal MG064	-
K05_orf250	-	-
P01_orf140	-	-
P01_orf199	-	-
P01_orf243	-	-
P02_orf103b	-	-
P02_orf126	-	-
P02_orf143	-	-
P02_orf147	-	-
P02_orf163	-	-
P02_orf196	-	-
P02_orf253	-	-
P02_orf474	-	-
R02_orf101	-	-
R02_orf105	-	-
R02_orf140	-	-
R02_orf150	-	-
R02_orf183o	-	-
R02_orf254	-	-
R02_orf264	-	-
R02_orf329	-	-
R02_orf440	-	-
VXpSPT7_orf112	-	-
• **hypothetical ORFs derived from repetitve DNA elements [46]**		
A05_orf139	-	-
A19_orf211	-	-
A65_orf115	-	-
B01_orf147	-	-
C09_orf140o	-	-
C09_orf149a	-	-
E07_orf163	-	-
F11_orf148o	-	-
G12_orf168	-	-
H08_orf157a	marginal MG321	-
H91_orf180	-	-
P01_orf199	-	-
P02_orf103b	-	-
P02_orf196	-	-
R02_orf138	-	-
R02_orf140	-	-
R02_orf183o	-	-
C09_orf149b	-	adhesin P1 (group 2) homolog; MYCPN
H08_orf329V	MG321	adhesin P1 (group 2) homolog; MYCPN
A65_orf465V	MG191	adhesin P1 (group 2) homolog; MYCPN
E07_orf413	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
E07_orf256L	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
A05_orf278	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
H08_orf270	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
P02_orf422V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN

Table 1. *Continued*

P02_orf422V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
P02_orf527V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
F11_orf533L	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
P01_orf208V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
GT9_orf438V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
GT9_orf127	-	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
GT9_orf313	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
C09_orf428V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
A19_orf737V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
E07_orf221V	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
R02_orf347L	MG191	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
G12_orf325	MG371	hypothetical 28K protein (P1 operon) homolog; MYCPN
E07_orf224	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
E07_orf434	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
C09_orf272	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
A05_orf493	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
R02_orf301	-	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
R02_orf173	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
H08_orf445	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
P02_orf381	(MG192)	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
H91_orf322	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
H91_orf272	MG192	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN

- **RNA - rRNA [3]**

5S rRNA
16S rRNA
23S rRNA

- **RNA - tRNA [33 tRNAs in 14 genes/operons]**

Arg-tRNA gene (CGA); MYCPN
Arg-tRNA gene (CGC); MYCPN
Arg-tRNAgene (AGA); MYCPN
Asn-tRNA(AAC), Glu-tRNA(GAA), Thr-tRNA(ACG), Val-tRNA(GTA), Thr-tRNA(ACA), Lys-tRNA(AAG), Leu-tRNA(CTA) genes; MYCPN
Cys-tRNA(TGC), Pro-tRNA(CCA), Met-tRNA(ATG), Ile-tRNA(ATG), Ser-tRNA(TCA), fMet-tRNA(ATG), Asp-tRNA(GAC) and Phe-tRNA(TTC) genes; MYCPN
Gly-tRNA(GGC) gene; MYCPN
His-tRNA(CAC) gene; MYCPN
Ile-tRNA(ATC), Ala-tRNA(GCA) genes; MYCPN
Thr-tRNA(GGU) gene; MYCPN
Ser-tRNA (AGC) gene; MYCPN
Ser-tRNA(TCC), Ser-tRNA(TCG) genes; MYCPN
Trp-tRNA (TGA)gene; MYCPN
Trp-tRNA(TGG) gene; MYCPN
Tyr-tRNA (TAC), Glu-tRNA (CAA), Lys-tRNA (AAA), Leu-tRNA (TTA), Gly-tRNA (GGA) genes; MYCPN

- **RNA - other [3]**

4.5S RNA; MYCPN
10sa RNA; MYCGE
RNaseP RNA; MYCGE

MG is the name of the corresponding ORF in *M.genitalium* (9).

coding densities have been also estimated for the smaller *M.genitalium* genome (9) and for the genome of *Haemophilus influenzae* which is more than twice as large (30). The length of the proposed proteins in *M.pneumoniae* ranges from 37 (4.3 kDa) to 1882 (209.4 kDa) amino acids (Fig. 3). One of the largest proteins is the cytadherence accessory protein HMW2 (F10_orf1818) and the smallest identified protein is the 37 amino acid ribosomal protein L36 (GT9_orf37). For practical reasons we introduced at the beginning of the sequence analysis a cut-off point of 100 amino acids for proposed proteins unless we found smaller proteins such as some of the ribosomal proteins during the initial BLASTX homology search. All intergenic or non coding regions were reanalyzed with a cut-off point of 50 amino acids and searches were done for specific small proteins. However, we cannot exclude the possibility that some of the smaller proteins, not showing similarities to known proteins from other organisms, have been missed in our analysis.

The codon usage of *M.pneumoniae* is summarized in Table 3. We compared it for all proposed genes, for the subsets of genes with a low G+C (content below 35 mol%) and high G+C content (between 50 and 56 mol%) and for all 50 ribosomal protein genes (42.8 mol%) as an example for frequently translated genes. Codon usage of the low and high G+C content subfractions is clearly influenced by the DNA composition, favouring either codons with G/C or A/T at the third position. The codon usage pattern differs also for the complete genome and for genes which are frequently expressed like the ones coding for ribosomal proteins.

The most frequently used codons are AUU (Ile, 4.6%); AAA (Lys, 4.6%); UUU (Phe, 4.3%); GAA (Glu, 4.2%) and UUA (Leu, 3.9%) and the most common amino acids are Leu (10.3%), Lys (8.5%), Ile (6.6%), Ala (6.6%) and Val (6.5%). The high value for Lys is in agreement with the relative high percentage of proposed proteins with calculated isoelectric points between pH 9 and 12 (Fig. 4). The least frequently used codons are UGC (Cys, 0.2%); CGA (Arg, 0.25%); AGG (Arg, 0.29%); AGA (Arg, 0.4%) and UGU (Cys, 0.55%).

All *M.pneumoniae* gene products were classified (Table 1 and 2), with some minor modifications, in accordance with criteria introduced for *Escherichia coli* (31) and adapted for the classification of putative genes from *H.influenzae*. We added

Table 2. Summary of the functional classification of the ORFs

• Biosynthesis of cofactors, prosthetic groups and carrier	**8**
Folic acid	5
Heme and porphyrin	1
Thioredoxin	2
• Cell envelope	**54**
Membranes, lipoproteins and porines	42
Surface structures and cytadherence	8
Surfaces polysaccharides, lipopolysaccharides and antigens	4
• Cellular processes	**20**
Cell division	2
Cell killing	1
Chaperones	7
Detoxification	1
Protein and peptide secretion	9
• Central intermediary metabolism	**6**
Other	5
Phosphorous compounds	1
• Energy metabolism	**39**
Aerobic	3
Amino acids and amines	5
Anaerobic	1
ATP-proton motive force interconversion	9
Glycolysis	10
Pentose Phosphate pathway	2
Pyruvate DHase	4
Sugars	5
• Fatty acid and phospholipid metabolism	**9**
• Purines, pyrimidines, nucleosides and nucleotides	**18**
2'-Deoyribonucleotide metabolism	3
Nucleotide and nucleoside interconversions	2
Purine ribonucleotide biosynthesis	3
Salvage of nucleosides and nucleotides	8
Sugar-nucleotide biosynthesis and conversions	2
• Pyridine nucleotide metabolism	**1**
• Regulatory function	**8**
• Replication	**46**
DNA replication, restriction, modification, recombination and repair	46
• Transcription	**13**
Degradation of RNA	2
RNA synthesis, modification and DNA transcription	11
• Translation	**99**
Amino acyl tRNA synthetases and tRNA modification	24
Degradation of proteins, peptides and glycopeptides	8
Protein modification and translation factors	15
Ribosomal proteins: synthesis and modification	52
• Transport and binding proteins	**44**
ABC transport	34
PTS transport	7
Other transport systems	3
• Other categories	**191**
Adaptations and atypical conditions	3
Other	188
• hypothetical ORFs derived from repetitve DNA elements	**46**
• no classification so far	**86**
• RNA	**39**
rRNA	**3**
tRNA	**33**
other	**3**

'cytadherence associated proteins' to the category of cell envelope–surface structures, since evidence is mounting, that *M.pneumoniae* possesses a cytoskeleton-like organization which stabilizes the bacterium and protects it against osmotic lysis (2). The category of transport and binding proteins was altered by subdivision into three groups namely, into PTS-, ABC- and other transport systems. To facilitate the orientation on the gene map we added a list which contains all proposed ORFs and RNAs in nummerical order (Table 4).

More details on this very general analysis will be made public on the www (http://www.zmbh.uni-heidelberg.de/M_pneumoniae).

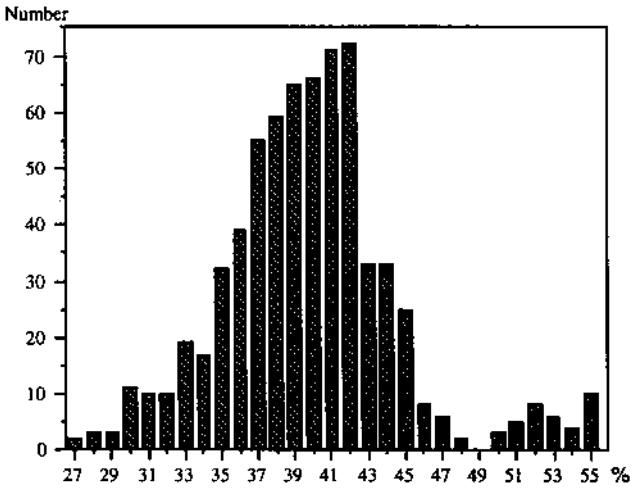


Figure 2. Distribution of the G+C content of the coding sequences of all *M.pneumoniae* ORFs.

DNA replication and repair

The central enzyme for DNA replication in bacteria is the DNA polymerase III holoenzyme (32), which consists of 10 subunits in *E.coli*, a DNA polymerase subunit α and nine accessory proteins (ε, υ, τ, γ, δ, δ′, χ, ψ and β). *Mycoplasma pneumoniae* codes for two potential α subunits (the gene name in the literature is either dnaE or polC). Both proposed α subunits, A19_orf872 and B01_orf1443, differ in length and also in their degree of similarity to the α subunits from *E.coli* and *Bacillus subtilis*. The protein from B01_orf1443 shares the highest similarity with the α subunit from Gram-positive bacteria including the motif for a 3′–5′ exonuclease activity which is typical for these bacteria. In contrast, the orf A19_orf872 is most similar to the α subunit from *E.coli* and does not contain a 3′–5′ exonuclease domain. The 3′–5′ exonuclease activity in *E.coli* is encoded by a separate gene (dnaQ), which has not been found in *M.pneumoniae*. Of the other subunits which build the DNA polymerase III holoenzyme in *E.coli* (32) only the subunits β (dnaN), δ′(holB), γ and τ (dnaX) are present in *M.pneumoniae*, indicating a simplified replication complex compared with the Gram-negative bacteria *E.coli* and *H.influenzae*. Presently, it cannot be excluded that other proteins replace these subunits in *M.pneumoniae*. A true comparison with a phylogenetically closer related Gram-positive bacterium like *B.subtilis* is not possible since the *Bacillus* DNA polymerase III holoenzyme complex has not been defined as yet and the nucleotide sequence of the entire *B.subtilis* genome has not been completed.

Mycoplasma pneumoniae does not code for a DNA polymerase I (polA)-like DNA repair enzyme. Instead, we find a truncated polA gene (A19_orf291) comprising only the 5′–3′ exonuclease domain, whereas in *E.coli* and *B.subtilis* the polA gene is much larger and codes for the 5′–3′ exonuclease and a 5′–3′ polymerase-specific domain.

Experimental results on DNA polymerase enzymatic activities in mycoplasmas are confusing. It was claimed that the DNA polymerase III of *Mollicutes* lacks the 3′–5′ exonuclease proofreading activity in general (33) and this was taken as an explanation for the observed genetic instability of many *Mollicutes* species (4). Recently, the nucleotide sequence of the polC gene of

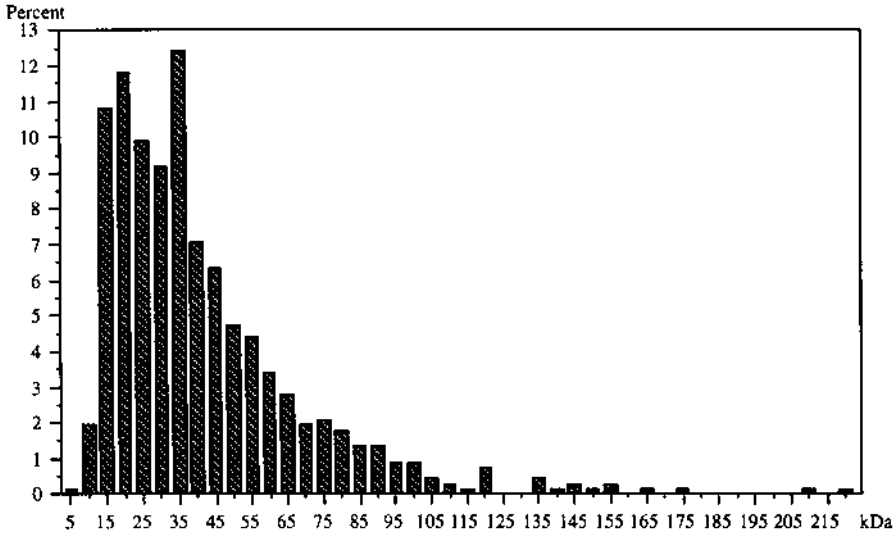


Figure 3. Distribution of all *M.pneumoniae* proteins according to their molecular weight.

Mycoplasma pulmonis and experimental results on enzyme purification and characterization of enzyme activities were published (34). The results indicated that the polC gene from *M.pulmonis* also codes for a 3′–5′ exonuclease, and that the size of the predicted PolC protein, 1435 amino acids, is very similar to the PolC homolog B01_orf1443 in *M.pneumoniae* and that the polymerase could be inhibited by compounds specific for PolC proteins of Gram-positive bacteria. Furthermore, the authors provided some experimental evidence for a second, smaller enzyme with DNA polymerase activity. Considering the characterization data of DNA polymerase activities in *M.pulmonis* and the nucleotide sequence data on DNA polymerase genes of *M.pneumoniae* and *M.genitalium* (9,35), one can conclude that at least these three *Mycoplasma* species have two DNA polymerase (polC) genes coding for a larger protein (≈1400 amino acids) with a 3′–5′ exonuclease activity and with the highest sequence similarities to the Gram-positive *B.subtilis* polymerase III. Therefore it is unlikely that an increased mutation frequency is caused by the DNA replication process. The nucleotide sequence of the smaller Pol III homolog (≈100 kDa) of *M.pneumoniae* and *M.genitalium* (9,35) resembles more the polC gene from the Gram-negative *E.coli*. This is also emphasized by the absence of the 3′–5′ exonuclease domain in the proposed genes. The gene for the smaller, Gram-negative typical PolC has not yet been found in *M.pulmonis*, but during the purification of the larger PolC, a second polymerase activity lacking exonuclease activity has been identified. The function of the exonuclease negative DNA polymerase can only be elucidated experimentally and it remains to be seen if it can substitute for the function of the polymerase I (PolA) in combination with the proposed 5′–3′ exonuclease of the truncated polA gene (A19_orf291). This topic has been also discussed for *M.genitalium* (35).

In addition to the DNA polymerase many more gene products are necessary for DNA replication, e.g. initiation, elongation and termination (32). The most obvious functions missing in *M.pneumoniae* according to the sequence analysis are an RNaseH for primer removal and a protein for the termination of replication.

The number of genes involved in DNA repair is considerably smaller in *M.pneumoniae* than in the 'standard' eubacteria *E.coli* and *B.subtilis* or even *H.influenzae* with the smaller genome.

Mycoplasma pneumoniae codes only for 13 of the genes known to be involved in excision repair of DNA, recombination and SOS repair. Thus the genes recB, recC, recD, recG and ruvC involved in recombination are missing as well as the genes recN, recO, recQ and recR involved in SOS repair in *E.coli*. Nevertheless, a rudimentary stock of enzymes has been conserved in *M.pneumoniae* to permit homologous recombination [RecA, Ssb, PolA (see above), GyrA, GyrB, RuvA and RuvB] (36), excision repair (37) and a kind of truncated SOS repair (38). In particular missing is the lexA gene which plays a central role in regulating the SOS response including the expression of the recA gene in other bacteria.

We were also unable to find components of the so called mismatch-repair system encoded by the mutS, mutL and mutH genes. Since bacteria which normally carry the mut genes show a reduced genetic stability, if these genes are mutated, it seems likely that the absence of these genes in mycoplasmas causes an increased mutation rate (65).

Transcription

The DNA dependent RNA polymerase of *M.pneumoniae* is coded by the conserved genes rpoA (α subunit), rpoB (β subunit), rpoC (β′ subunit) and rpoE (δ′ subunit). The only sigma factor found (H91_orf499) shares the highest similarity with the sigma factor SigA from *B.subtilis* (39). Presently, not enough experimental data are available for defining promoter sequences in *M.pneumoniae*. The promoter of only three genes/operons have been determined experimentally by primer extension. These genes are the P1 operon (14), the ribosomal RNA operon (40) and F10_orf405 (27). The –10 region and to a lesser extent the –35 region of these three examples are comparable with consensus promoters sequences in *B.subtilis* (41). Termination of transcription seems to be independent of the termination factor Rho, since the corresponding gene could not be found. Transcription stops on typical terminator sequences which are short interrupted palin-

Table 3. Codon usage of different sets of *M.pneumoniae* ORFs: all 677 ORFs; ORFs with a G+C content <35 mol%; codon usage of the adhesin P1 and ORF6 (high G+C content); ribosomal ORFs as examples for frequently expressed proteins

AmAcid	Codon	all MP ORFs(677) /1000	GC<35% /1000	high GC (P1+orf6) /1000	ribosomal ORFs /1000
Ala	GCA	13.76	14.92	8.43	14.90
Ala	GCC	16.50	8.09	27.75	16.95
Ala	GCG	11.05	4.43	22.48	13.12
Ala	GCT	25.20	22.80	25.64	30.62
Arg	AGA	4.02	11.22	2.46	5.19
Arg	AGG	2.84	3.70	4.21	1.37
Arg	CGA	2.48	3.55	2.81	3.42
Arg	CGC	10.72	4.59	14.75	22.83
Arg	CGG	5.00	0.94	5.27	8.20
Arg	CGT	9.68	5.63	6.32	21.46
Asn	AAC	37.01	27.91	41.80	41.69
Asn	AAT	25.09	45.50	24.24	15.72
Asp	GAC	19.16	13.88	25.99	14.63
Asp	GAT	30.40	39.18	32.31	19.68
Cys	TGC	2.09	2.82	0.00	2.32
Cys	TGT	5.39	5.48	0.00	3.96
Gln	CAA	37.90	39.55	31.96	35.95
Gln	CAG	15.65	7.46	21.07	8.34
Glu	GAA	42.01	53.22	20.02	39.64
Glu	GAG	14.71	12.47	12.29	11.34
Gly	GGA	6.38	9.29	8.43	7.52
Gly	GGC	11.81	9.34	22.13	12.17
Gly	GGG	8.95	2.30	25.99	8.61
Gly	GGT	27.90	22.33	27.75	34.86
His	CAC	11.86	6.16	8.08	16.54
His	CAT	6.17	6.16	2.81	4.24
Ile	ATA	5.46	12.84	1.40	1.78
Ile	ATC	14.39	13.10	11.59	13.94
Ile	ATT	45.99	48.21	16.16	47.57
Leu	CTA	10.62	10.64	3.86	8.88
Leu	CTC	12.23	6.47	26.69	13.81
Leu	CTG	9.54	5.17	10.89	6.01
Leu	CTT	10.06	18.10	8.78	7.38
Leu	TTA	39.24	46.54	19.32	34.03
Leu	TTG	21.48	17.48	22.48	16.54
Lys	AAA	46.27	73.20	24.24	61.92
Lys	AAG	39.08	29.84	33.02	63.01
Met	ATG	15.60	13.98	7.38	21.32
Phe	TTC	12.75	16.23	10.89	7.52
Phe	TTT	43.03	53.17	25.64	24.06
Pro	CCA	10.86	9.76	16.51	12.03
Pro	CCC	9.05	3.13	23.18	7.11
Pro	CCG	6.65	2.40	14.05	7.52
Pro	CCT	8.30	9.86	9.13	9.16
Ser	AGC	10.62	10.49	11.94	8.20
Ser	AGT	21.04	21.76	28.10	12.85
Ser	TCA	8.74	13.20	8.43	8.61
Ser	TCC	9.59	6.73	22.48	9.84
Ser	TCG	6.43	3.18	15.10	5.06
Ser	TCT	8.16	15.03	5.97	6.15
Thr	ACA	10.38	15.18	8.43	8.47
Thr	ACC	21.92	11.74	45.66	27.88
Thr	ACG	7.90	3.60	18.97	6.56
Thr	ACT	19.32	24.16	10.89	17.22
Trp	TGA	6.06	8.77	9.83	2.32
Trp	TGG	5.82	3.60	9.13	4.10
Tyr	TAC	17.94	15.34	16.51	13.67
Tyr	TAT	14.26	20.04	10.89	9.16
Val	GTA	13.73	11.64	7.73	21.05
Val	GTC	11.03	4.85	15.45	8.47
Val	GTG	18.73	6.37	29.50	21.46
Val	GTT	21.17	27.50	14.05	23.10
xxx	TAA	2.05	2.97	0.35	1.91
xxx	TAG	0.78	0.83	0.35	5.06

dromic regions followed by a run of U residues. The Nus transcription termination factors, of which NusA (E07_orf540) and NusG (D09_orf320) are present, may play a role in the termination of transcription. NusB and NusC are absent. NusA is involved in termination and NusG in antitermination in other bacteria. Finally, GreA promotes elongation by the RNA polymerase by utilizing a novel transcript-cleavage reaction (42).

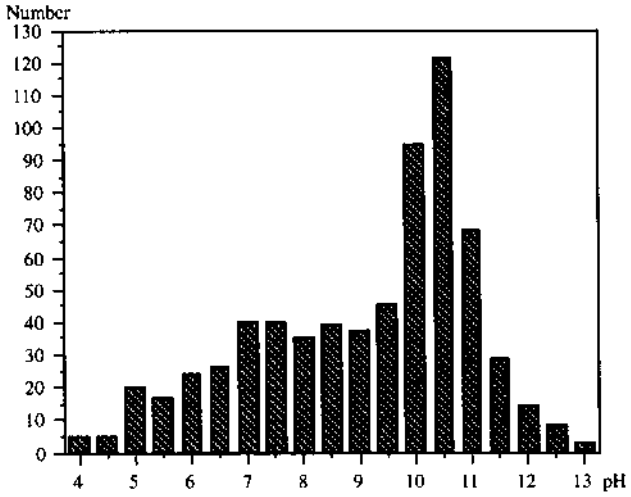


Figure 4. Distribution of all *M.pneumoniae* proteins according to their predicted isoelectric point (IP).

Gene expression and regulation

Regulation of gene expression in *M.pneumoniae* has not been studied so far. Therefore we do not know how this bacterium coordinates the synthesis of those gene products which are essential for reproduction. Also, *M.pneumoniae* has to sense and respond to environmental changes. This requires a signal transduction system. The presence of only one sigma factor (sigA, H91_orf499) which is also the only one of all proposed proteins showing the characteristic helix–turn–helix (HTH) motif, suggests that the response to external stimuli is not controlled by the level of expression of alternative sigma factors.

The presence of a *cis*-acting conserved palindromic repeated sequence in front of four heat shock genes, similar to the 'CIRCE' element first identified in *B.subtilis* (43) and the identification of the proposed repressor (C09_orf351, hrcA), indicates that the heat shock response in *M.pneumoniae* is regulated by the interaction of this repressor with the CIRCE element, and provides an example for a negative regulation of gene expression in *M.pneumoniae*.

The two-component signal transduction system (44), consisting of a sensor and a response regulator, which has been found in many prokaryotic and eukaryotic organisms is believed to be essential for all cells. Nevertheless, based on sequence similarity we were unable to detect any such system in *M.pneumoniae*.

Concerning other proteins with regulatory functions we identified several GTP-binding proteins and other proteins like the virulence associated protein vacB (K04-orf726). These regulatory proteins act by unknown mechanisms.

Translation

The translation machinery of *M.pneumoniae* is rather extensive. About 15% of all proposed ORFs, are involved in translation including 19 tRNA synthetases, 50 ribosomal proteins, various factors and enzymes, 33 tRNAs, one ribosomal RNA operon with one copy of each 5S, 16S and 23S rRNA (45), and a gene coding for the 10Sa RNA. The conservation of the 10Sa RNA which functions as tRNA and mRNA and is implicated in *trans*-translation (66), is interesting in evolutionary terms. Three exceptions are

Table 4. List of the proposed ORFs, RNAs and REPs in numerical order starting with E07_orf540o on the gene map (Fig. 1)

Number	Genome Position	Name	Annotation
001	663**815435 (rcf)	E07_orf540o	N-utilization substance protein A homolog (nusA); BACSU
002	4081..340	E07_orf1113	MG140 homolog, MYCGE
003	6641..4257	E07_orf794	putative lipoprotein, MG260 homolog, MYCGE
004	7325..8924	E07_orf133	-
005	8482..7808	E07_orf224	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
	8620..7896	REPMP5	repetitive DNA sequence REPMP5
006	9614..8310	E07_orf434	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
007	10589..10167	E07_orf140	-
008	12589..11132	E07_orf485	MG260 homolog, MYCGE
009	13393..12596	E07_orf265	MG260 homolog, MYCGE
010	14250..13711	E07_orf179	-
011	15843..14602	E07_orf413	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	16274..14754	REPMP2/3	repetitive DNA sequence REPMP2/3
012	16944..16417	E07_orf175	-
013	20717..17061	E07_orf1218	hypothetical 130K protein (orf6; P1 operon); MYCPN
	20717..18017	REPMP5	repetitive DNA sequence REPMP5
	23560..21780	REPMP2/3	repetitive DNA sequence REPMP2/3
014	25606..20723	E07_orf1627	ADP1_MYCPN adhesin P1 (orf5, P1 operon); MYCPN
	25606..24060	REPMP4	repetitive DNA sequence REPMP4
015	26393..25619	E07_orf324	hypothetical 28K protein (orf4, P1 operon); MYCPN
	26823..27081	REPMP1	repetitive DNA sequence REPMP1
016	26844..27335	E07_orf163	-
017	27573..28072	E07_orf166	-
018	28321..29007	E07_orf228	-
019	30544..29585	E07_orf319	sn-glycerol-3-phosphate transport system permease protein (ugpE); ECOLI
020	31505..30516	E07_orf329	sn-glycerol-3-phosphate transport system permease protein (ugpA); ECOLI
021	33258..31498	E07_orf586	sn-glycerol-3-phosphate transport system permease protein (ugpC); ECOLI
022	34187..33282	E07_orf301	putative lipoprotein, MG186 homolog, MYCGE
	35192..36457	REPMP2/3	repetitive DNA sequence REPMP2/3
023	35415..34645	E07_orf256L	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
024	36396..35731	E07_orf221V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	37389..37148	REPMP1	repetitive DNA sequence REPMP1
025	37422..37000	C09_orf140o	-
	38383..37821	REPMP2/3	repetitive DNA sequence REPMP2/3
026	38832..38383	C09_orf149b	adhesin P1 (group 2) homolog; MYCPN
027	39981..39532	C09_orf149a	-
	40650..39538	REPMP4	repetitive DNA sequence REPMP4
028	41980..41438	C09_orf180	-
029	42851..42372	C09_orf159	MG207 homolog, MYCGE
030	44647..42887	C09_orf586L	excinuclease ABC subunit C (uvrC); BACSU
031	44679..45734	C09_orf351	protein (hrcA) homolog, BACSU
032	48090..45721	C09_orf789	topoisomerase IV subunit A (parC), BACSU
033	49997..48090	C09_orf655	topoisomerase IV subunit B (parE), BACSU
	50032..50105	mpg1	Thr-tRNA(GGU) gene; MYCPN
034	50488..50133	C09_orf121	MG203 homolog, MYCGE
035	51141..50488	C09_orf217	heat shock protein GrpE, HAEIN
036	53896..51164	C09_orf910	Dnal homolog protein, MYCCA
037	54231..54562	C09_orf143b	MG199 homolog, MYCGE
038	55020..54657	C09_orf127	ribosomal protein L20 (rplZ0); MYCPE
039	55310..55031	C09_orf99	ribosomal protein L35 (rpL35); BACST
040	55821..55216	C09_orf201	translation initiation factor IF3 (infC); MYCPE
041	57713..55911	C09_orf800	carnitine palmitoyltransferase II precursor(cpt2); HUMAN
042	58334..57703	C09_orf223	-
043	59315..58923	C09_orf130b	-
044	61443..60175	C09_orf422	-
045	64103..61943	C09_orf718	-
046	64524..64027	C09_orf165	-
047	66438..65204	C09_orf404	-
048	67175..66420	C09_orf251	-
049	69705..67288	C09_orf805	phenylalanyl-tRNA synthetase beta chain (pheT); BACSU
050	70733..69708	C09_orf341	phenylalanyl-tRNA synthetase alpha-subunit (pheS); BACSU
051	71881..71567	C09_orf104	(MG191 homolog, MYCGE)
052	71891..72409	C09_orf172	-
053	73896..73078	C09_orf272	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
	74668..72883	REPMP5	repetitive DNA sequence REPMP5
054	75958..74712	C09_orf428V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	76038..74736	REPMP4	repetitive DNA sequence REPMP4
	76973..76691	REPMP1	repetitive DNA sequence REPMP1
055	77006..76455	R02_orf183o	-
056	78388..77345	R02_orf347L	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	79072..77697	REPMP2/3	repetitive DNA sequence REPMP2/3
057	79517..79074	R02_orf147	MG280 homolog, MYCGE
058	81440..79815	R02_orf541	putative lipoprotein, MG280 homolog, MYCGE
059	82410..81616	R02_orf264	-
060	83174..82410	R02_orf254	-
	83460..83358	5s rRNA	5S rRNA
	86408..83682	23s rRNA	23S rRNA
	88155..86632	16s rRNA	16S rRNA
061	90177..89755	R02_orf140	-
	90202..89903	REPMP1	repetitive DNA sequence REPMP1
062	91516..90681	R02_orf301	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
063	91892..91371	R02_orf173	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
064	92826..92230	R02_orf138	-
	93192..90643	REPMP5	repetitive DNA sequence REPMP5
065	93692..92703	R02_orf329	-
066	94856..93847	R02_orf335	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
067	95651..95346	R02_orf101	-
068	97118..96666	R02_orf150	-
069	97607..97290	R02_orf105	-
070	99191..97868	R02_orf440	-
071	100872..99258	R02_orf524	MG068 homolog, MYCGE
072	102523..100922	R02_orf533	putative lipoprotein, MG067 homolog, MYCGE
073	104479..102533	R02_orf648	transketolase 1 (TK 1; dxsB); RHOSH
074	105897..104500	R02_orf465	glutamine transport ATP-binding protein (glnQ); ECOLI
075	110057..105897	R02_orf1386V	MG064 homolog, MYCGE
076	111196..110094	R02_orf300	1-phosphofructokinase (fruK); HAEIN
077	113275..111189	R02_orf694	fructose-permease IIBC component (fruA); ECOLI
	113324..113432	mpgpb	Ser-tRNA gene (AGC); MYCPN
078	113856..115265	R03_orf469	MG061 homolog, MYCGE
079	115471..117165	R03_orf564o	hexosephosphate transport protein (uhpT); SALTY
080	118116..117237	D09_orf299	hypothetical protein (ywdF) homolog, BACSU
081	118123..118586	D09_orf147	hypothetical protein (A43299) homolog, ENTHR
082	118375..119539	D09_orf388	phosphoribosylpyrophosphate synthetase (prs); SYNP

Table 4. *Continued*

083	119518..120054	D09_orf178	hypothetical protein (yabF) homolog; BACSU
084	120036..120866	D09_orf276	hypothetical protein (yabC) homolog; BACSU
085	120853..121236	D09_orf127a	-
086	121404..121781	D09_orf125	MG055 homolog, MYCGE
087	121789..122751	D09_orf320	transcription antitermination factor (nusG); BACSU
088	124383..122719	D09_orf554	phosphomannomutase (cpsG); MYCPI
089	124774..124373	D09_orf133	cytidine deaminase (cdd); MYCPI
090	126050..124785	D09_orf421	thymidine phosphorylase (deoA); MYCPI
091	126711..126037	D09_orf224	deoxyribose-phosphate aldolase (deoC); MYCPN
092	127431..126715	D09_orf238	purine-nucleoside phosphorylase (deoD); BCOLI
093	127487..128839	D09_orf450	signal recognition particle protein (ffh); MYCMY
094	130278..129127	D09_orf383	S-adenosylmethionine synthetase 2 (metX); BCOLI
095	131221..130262	D09_orf319	o-sialoglycoprotein endopeptidase (gcp); PASHA
096	132678..131221	D09_orf485	putative lipoprotein, MG045 homolog, MYCGE
097	133523..132663	D09_orf286a	spermidine/putrescine transport system permease (potJ); BCOLI
098	134376..133516	D09_orf286b	spermidine/putrescine transport system permease (potB); HAEIN
099	136060..134378	D09_orf560L	spermidine/putrescine transport ATP-binding prot (potA); BCOLI
100	137837..137466	D09_orf123	putative lipoprotein
101	139642..139376	D09_orf88	phosphocarrier protein HPr (ptsH); MYCCA
102	141633..139660	D09_orf657	putative lipoprotein, MG040 homolog, MYCGE
103	141816..142970	D09_orf384	aerobic glycerol-3-phospate dehydrogenase (glpD); BCOLI
104	142961..144487	D09_orf508	glycerol kinase (glpK), HAEIN
105	146845..144947	D09_orf632	MG288 homolog, MYCGE
106	148578..147022	D09_orf518	MG096 homolog, MYCGE
107	150522..149167	D09_orf451	pre-B cell enhancing factor homolog (pbeF); HUMAN
108	152171..150498	D09_orf557	aspartyl-tRNA synthetase (aspS); THEAQ
109	153387..152143	B01_orf414o	histidyl-tRNA synthetase (hisS); STREQ
110	153414..153989	B01_orf191	thymidine kinase (tdk); BACSU
111	154830..154036	B01_orf264	glycerol uptake facilitator (glpF); BACSU
112	157172..155154	B01_orf672	MG032 homolog, MYCGE
113	157794..157234	B01_orf186L	MG032 homolog, MYCGE
114	158048..158359	B01_orf103b	-
115	159270..158254	B01_orf338	MG032 homolog, MYCGE
116	159672..160020	B01_orf116L	-
	160267..160532	REPMP1	repetitive DNA sequence REPMP1
117	160694..160251	B01_orf147	-
118	162883..160862	B01_orf673	MG032 homolog, MYCGE
119	165055..163055	B01_orf666	MG032 homolog, MYCGE
120	165333..169664	B01_orf1443	DNA polymerase III (dnaE) alpha chain (3'-5' exonuclease); BACSU
121	169788..170324	B01_orf178	uracil phosphoribosyltransferase (upp); STRSL
122	170328..170654	B01_orf108	hypothetical protein (gi: 606093) homolog; BCOLI
123	171489..170678	B01_orf203	MG028 homolog, MYCGE
124	171995..171489	B01_orf168	MG027 homolog, MYCGE
125	172485..171913	B01_orf190	elongation factor P (efp) homolog; HAEIN
126	173405..172506	B01_orf299V	TraH protein; YEREN
127	173438..174262	B01_orf274	-
128	175353..174265	B01_orf362	hypothetical protein (yyaF) homolog; BACSU
129	176220..175354	B01_orf288	fructose-bisphosphate aldolase (tsr); BACSU
130	176660..176220	B01_orf146	DNA-directed RNA polymerase delta subunit (rpoE); BACSU
131	178219..176681	B01_orf512	methionyl-tRNA synthetase (metS); BACST
132	179148..178219	B01_orf309	proline iminopeptidase (pip); NEIGO
133	180304..179132	D12_orf390o	heat shock protein DnaJ; BACSU
134	183442..180350	D12_orf1030	hypothetical helicase (yb95) homolog; YEAST
135	185356..183452	D12_orf634	transport ATP-binding protein (msbA); HAEIN
136	187139..185268	D12_orf623	transport ATP-binding protein (pmd1); SCHPO
	187233..187390	mpigi	Ile-tRNA(ATC), Ala-tRNA(GCA) genes; MYCPN
137	187475..188284	D12_orf269	5,10-methylene-tetrahydrofolate dehydrogenase (mtd1); HAEIN
138	188259..189125	D12_orf288	ribosomal protein S6 modification protein (rimK); ECOLI
139	189125..189982	D12_orf285	MG011 homolog, MYCGE
140	190597..189959	D12_orf212	DNA primase motif (dnaG); CLOAB
141	191472..190699	D12_orf257	-
142	192199..192906	D12_orf235	-
143	192931..193626	D12_orf231	putative lipoprotein
144	194207..193812	D12_orf131	-
145	195189..194404	D12_orf261	hypothetical protein (yabD) homolog; BACSU
146	196517..195189	D12_orf442	possible thiophene and furan oxidation protein (tdhF); BACSU
147	197280..196519	D12_orf253	DNA polymerase III subunit delta' (holB); BCOLI
148	197885..197253	D12_orf210	thymidylate kinase (CDC8) homolog, MYCGE
149	199152..197890	D12_orf420	seryl-tRNA synthetase (serS); BACSU
150	201643..199124	K05_orf839o	DNA gyrase subunit A (gyrA); STAAU
151	203595..201643	K05_orf650	DNA gyrase subunit B (gyrB); MYCPN
152	204626..203697	K05_orf309	DnaJ homolog protein; YEAST
153	205772..204630	K05_orf380	DNA polymerase III beta subunit (dnaN); STAAU
154	206520..207332	K05_orf270	protein (soj) homolog; BACSU
155	207319..208071	K05_orf250	-
156	208071..209390	K05_orf439	chromosomal replication initiator protein (dnaA); MYCCA
157	209458..210312	K05_orf284	sulfate transport ATP-binding protein (cysA); SYNP
158	210318..215966	K05_orf1882	-
159	215968..216987	K05_orf339	protein (devA) homolog; ANASP
160	217010..217156	K05_orf48	ribosomal protein L34 (rpL34); PROMI
161	217146..217502	K05_orf118V	RNaseP C5 chain (rnpA); MYCCA
162	217483..218640	K05_orf385	hypothetical protein 1 (S42122); MYCCA
163	218633..219424	K05_orf263V	S-adenosylmethionine-6-N',N'-adenosyl(rRNA) dimethyltransferase (ksgA); ECOLI
164	219411..220865	K05_orf484	glutamyl-tRNA synthetase (gltX); BACST
165	220846..222123	K05_orf425	MG461 homolog, MYCGE
166	223000..222680	K05_orf106	-
167	223391..223696	K05_orf101a	-
168	225039..224101	K05_orf312	L-lactate dehydrogenase (ldh); MYCHY
169	225210..225719	K05_orf169	hypothetical protein (HI0671) homolog; HAEIN
170	225719..226246	K05_orf175	hypoxanthine-guanine phosphoribosyltransferase (hpt); LACLA
171	226427..228556	K05_orf709	cell division protein (ftsH); BACSU
172	229109..230146	K05_orf345	MG456 homolog, MYCGE
173	231385..230186	K05_orf399	tyrosyl tRNA synthetase (tyrS); BACCA
174	231411..231833	K05_orf140	osmotically inducible protein (osmC); ECOLI
175	232705..231830	K05_orf291	UDP-glucose pyrophosphorylase (gtaB); BACSU
176	233448..232693	K05_orf251	MG452 homolog, MYCGE
177	233533..234717	K05_orf394	elongation factor TU (tuf); MYCGE
178	234876..235589	K05_orf237	homolog (degV) protein; BACSU
179	235596..236300	K05_orf234	MG449 homolog, MYCGE
180	236264..236719	K05_orf151	pilB homolog (fragment); HAEIN
181	236870..238369	K05_orf499	MG447 homolog, MYCGE
182	238451..238717	K05_orf88	ribosomal protein S16 (BS17); BACSU
183	238783..239415	K05_orf210	tRNA (guanine-N1)-methyltransferase (trmD); HUMAN
184	239399..239758	K05_orf119	ribosomal protein L19 (rpL19); BACST

Table 4. *Continued*

185	239774..240979	K05_orf401	hypothetical protein (P27712); SPICI
186	240948..241763	K05_orf271	MG442 homolog, MYCGE
187	242850..242236	B09_orf204o	protein P30, MYCPN
188	243127..243516	B09_orf129	putative lipoprotein
189	244320..243889	B09_orf143V	PTS system mannitol-specific component IIA (EIIA-MTL)(mtlF); STRMU
190	245395..244301	B09_orf364	mannitol-1-phosphate 5-dehyrogenase (EC 1.1.1.17)(mtlD); STRMU
191	246521..245382	B09_orf379	PTS system mannitol-specific component IIA (EIIA-MTL)(mtlA); STACA
192	247519..247824	B09_orf101	putative lipoprotein
193	247809..248219	B09_orf136L	-
194	249106..249516	B09_orf136	MG441 homolog, MYCGE
195	249627..250499	B09_orf290	putative lipoprotein, MG439 homolog, MYCGE
196	250522..251355	B09_orf277	putative lipoprotein, MG440 homolog, MYCGE
197	251355..252206	B09_orf283a	putative lipoprotein, MG439 homolog, MYCGE
198	252209..253060	B09_orf283b	putative lipoprotein, MG439 homolog, MYCGE
199	252981..253889	B09_orf302	MG440 homolog, MYCGE
200	253889..254782	B09_orf279	putative lipoprotein, MG439 homolog, MYCGE
201	254731..255561	B09_orf276	putative lipoprotein, MG440 homolog, MYCGE
202	255561..256463	B09_orf300	putative lipoprotein, MG439 homolog, MYCGE
203	256471..257334	B09_orf287o	MG439 homolog, MYCGE
204	258458..257331	E30_orf375	MG438 homolog, MYCGE
205	259665..258478	E30_orf395	CDP-diglyceride synthetase (cdsA); HAEIN
206	260219..259665	E30_orf184	ribosome releasing factor (frr); HAEIN
207	261354..260296	E30_orf352	-
208	262455..261910	C12_orf181o	-
209	263280..262537	C12_orf247	-
210	264090..263383	C12_orf235	uridylate kinase (pyrH); ECOLI
211	264988..264092	C12_orf298	elongation factor Ts (tsf); SPICI
212	265075..266289	C12_orf404	hypothetical protein (yfiB) homolog; SPICI
213	266342..267076	C12_orf244	triosephosphate isomerase (tim); ECOLI
214	267069..268595	C12_orf508	phosphoglycerate mutase (pgm); BACSU
215	268600..270318	C12_orf572	PEP-dependent HPr protein kinase phosphoryltransferase (Enzyme I) (ptsI); STRSL
216	270833..270315	C12_orf172	MG428 homolog, MYCGE
217	271393..270968	C12_orf141	MG427 homolog, MYCGE
218	271634..271437	C12_orf65	ribosomal protein L28 (rpL28); BACSU
219	273008..271656	C12_orf450	ATP-dependent RNA helicase (deaD); HAEIN
220	273166..273426	C12_orf86	ribosomal protein S15 (BS18); BACST
221	273431..275116	C12_orf561	MG423 homolog, MYCGE
222	275162..590313	C12_orf839	MG422 homolog, MYCGE
223	277659..280505	C12_orf948L	excinuclease ABC subunit A (uvrA); ECOLI
224	280514..282559	C12_orf681	DNA polymerase III subunit gamma and tau (dnaX); ECOLI
225	282590..283030	C12_orf146	ribosomal protein L13 (rpL13); ECOLI
226	283036..283434	C12_orf132	ribosomal protein S9 (rpS9); BACST
227	283864..284613	C12_orf249	restriction-modification enzyme subunit S1B (hsdS); MYCPU
228	284699..285703	C12_orf334	MG413 homolog, MYCGE
229	285639..286673	C12_orf344	MG415 homolog, MYCGE
230	286788..289781	C12_orf997	MG414 homolog, MYCGE
231	290023..291180	C12_orf385	MG412 homolog, MYCGE
232	291180..293135	C12_orf651V	phosphate transport system permease protein (pstA); ECOLI
233	293120..294109	C12_orf329	phosphate transport ATP-binding protein (pstB); ECOLI
234	294112..294789	C12_orf225	phosphate transport system regulatory protein (phoU); ECOLI
235	295259..294786	C12_orf157	peptide methionine sulfoxide reductase (pmsR), ECOLI
236	295314..296684	C12_orf456	enolase (eno) (EC 4.2.1.11); PLAFA
238	297129..298010	C12_orf293o	ATP synthase A chain (atpB); MYCGA
237	297163..296690	C12_orf157L	ATP synthase protein I (atpI); MYCGA
239	298013..298330	D02_orf105	ATP synthase C chain (atpE); MYCGA
240	298333..298956	D02_orf207	ATP synthase B chain (atpF); MYCGA
241	298949..299485	D02_orf178	ATP synthase delta chain (atpH); MYCGA
242	299488..301044	D02_orf518	ATP synthase alpha chain (atpA); MYCGA
243	301044..301883	D02_orf279	ATP synthase gamma chain (atpG); MYCGA
244	301883..303310	D02_orf475	ATP synthase beta chain (atpD); MYCGA
245	303313..303714	D02_orf133a	ATP synthase epsilon chain (atpC); MYCGA
246	303714..305423	D02_orf569	MG397 homolog, MYCGE
247	305423..305881	D02_orf152	galactose-6-phosphate isomerase subunit (lacA); STRMU
248	305799..306167	D02_orf122a	-
249	306393..306761	D02_orf122b	-
250	306862..308427	D02_orf521	putative lipoprotein, MG395 homolog, MYCGE
251	308950..310011	D02_orf353V	MG068 homolog, MYCGE
252	310168..310821	D02_orf217L	putative lipoprotein, MG395 homolog, MYCGE
253	310962..311435	D02_orf157L	MG395 homolog, MYCGE
254	311648..313243	D02_orf531	putative lipoprotein, MG395 homolog, MYCGE
255	313301..313753	D02_orf150	MG068 homolog, MYCGE
256	313629..314672	D02_orf347	MG067 homolog, MYCGE
257	314746..315654	D02_orf302	putative lipoprotein, MG068 homolog, MYCGE
258	315716..316123	D02_orf135L	MG067 homolog, MYCGE
259	316627..317304	D02_orf225L	MG068 homolog, MYCGE
260	317742..319061	D02_orf439	putative lipoprotein, MG068 homolog, MYCGE
261	319237..320034	D02_orf265V	MG068 homolog, MYCGE
262	320102..320524	D02_orf140	MG395 homolog, MYCGE
263	320666..320995	D02_orf109	-
264	321313..321011	D02_orf100	-
265	321751..322791	D02_orf346	MG068 homolog, MYCGE
266	322953..324173	D02_orf406	serine hydroxymethyltransferase (glyA); ACTAC
267	324608..324994	D02_orf128	-
268	325182..325532	D02_orf116	heat shock protein GroES; BACSU
269	325535..327166	D02_orf343	heat shock protein GroEL; BACSU
270	327180..328517	D02_orf445	nonspecified aminopeptidase; MYCSA
271	328621..330603	D02_orf660	lactococcin transport ATP-binding protein (lcnDR3); LACLA
272	330605..330994	D02_orf129	MG389 homolog, MYCGE
273	331116..331442	D02_orf108	MG388 homolog, MYCGE
274	331430..332305	D02_orf291	GTP-binding protein era homolog; STRMU
275	332405..335515	D02_orf1036o	protein P200; MYCPN
276	335519..336232	H03_orf237	glycerophosphoryl diester phosphodiesterase (glpQ); STAAU
277	336402..336860	H03_orf152	-
278	337074..338129	H03_orf351	NADP-dependent alcohol dehydrogenase (adh); THEBR
279	338333..339634	H03_orf433	GTP-binding protein (obg); BACSU
280	339627..340373	H03_orf248	probable NH(3)-dependent NAD(+) synthetase (outB); BACSU
281	341011..340370	H03_orf213	uridine kinase (udk); HAEIN
282	341065..342381	H03_orf438	arginine deiminase (arcA); PSEPU
	342382..342432	mptgab	Arg-tRNAgene (AGA); MYCPN
283	343166..342439	H03_orf235	MG381 homolog, MYCGE
284	343695..343120	H03_orf191	glucose inhibited division protein (gidB); ECOLI
285	345526..343688	H03_orf612	glucose inhibited division protein (gidA); ECOLI
286	345554..347167	H03_orf537	arginyl-tRNA synthetase (argS); BRELA
287	347210..347791	H03_orf193o	MG377 homolog (put. zinc protease), MYCGE

Table 4. *Continued*

288	347793..348107	G12_orf104	MG376 homolog, MYCGE
289	348107..349801	G12_orf564	threonyl-tRNA synthetase (thrSv); BACSU
290	349794..350603	G12_orf269	MG374 homolog, MYCGE
291	350610..351455	G12_orf281	MG373 homolog, MYCGE
292	351442..352605	G12_orf387	MG372 homolog, MYCGE
293	352598..353575	G12_orf325	hypothetical 28K protein (P1 operon) homolog; MYCPN
294	353562..354542	G12_orf326	hypothetical protein (HI0176) homolog; HAEIN
295	354597..356273	G12_orf558	MG369 homolog, MYCGE
296	356273..357259	G12_orf328a	fatty acid/phospholipid synthesis protein (plsX); ECOLI
297	357249..358097	G12_orf282a	ribonuclease III (rnc); ECOLI
298	360075..358081	G12_orf664	MG366 homolog, MYCGE
299	361010..360075	G12_orf311	methionyl-tRNA formyltransferase (fmt); ECOLI
300	361671..361015	G12_orf218	MG364 homolog, MYCGE
301	361732..361995	G12_orf87	ribosomal protein S20 (rpsT); ECOLI
302	362178..362005	G12_orf57	ribosomal protein L32 (rpL32); HAEIN
303	362553..362185	G12_orf122	ribosomal protein L7/L12 ('A' type) (rpL7/L12); MICLU
304	363076..362591	G12_orf161	ribosomal protein L10 (rplJ0); THEMA
305	363194..364432	G12_orf412	UV protection protein (mucB); ECOLI
306	365341..364418	G12_orf307	Holliday junction DNA helicase (ruvB); HAEIN
307	365936..365316	G12_orf206	Holliday junction DNA helicase (ruvA); ECOLI
308	366364..365942	G12_orf140b	-
309	366705..367877	G12_orf390	acetate kinase (ackA); BACSU
310	367885..368733	G12_orf282b	LicA protein (licA) homolog; HAEIN
311	368909..371056	G12_orf715	ATP-dependent protease binding subunit (clpB) homolog; HAEIN
312	371463..371053	G12_orf136	MG354 homolog, MYCGE
313	371612..371941	G12_orf109	MG353 homolog, MYCGE
314	373019..372465	G12_orf184	inorganic pyrophosphatase (ppa); THEAC
315	373074..373751	G12_orf225	-
316	374992..374006	G12_orf328b	MG350 homolog, MYCGE
317	376214..374973	G12_orf413	MG349 homolog, MYCGE
318	376807..377313	G12_orf168	-
	376824..377060	REPMP1	repetitive DNA sequence REPMP1
319	377903..378820	G12_orf305	putative lipoprotein, MG348 homolog, MYCGE
	378870..378945	mptgh	His-tRNA(CAC) gene; MYCPN
320	379607..378975	G12_orf210V	hypothetical protein (HI0340) homolog; HAEIN
321	380098..379598	G12_orf166b	hypothetical protein (ygl3) homolog; BACST
322	380141..382726	G12_orf861	isoleucine-tRNA ligase (ileS); STAAU
323	382844..383662	G12_orf272V	triacylglycerol lipase (lip) 3; MYCMY
324	383665..384711	G12_orf348	MG343 homolog, MYCGE
325	385804..386304	G12_orf166a	MG342 homolog, MYCGE
326	386397..390572	G12_orf1391o	RNA polymerase beta subunit (rpoB); BACSU
327	390576..394448	F04_orf1290	DNA-directed RNA polymerase beta' chain (rpoC); THEMA
328	394610..394972	F04_orf120	-
329	395489..395941	F04_orf150	-
330	396719..397183	F04_orf154	MG288 homolog, MYCGE
331	397214..397996	F04_orf260V	MG288 homolog, MYCGE
332	398608..399984	P02_orf458	MG096 homolog, MYCGE
333	401014..402297	P02_orf427	MG288 homolog, MYCGE
334	402844..404373	P02_orf509	MG288 homolog, MYCGE
335	405492..404401	P02_orf363V	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
336	407993..405612	P02_orf793	putative lipoprotein, MG260 homolog, MYCGE
337	408909..409670	P02_orf253	-
338	410138..409738	P02_orf126	-
339	411833..410688	P02_orf381	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
	412343..410580	REPMP5	repetitive DNA sequence REPMP5
340	413656..412388	P02_orf422V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	413701..412404	REPMP4	repetitive DNA sequence REPMP4
341	414691..414101	P02_orf196	-
	414718..414417	REPMP1	repetitive DNA sequence REPMP1
342	416640..415057	P02_orf527V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	416770..415161	REPMP2/3	repetitive DNA sequence REPMP2/3
343	417279..416788	P02_orf163	-
344	417961..417233	P02_orf242	L-ribulose-5-phosphate 4-epimerase (araD); ECOLI
345	418272..418703	P02_orf143	-
346	419131..421113	P02_orf660	hypothetical protein (yjfS) homolog; ECOLI
347	421405..421884	P02_orf159	hypothetical phosphotransferase protein (yjfU) homolog; ECOLI
348	421886..422542	P02_orf218	hypothetical protein (yjfV) homolog; ECOLI
349	422478..423395	P02_orf305	hypothetical protein (yjfW) homolog; ECOLI
350	424958..423534	P02_orf474	-
351	425032..426042	P02_orf336	recombination protein (recA); STAAU
352	426558..430460	P02_orf1300	putative lipoprotein, MG338 homolog, MYCGE
353	431060..430638	P02_orf140	MG337 homolog, MYCGE
354	432289..431063	P02_orf408	nitrogen fixation protein (nifS); HAEIN
356	432878..433828	P02_orf316	MG338 homolog, MYCGE
355	432936..432493	P02_orf147	-
	434119..434385	REPMP1	repetitive DNA sequence REPMP1
357	434245..434556	P02_orf103b	-
358	436086..435061	P01_orf341	hypothetical protein (yibD) homolog; ECOLI
359	436374..436955	P01_orf193	hypothetical protein (yihA) (era like) homolog; ECOLI
360	436939..439455	P01_orf838	valyl-tRNA synthetase (valS); BACST
361	439483..440076	P01_orf197	hypothetical protein (HI1366) homolog; HAEIN
362	440080..440787	P01_orf235	hypothetical protein (HI0315) homolog; HAEIN
363	440790..441419	P01_orf209	MG331 homolog, MYCGE
364	441446..442099	P01_orf217	cytidylate kinase (cmk); BACSU
365	442572..443450	P01_orf292	hypothetical protein (HI0136) (era like) homolog; HAEIN
366	443807..446908	P01_orf1033	MG328 homolog, MYCGE
367	446895..447701	P01_orf268	triacylglycerol lipase (lip) 2; MYCMY
368	447707..448588	P01_orf293	homolog (degV) protein; BACSU
369	448607..448768	P01_orf53	ribosomal protein L33 (rpL33); BACST
370	448768..449832	P01_orf354	X-Pro dipeptidase (pepX); LACDE
371	449873..450604	P01_orf243	-
	450647..451033	10saRNA	10saRNA; MYCGE
	451297..451058	rnpB RNA	RNaseP RNA; MYCGE
372	452076..451450	P01_orf208V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
373	452813..453118	P01_orf101	putative lipoprotein
374	453148..453570	P01_orf140	-
375	453614..454213	P01_orf199	-
	454252..453959	REPMP1	repetitive DNA sequence REPMP1
376	455967..454630	H08_orf445	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
377	456734..456261	H08_orf157a	-
	456769..454719	REPMP5	repetitive DNA sequence REPMP5
378	457621..456809	H08_orf270	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	457770..456825	REPMP4	repetitive DNA sequence REPMP4
379	458468..457773	H08_orf231	hypothetical protein (yzzC) homolog; BACSU

Table 4. *Continued*

380	458503..460030	H08_orf565	Na(+) translocating ATPase subunit J (ntpJ); ENTHR
	460165..460885	nptg>	Asn-tRNA(AAC), Glu-tRNA(GAA), Thr-tRNA(ACG), Val-tRNA(GTA), Thr-tRNA(ACA), Lys-tRNA(AAG), Leu-tRNA(CTA) genes; MYCPN
381	460960..462735	H08_orf591	MG321 homolog, MYCGE
382	462656..463129	H08_orf157b	MG321 homolog, MYCGE
383	463071..464060	H08_orf329V	adhesin P1 (group 2) homolog; MYCPN
384	464443..467460	H08_orf1005	putative lipoprotein, MG321 homolog, MYCGE
	467634..467717	nptga	Ser-tRNA(TCC), Ser-tRNA(TCG) genes; MYCPN
385	467786..468849	H08_orf287	(cytochrome C oxidase polypeptide I (coxD); BACSU)
386	468738..469319	H08_orf193	MG319 homolog, MYCGE
387	469340..470164	H08_orf274	30K adhesin-related protein; MYCPN
388	470138..472196	H08_orf672	cytadherence accessory protein (hmw3); MYCPN
389	472236..473345	H08_orf369	(competence locus E (comEJ); BACSU)
390	473224..474168	H08_orf314	MG315 homolog, MYCGE
391	474180..475526	H08_orf448	MG314 homolog, MYCGE
392	475643..476454	H08_orf263	MG313 homolog, MYCGE
393	476498..479554	H08_orf1018	cytadherence accessory protein (hmw1); MYCPN
394	479577..480194	H08_orf205	ribosomal protein S4 (rpS4); BACSU
395	481119..485096	H08_orf1325	putative lipoprotein, MG309 homolog, MYCGE
396	481124..480255	H08_orf289	triacylglycerol lipase (lip) 3; Mycoplasma sp
397	485303..486332	H08_orf409	ATP-dependent RNA helicase (deaD); ECOLI
398	486317..486769	H08_orf150	putative lipoprotein, MG307 homolog, MYCGE
399	487390..487082	H08_orf102	-
400	487860..490040	H08_orf726	MG307 homolog, MYCGE
401	490196..490909	H08_orf237	putative lipoprotein, MG307 homolog, MYCGE
402	490965..492002	H08_orf345	MG307 homolog, MYCGE
403	492220..493958	H08_orf571o	MG307 homolog, MYCGE
404	494247..497981	A05_orf1244	putative lipoprotein, MG307 homolog, MYCGE
405	497991..499178	A05_orf395	MG306 homolog, MYCGE
406	499234..501021	A05_orf595	heat shock protein DnaK, ERYRH
407	501179..501991	A05_orf270L	abc transport ATP-binding protein (chiD), SALTY
408	501886..503034	A05_orf382	abc transport ATP-binding protein (artP); ECOLI
409	503024..503977	A05_orf317	MG302 homolog, MYCGE
410	504008..505021	A05_orf337	glyceraldehyde-3-phosphate dehydrogenase(gapL CLOPA
411	505024..506253	A05_orf409	phosphoglycerate kinase (pgk); THEMA
412	506291..507253	A05_orf320	phosphotransacetylase (pta); BACSU
413	508131..507259	A05_orf290	hypothetical protein (yidA) homolog; ECOLI
414	508316..511264	A05_orf983	P115 protein homolog (SOC3); MYCHR
415	511270..512316	A05_orf348	cell division protein (fluY); ECOLI
416	512297..512605	A05_orf100	hypothetical 13.2 KD protein homolog (ylxM); BACSU
417	512605..512994	A05_orf129	MG296 homolog, MYCGE
418	512995..514107	A05_orf370	hypothetical protein (HI0174); HAEIN
419	514238..515665	A05_orf475	MG294 homolog (put. permease), MYCGE
420	515658..516383	A05_orf241a	glycerophosphoryl diester phosphodiesterase (glpQ); BACSU
421	516435..519137	A05_orf900	alanyl-tRNA synthetase (alaS); ECOLI
422	521188..519560	A05_orf542	transport system permease protein P99; MYCHR
423	521915..521181	A05_orf244	ATP-binding protein P29; MYCHR
424	523050..521908	A05_orf380V	high affinity transport system protein P37; MYCHR
425	524782..523301	A05_orf493	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
426	524892..525311	A05_orf139	-
	525343..523309	REPMP5	repetitive DNA sequence REPMP5
427	525388..526224	A05_orf278	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	526357..525404	REPMP4	repetitive DNA sequence REPMP4
428	526818..527576	A05_orf252	putative lipoprotein, MG440 homolog, MYCGE
	528050..527890	REPMP1	repetitive DNA sequence REPMP1
429	528364..527718	F11_orf148c	-
	528391..528045	REPMP1	repetitive DNA sequence REPMP1
430	530128..528527	F11_orf533L	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	530001..528684	REPMP2/3	repetitive DNA sequence REPMP2/3
431	532483..530201	F11_orf760	putative lipoprotein, MG260 homolog, MYCGE
432	532711..535350	F11_orf879	-
	535466..535390	nptgwa	Trp-tRNA (TGA) gene; MYCPN
433	535709..535455	F11_orf84	(acyl carrier protein; STRGA)
434	536337..535744	F11_orf197	MG286 homolog, MYCGE
435	537384..536344	F11_orf346	MG285 homolog, MYCGE
436	537733..537365	F11_orf122a	MG284 homolog, MYCGE
437	539329..537878	F11_orf483	putative prolyl-tRNA synthetase (proS); YEAST
438	539611..540093	F11_orf160	transcription elongation factor (greA); RICPR
	540123..540573	nptma	Tyr-tRNA (TAC), Glu-tRNA (CAA), Lys-tRNA (AAA), Leu-tRNA (TTA), Gly-tRNA (GGA) genes; MYCPN
439	540861..542609	F11_orf582	MG281 homolog, MYCGE
440	542631..543534	F11_orf287	MG280 homolog, MYCGE
441	543534..544190	F11_orf218	MG279 homolog, MYCGE
442	546388..544187	F11_orf733	antagen response protein (spoT); ECOLI
443	546644..549307	F11_orf887	MG277 homolog, MYCGE
444	549434..549875	F11_orf133	adenine phosphoribosyltransferase (apt); HAEIN
445	549943..551382	F11_orf479	NADH oxidase (nox); ENTFA
446	551403..552479	F11_orf358a	pyruvate dehydrogenase E1-alpha subunit (pdhA); ACHLA
447	552501..553484	F11_orf327	pyruvate dehydrogenase E1-beta subunit (pdhB); ACHLA
448	553803..555011	F11_orf402	dihydrolipoamide acetyltransferase component (E2) (pdhC); ACHLA
449	555012..556385	F11_orf457	dihydrolipoamide dehydrogenase (pdhD); BACST
450	556412..557431	F11_orf339	lipoate protein ligase (lplA); ECOLI
451	557803..558879	F11_orf358b	MG269 homolog, MYCGE
	558904..558982	4.5s RNA	4.5S RNA; MYCPN
452	559027..559716	F11_orf229	hypothetical protein (yaaF) homolog; BACSU
453	559731..560095	F11_orf114	MG267 homolog, MYCGE
454	560096..562477	F11_orf793o	leucyl-tRNA synthetase (leuS); BACSU
455	562480..563328	A19_orf282	hypothetical protein (yidA) homolog; ECOLI
456	563860..563258	A19_orf200	hypothetical protein (HI0896) homolog; HAEIN
457	564732..563854	A19_orf292	hypothetical protein (yidA) homolog; ECOLI
458	565711..564878	A19_orf277	formamidopyrimidine-DNA glycosylase (fpg); BACFI
459	566586..565711	A19_orf291	DNA polymerase I (polA, 5'-3' exonuclease) homolog; STRPN
460	569208..566590	A19_orf872	DNA polymerase III alpha subunit (dnaE); HAEIN
	569534..569598	nptga	Arg-tRNA gene (CGA); MYCPN
461	569863..573285	A19_orf1140	-
462	573664..574053	A19_orf129	-
463	574399..575088	A19_orf229V	-
464	576117..576731	A19_orf204	-
465	576517..576742	A19_orf591	-
466	578671..579306	A19_orf211	-
	579725..578587	REPMP4	repetitive DNA sequence REPMP4
	581534..580008	REPMP2/3	repetitive DNA sequence REPMP2/3
467	581562..579349	A19_orf737V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
468	582203..582964	H91_orf253	putative lipoprotein
469	583638..583096	H91_orf180	-

Table 4. *Continued*

	583663..583391	REPMP1	repetitive DNA sequence REPMP1
470	585295..584327	H91_orf322	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
471	586044..585226	H91_orf272	hypothetical 130K protein homolog (orf6, P1 operon); MYCPN
	586110..584114	REPMP5	repetitive DNA sequence REPMP5
472	586934..586128	H91_orf268	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
473	589311..587278	H91_orf677	MG260 homolog, MYCGE
474	589658..589350	H91_orf102	putative lipoprotein, MG260 homolog, MYCGE
475	591151..589790	H91_orf453	possible protoporphyrinogen oxidase (hemK); ECOLI
476	592230..591151	H91_orf359V	peptide chain release factor 1 (RF1; prfA);BACSU
477	593524..592231	H91_orf97	ribosomal protein L31 (rpL31); ECOLI
478	593345..592569	H91_orf258	MG256 homolog, MYCGE
	593426..593353	mpgg	Trp-tRNA(TGG) gene; MYCPN
479	593579..593575	H91_orf534	MG255 homolog, MYCGE
	595211..595283	mpgg	Gly-tRNA(GGC) gene; MYCPN
480	595347..597323	H91_orf658	DNA ligase (lig); ECOLI
481	597304..598617	H91_orf437	cysteinyl-tRNA synthetase (cysS); BACSU
482	598620..599348	H91_orf242a	hypothetical protein (yacO) (rRNA methylase) homolog; BACSU
483	599370..600719	H91_orf449	glycyl-tRNA synthetase (grsI); YEAST
484	600703..602565	H91_orf620	DNA primase (dnaG); BACSU
485	603618..604117	H91_orf499	RNA polymerase sigma-A factor (sigA); BACSU
486	604101..604742	H91_orf213	MG248 homolog; MYCGE
487	604748..605467	H91_orf239	hypothetical protein (ygiH) homolog; ECOLI
488	606304..605459	H91_orf281	MG246 homolog, MYCGE
489	606788..606294	H91_orf164	5-formyl tetrahydrofolate cyclo-ligase (HI0858) homolog; HAEIN
490	608873..607743	H91_orf376	Type I restriction enzyme (hsdR) homolog; ECOLI
491	609427..609080	H91_orf115	-
492	610177..609557	H91_orf206	Type I restriction enzyme (hsdR) homolog; ECOLI
493	611772..611122	H91_orf216	-
494	612987..611995	H91_orf330	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
495	614997..613366	H91_orf543	type I restriction enzyme (hsdM); ECOLI
496	617285..615138	H91_orf715	DNA helicase II (uvrD1); HAEIN
497	618937..617348	H91_orf529	DNA helicase (pcrA) homolog; STAAU
498	619615..618941	H91_orf224	MG243 homolog, MYCGE
499	621513..619615	H91_orf632c	MG242 homolog, MYCGE
500	622381..621516	H91_orf621	MG241 homolog, MYCGE
501	623425..624500	H91_orf291	MG240 homolog, MYCGE
502	626726..624501	H91_orf741	-
503	627693..626713	H91_orf326	protein (bcrA) homolog; BACLI
504	629948..627696	H91_orf750	putative ABC transport permease
505	632530..630143	H91_orf795	ATP-dependent protease (lon); BACSU
506	633935..632801	H91_orf444	trigger factor (tig); HAEIN
507	634844..633960	H91_orf294	MG237 homolog, MYCGE
508	635310..634834	H91_orf158	MG236 homolog, MYCGE
509	636124..635264	H91_orf286	endonuclease IV (nfo); ECOLI
510	636431..636117	H91_orf104	ribosomal protein L27 (rpL27); BACSU
511	636726..636424	H91_orf100a	hypothetical protein (ysxB) homolog; BACSU
512	637021..636719	H91_orf100b	ribosomal protein L21 (rpL21); BACSU
513	639333..637968	H91_orf721	ribonucleoside-diphosphate reductase (nrdE); SALTY
514	639818..639357	H91_orf153	MG230 homolog, MYCGE
515	640840..639821	H91_orf339	ribonucleotide reductase 2 (nrdF); SALTY
516	641329..640847	H91_orf160	dihydrofolate reductase (EC 1.5.1.3)(dhfr); LACLA
517	642317..641331	H91_orf328	thymidylate synthase (thyA); STAAU
518	644200..642689	H91_orf503	general amino acid permease GAP1 homolog; YEAST
519	645650..644175	H91_orf491	hypothetical protein (gi: 710640) homolog (put. amino acid permease); CLOPE
520	646835..645693	H91_orf380	cell division protein (ftsZ); BACSU
521	648100..646841	H91_orf419	MG223 homolog, MYCGE
522	649029..648103	H91_orf308	hypothetical protein (yabC) homolog; ECOLI
523	649444..649019	H91_orf141b	hypothetical protein (yabB) homolog; ECOLI
	649775..649699	mpgas	Arg-tRNA gene (CGC); MYCPN
524	649845..650117	H91_orf90	MG220 homolog, MYCGE
525	650856..650200	H91_orf218	-
526	651919..650846	H91_orf357	-
527	657390..651934	H91_orf1818	cytadherence accessory protein (hmw2); MYCPN
528	658627..657410	H91_orf405	protein P65; MYCPN
529	660458..658761	H91_orf565	-
530	661390..660461	H91_orf309	carbamate kinase (EC 2.7.2.2) (arcC); PSEAE
531	662214..661393	H10_orf273c	ornithine carbamoyl transferase (arc1); ECOLI
532	663058..662462	H10_orf198	arginine deiminase (arcA); MYCCA
533	663675..662959	H10_orf238	arginine deiminase (arcA); MYCCA
	664617..663872	mpgs	Cys-tRNA(TGC), Pro-tRNA(CCA), Met-tRNA(ATG), Ile-tRNA(ATG), Ser-tRNA(TCA), fMet-tRNA(ATG), Asp-tRNA(GAC) and Phe-tRNA(TTC) genes; MYCPN
534	666181..664655	H10_orf508	pyruvate kinase (pyk); LACLA
535	667173..666187	H10_orf328	6-phosphofructokinase (pfk); ECOLI
536	667819..667193	H10_orf208	hypothetical protein (P35155) homolog; BACSU
537	669323..667803	H10_orf506	dihyrofolate reductase (dyr) homolog protein; HNTFC
538	670124..669324	H10_orf266	1-acyl-sn-glycerol-3-phosphate acyltransferase (slc8); YEAST
539	670471..670112	H10_orf119	-
540	670923..670474	H10_orf149	MG211 homolog, MYCGE
541	671792..671130	H10_orf220L	-
542	672461..671841	H10_orf206	-
543	672500..673054	H10_orf184	prolipoprotein signal peptidase (lsp); STACA
544	673054..673983	H10_orf309	hypothetical protein (yceC) homolog; ECOLI
545	673967..674557	H10_orf196	MG208 homolog, MYCGE
546	674987..674550	H10_orf145L	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
547	675689..675126	H10_orf187V	Hsd51B protein homolog; MYCPU
548	678142..675779	A65_orf787c	putative lipoprotein, MG260 homolog, MYCGE
549	679094..678738	A65_orf118	-
	680988..679736	REPMP2/3	repetitive DNA sequence REPMP2/3
550	681222..679825	A65_orf465V	adhesin P1 (group 2) homolog; MYCPN
551	682245..681325	A65_orf306	protein (prfB) homolog, ECOLI
552	685088..682704	A65_orf794	putative lipoprotein, MG260 homolog, MYCGE
	686360..686126	REPMP1	repetitive DNA sequence REPMP1
553	686379..686032	A65_orf115	-
554	688090..687590	A65_orf166	MG260 homolog, MYCGE
555	689578..688445	A65_orf377	MG260 homolog, MYCGE
556	691488..689789	A65_orf569	MG139 homolog, MYCGE
557	693374..691629	A65_orf581	GTP-binding membrane protein (lepA); HAEIN
558	694573..693374	A65_orf399V	YrfE protein homolog; ECOLI
559	696002..694533	A65_orf489	lysyl-tRNA synthetase (lysS); BACSU
560	696047..696904	A65_orf285	MG135 homolog, MYCGE
561	697178..696876	A65_orf100	hypothetical protein (yaaK) homolog; BACSU
562	697200..698000	A65_orf266	MG133 homolog, MYCGE
563	697969..698403	A65_orf144	hypothetical protein (hit1) homolog; YEAST
564	701122..700367	A65_orf251a	putative lipoprotein, MG440 homolog, MYCGE

Table 4. *Continued*

565	703155..701674	A65_orf493	hypothetical protein (yxr1) homolog; MYCMY
566	703498..703145	A65_orf117	MG129 homolog, MYCGE
567	704277..703498	A65_orf259	hypothetical protein (HI0072) homolog; HAEIN
568	704714..704277	A65_orf145	hypothetical protein (ygl1) homolog; STRVR
569	704771..705811	A65_orf346	tryptophanyl-tRNA synthetase (trpS); HAEIN
570	706664..705819	A65_orf281	hypothetical protein (gi: 973220) homolog; ECOLI
571	706984..706676	A65_orf102	thioredoxin (trx); YEAST
572	708477..707050	A65_orf475	MG123 homolog, MYCGE
573	710602..708467	A65_orf711	DNA topoisomerase I (topA); BACSU
574	711574..710639	A65_orf311	high affinity ribose transport protein (rbsC); HAEIN
575	713127..711574	A65_orf517	MG120 homolog, MYCGE
576	714862..713144	A65_orf572	hypothetical ABC transporter (yjcW) homolog; ECOLI
577	715893..714877	A65_orf338	UDP-glucose 4-epimerase (galE); STRTR
578	716545..715874	A65_orf223	MG117 homolog, MYCGE
579	717293..716538	A65_orf251b	MG116 homolog, MYCGE
580	718497..717814	A65_orf227	phosphatidylglycerophosphate synthase (pgsA); HAEIN
581	719821..718454	K04_orf455o	asparaginyl-tRNA synthetase (asnS); ECOLI
582	720475..719828	K04_orf215L	D-ribulose-5-phosphate 3 epimerase (cfxE); ALCEU
583	721745..720453	K04_orf430	phosphoglucose isomerase B (pgiB); BACST
584	722603..721767	K04_orf278L	hypothetical protein (yjeQ) homolog; ECOLI
585	723759..722590	K04_orf389	probable protein serine/threonine kinase (YKT3); CAEEL
586	724529..723750	K04_orf259	protein phosphatase 2C homolog (ptc1); YEAST
588	725070..725720	K04_orf216	polypeptide deformylase (def); HAEIN
587	725248..724529	K04_orf239	5'guanylate kinase (gmk); HAEIN
589	726297..725689	K04_orf202	MG105 homolog, MYCGE
590	728477..726297	K04_orf726	virulence associated protein homolog (vacB); HAEIN
591	729593..728751	K04_orf280	MG103 homolog, MYCGE
592	730530..729583	K04_orf315	thioredoxin reductase (trxB); EUBAC
593	731191..730523	K04_orf222	MG101 homolog, MYCGE
594	732602..731166	G07_orf478o	protein (pet112) homolog; YEAST
595	734028..732592	G07_orf478V	amidase homolog (S47454); YEAST
596	735470..734031	G07_orf479	MG098 homolog, MYCGE
597	736390..735668	G07_orf240	uracil DNA glycosylase (ung); ECOLI
598	737668..736415	G07_orf417	MG288 homolog, MYCGE
599	739760..738396	G07_orf454	putative lipoprotein, MG095 homolog, MYCGE
600	741185..739764	G07_orf473	replicative DNA helicase (dnaC); BACSU
601	741621..741172	G07_orf149	ribosomal protein L9 (rplL9); BACST
602	741938..741624	G07_orf104b	ribosomal protein S18 (rpS18); ECOLI
603	742428..741928	G07_orf166	single-stranded DNA binding protein (ssb); HAEIN
604	743075..742428	G07_orf215	ribosomal protein S6 (rpS6); ECOLI
605	745198..743132	G07_orf688	elongation factor G (fus); THEAQ
606	745688..745221	G07_orf155	ribosomal protein S7 (rpS7); BACST
607	746161..745742	G07_orf139	ribosomal protein S12 (rpS12); BACST
608	747359..746190	G07_orf389b	prolipoprotein diacylglyceryl transferase (lgt); ECOLI
609	748287..747349	G07_orf312	MG085 homolog, MYCGE
610	749157..748288	G07_orf289	hypothetical protein (yacA) homolog; BACSU
611	749716..749150	G07_orf188	peptidyl-tRNA hydrolase homolog (pth); HAEIN
612	750396..749716	G07_orf226	ribosomal protein L1 (rplL1); BACST
613	750809..750396	G07_orf137	ribosomal protein L11 (RPL11); THEMA
614	753420..750865	G07_orf851	oligopeptide transport ATP-binding protein (oppF); BACSU
615	754654..753383	G07_orf423	oligopeptide transport ATP-binding protein (oppD); BACSU
616	755786..754656	G07_orf376	oligopeptide transport system permease protein (amiD); STRPN
617	756948..755779	G07_orf389a	oligopeptide transport system permease protein (oppB); BACSU
618	757224..757640	G07_orf138	MG076 homolog, MYCGE
619	760729..757637	G07_orf1030	protein P100; MYCPN
620	761241..760834	G07_orf135	MG074 homolog, MYCGE
621	763217..761244	G07_orf657	excinuclease ABC subunit B (uvrB); ECOLI
622	765618..763192	G07_orf808	preprotein translocase (secA); BACSU
623	768223..765605	G07_orf872V	MG(2+) transport ATPase, P-typ 1 (mgtA); ECOLI
624	769100..768216	G07_orf294	ribosomal protein S2 (rpS2); SPIPL
625	772532..769710	GT9_orf940o	PTS system, glucose-specific IIABC component (EIIABC-GLC); BACSU
626	772584..772925	GT9_orf113	-
627	774296..772980	GT9_orf438V	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	774345..773095	REPMP4	repetitive DNA sequence REPMP4
628	775203..774757	GT9_orf148	MG260 homolog, MYCGE
	775230..774929	REPMP1	repetitive DNA sequence REPMP1
629	775949..775566	GT9_orf127	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
630	776809..775868	GT9_orf313	ADP1_MYCPN adhesin P1 precursor homolog; MYCPN
	777250..775724	REPMP2/3	repetitive DNA sequence REPMP2/3
631	778005..777289	GT9_orf238	type I restriction enzyme ecoKI specificity protein (hsdS) homolog; HAEIN
632	780875..778479	GT9_orf798	putative lipoprotein, MG260 homolog, MYCGE
633	783441..781159	GT9_orf760	putative lipoprotein, MG185 homolog, MYCGE
634	784494..783535	GT9_orf319V	adenine-specific methyltransferase EcoRI (mte1); ECOLI
635	786329..784494	GT9_orf611	oligoendopeptidase F (pepF); LACLA
636	787053..786322	GT9_orf243V	pseudouridylate synthase I (hisT); ECOLI
637	788350..787046	GT9_orf434	MG181 homolog, MYCGE
638	789254..788343	GT9_orf303	histidine transport ATP-binding protein (hisP); ECOLI
639	790066..789242	GT9_orf274	sulfate transport ATP-binding protein (cysA); SYNP
640	790424..790050	GT9_orf124a	ribosomal protein L17 (rplL17); BACSU
641	791410..790427	GT9_orf327	RNA polymerase alpha core subunit (rpoA); BACSU
642	791781..791416	GT9_orf121	ribosomal protein S11 (rpS11); BACST
643	792155..791781	GT9_orf124b	ribosomal protein S13 (rpS13); BACSU
644	792268..792155	GT9_orf37	ribosomal protein L36 (rpL36); CHLTR
645	792515..792279	GT9_orf78	initiation factor 1 (infA); BACSU
646	793261..792515	GT9_orf248	methionine amino peptidase (map); BACSU
647	793908..793261	GT9_orf215	adenylate kinase (adk); BACST
648	795335..793902	GT9_orf477	preprotein translocase subunit (secY); MYCCA
649	795790..795335	GT9_orf151	ribosomal protein L15 (rpL15); MYCCA
650	796453..795794	GT9_orf219	ribosomal protein S5 (rpS5); BACSU
651	796807..796457	GT9_orf116b	ribosomal protein L18 (rpL18); BACST
652	797362..796808	GT9_orf184	ribosomal protein L6 (rpL6); MYCCA
653	797797..797369	GT9_orf142	ribosomal protein S8 (rpS8); MYCCA
654	797976..797791	GT9_orf61	ribosomal protein S14 (rpS14); MYCCA
655	798520..797978	GT9_orf180b	ribosomal protein L5 (rpL5); HAEIN
656	798858..798523	GT9_orf111a	ribosomal protein L24 (rpL24); BACST
657	799226..798858	GT9_orf122	ribosomal protein L14 (rpL14); BACST
658	799487..799230	GT9_orf85	ribosomal protein S17 (rpS17); MYCCA
659	799822..799487	GT9_orf111b	ribosomal protein L29 (rpL29); THEMA
660	800241..799822	VXpSPT7_orf139o	ribosomal protein L16 (rpL16); MYCCA
661	801062..800241	VXpSPT7_orf273	ribosomal protein S3 (rpS3); MYCCA
662	801618..801064	VXpSPT7_orf184	ribosomal protein L22 (rpL22); HAEIN
663	801808..801545	VXpSPT7_orf87	ribosomal protein S19 (rpS19); MYCBO
664	802671..801808	VXpSPT7_orf287a	ribosomal protein L2 (rpL2); MYCCA
665	803384..802671	VXpSPT7_orf237	ribosomal protein L23 (rpL23); THEMA

Table 4. *Continued*

666	804025..803387	VXpSPT7_orf212	ribosomal protein L4 (rpl.4); MYCCA
667	804888..804025	VXpSPT7_orf287b	ribosomal protein L3 (rpL3); MYCCA
668	805228..804902	VXpSPT7_orf108	ribosomal protein S10 (rpS10); THEMA
669	805660..805322	VXpSPT7_orf112	-
670	806869..805907	VXpSPT7_orf320	putative lipoprotein, MG149 homolog, MYCGE
671	808328..806991	VXpSPT7_orf445	MG148 homolog, MYCGE
672	809615..808482	VXpSPT7_orf377	MG147 homolog, MYCGE
673	810876..809602	VXpSPT7_orf424	hemolysin (hlyC) homolog protein; HAEIN
674	811711..810902	VXpSPT7_orf269	hypothetical protein (yaaC) homolog; PSEFL
675	812932..811724	VXpSPT7_orf402	MG144 homolog, MYCGE
676	813298..812948	VXpSPT7_orf116	ribosome binding factor A homolog (rbfA); ECOLI
677	815154..813301	VXpSPT7_orf617	protein synthesis initiation factor 2 (infB); BACST

noteworthy: the lack of the ribosomal protein S1, of the peptide chain release factor 2 (RF2) and of the glutaminyl-tRNA synthetase. So far, quite a number of Gram-positive bacteria including *Bacillus* or *Lactobacillus* species also lack the S1 protein and the glutaminyl-tRNA synthetase (46).

One of the functions of the S1 protein is to bind the mRNA to the 30S small ribosomal subunit. Therefore, it was argued that ribosomal binding sites in front of many genes (47) of *B.subtilis* compensate for the missing S1 protein. The Shine–Dalgarno sequences are so well conserved, that they could be used routinely as a good indicator for proposing ORFs in the *B.subtilis* genome sequencing projects, but this does not apply to *M.pneumoniae*. The Shine–Dalgarno sequence is in many instances not well conserved or missing altogether, even in genes for which we know the translational initiation sites from independent studies.

Of the 20 standard tRNA-synthetases, the glutaminyl-tRNA synthetase is the only one not detected in *M.pneumoniae*. Studies on tRNA synthetases in Gram-positive bacteria have indicated that this enzyme is dispensable. *Bacillus subtilis* solves this problem by charging the $tRNA^{Gln}$ first with glutamate which is subsequently converted to glutamine by an amido transferase. The glutamyl tRNA synthetase aminoacylates both $tRNA^{Glu}$ and $tRNA^{Gln}$. The corresponding amido transferase has not yet been identified in *M.pneumoniae*, therefore it is still an open question as to how glutamine is bound to its tRNA.

Finally, the modified codon usage by *M.pneumoniae*, reading UGA as tryptophan instead of a stop codon, requires the absence of the peptide chain release factor 2 (RF2) and the presence of the release factor 1 (RF1). The latter recognizes the stop codons UAG and UAA and RF2 the stop codons UGA and UAA. Since the UGA codon is frequently located within a gene it is essential to exclude RF2 to prevent the premature termination of proteins.

Surface structure, cytadherence-associated proteins and cell division

This category comprises the adhesins and the cytadherence associated proteins, including the components of the cytoskeleton-like structure, the function of which is probably to stabilize and maintain the shape of the wall-less mycoplasma, to direct proteins to certain regions in the membrane and to keep them in these positions (2). Adherence to the receptor(s) of the host cell depends on the tip structure. The correct assembly of the adhesin P1 (E07_orf1627) and the 30 kDa adhesin-related protein on the tip structure (H08_orf274) is necessary for attachment. The tip structure is an interesting example for bacterial cellular asymmetry (48).

The cytadherence-associated proteins were originally defined by hemadsorption-negative mutants which had lost certain proteins like the so called high molecular weight proteins HMW1, HMW2 and HMW3, the adhesin P1 and the proteins named A, B and C (2,28). B and C are most probably the gene products of the ORF6 gene of the P1 operon (40 kDa protein = C, 90 kDa protein = B). The gene for A is still unknown. Another criterion for a putative protein of the cytoskeleton-like structure is its partitioning into the Triton X-100 insoluble fraction after treating *M.pneumoniae* with this detergent. This fraction is ill defined and comprises ~50 proteins, of which only a subfraction is associated with the cytoskeleton and/or cytadherence. The following proteins have been identified as most likely components of a cytoskeleton (2): HMW1 (H08_orf1018), HMW2 (F10_orf1818; Krause, submitted), HMW3 (H08_orf672), P200 (D02_orf1036o) (49), P65 (F10_orf405) (27). These proteins, with the exception of HMW2, share some common peculiar features, like an extended acidic proline rich domain and an abnormal migration in SDS-PAGE (49). The adhesin P1 is mainly distributed in the membrane fraction and to a lesser extent in the Triton X-100 insoluble fraction (50).

A large number of proposed ORFs contain sequences with high similarities to subregions of either the P1 protein or the ORF6 gene product of the P1 operon. The coding DNA sequences correspond to the repetitive DNA sequences RepMP2/3 (P1), RepMP4 (P1) and RepMP5 (ORF6). Preliminary experiments indicate that the proposed ORFs are not expressed under standard laboratory conditions. It has been observed that another independent isolate of *M.pneumoniae*, the strain FH, carries a different copy of RepMP2/3, RepMP4 and RepMP5 in its P1 operon than the *M.pneumoniae* strain M129 which is the subject of this paper (51,52). All experimental data so far show that only the repetitive sequences which are part of the P1 operon are expressed. The exchange of these copies presumably takes place by gene conversion as was indicated by DNA sequence analysis of the corresponding RepMP5 sequences in *M.pneumoniae* strains M129 and FH. Different is the situation with RepMP1, copies of which seem to be part of several expressed proteins. RepMP1-specific antibodies recognize several proteins on western blots of *M.pneumoniae* protein extracts (26).

Only little is known about cell division in *M.pneumoniae*. The lack of mutants, especially of conditional mutants, has prevented a detailed analysis. So far, the two proteins FtsZ and FtsH are classified as cell division proteins in analogy to their function in other bacteria (53). Other genes involved in chromosome partitioning or septum formation have not been identified in *M.pneumoniae*. Interesting problems to study might include the possible interaction of FtsZ with components of the cytoskeleton-like structure, which seems to play a key role in cell division, or the effects of cellular asymmetry on cell division and the formation of daughter cells. Other genes known to be involved in cell division in *E.coli*, the muk and min genes or additional fts genes were not found in *M.pneumoniae* (53).

Lipoproteins

Altogether 46 proteins were identified as lipoproteins based on the following characteristic lipoprotein-specific features (54): (i) one or more basic amino acids among the first 5–7 amino acids of the N-terminus, (ii) a hydrophobic signal peptide and (iii) a cysteine residue immediately downstream of the signal peptide, which is available for modification by the transfer of the diacylglyceryl moiety from glycerophospholipid to its sulfhydryl group. The precursor prolipoprotein with the modified cysteine is subsequently cleaved in *M.pneumoniae* by a specific signal peptidase (signal peptidase II). The modified cysteine will then be the first amino

acid of the processed protein. The cleavage site including the cysteine and the three (positions –3, –2 and –1) upstream located amino acids, is to some extent conserved (–3: 37×L, 6×F, 1×A, 1×V; –2: 19×S, 10×A, 8×T, 6×V, 2×I; –1: 37×A, 7×S, 1×G).

The number of lipoproteins in *M.pneumoniae* is relatively high compared with the Gram-negative bacteria *E.coli* and *H.influenzae*. Even in the closely related *M.genitalium* only 21 putative lipoproteins could be found by analyses of the published data (9).

The lipoproteins of *M.pneumoniae* can be divided into six subgroups based on sequence similarities; also included in these groups are proteins with similarities to lipoproteins but without the lipoprotein signature at the N-terminal end. Quite a number of these proposed genes with high similarities are organized in tandem. For instance seven lipoproteins and one protein without the lipobox but with otherwise extended similarities are located between genome positions 249 627 and 256 463 (cosmid pcosMPE09). A gene family, with 13 proposed ORFs including five lipoproteins, is located between 306 862 and 320 524 (cosmid pcosMPD02). Presently it is unclear whether all of the proposed genes are expressed.

In vivo labelling of *M.pneumoniae* with ^{14}C-labelled palmitic acid and protein analysis by SDS–PAGE reveal, instead of the expected 46 lipoproteins, only between 20 and 25 lipoproteins (Pyrowolakis, unpublished data). This discrepancy could be explained either by a regulated expression which only allows some of the several tandemly organized lipoproteins to be synthesized or that the labelling with palmitic acid was not sensitive enough or that some lipoproteins carry fatty acids other than palmitic acid. Only four of all the proposed lipoproteins show significant similarities to other bacterial genes beside the ones from *M.genitalium*. These include A05_orf380V [high affinity transport system P37 with unknown specificity from *Mycoplasma hyorhinis* (55)], D09_orf384 (aerobic glycerol-3-phosphate dehydrogenase, glpD), H03_orf213 (uridine kinase) and D02_orf207 (ATP synthase b subunit (atpF).

The processing of the prolipoprotein to the mature lipoprotein in *E.coli* requires the three enzymes prolipoprotein diacylglyceryl transferase, prolipoprotein signal peptidase and apolipoprotein transacylase. We find in *M.pneumoniae* only the transferase which catalyzes the thioether linkage between the diacylglycerol and the cysteine and the peptidase which cleaves in front of the cysteine following the signal peptide. The transacylase could not be identified either in *M.pneumoniae* nor in *M.genitalium* (9). Therefore it is still an open question if a third fatty acid is linked to the cysteine by an amide bond as has been found for lipoproteins of *E.coli*.

The absence of a periplasmic space provides reasons for the existence of a large number of lipoproteins. For surface-exposed proteins which have to function on the outside, anchoring them via long chain fatty acids at the *M.pneumoniae* cell membrane is an efficient way. Already known examples are substrate-binding proteins of transport systems or proteins possibly involved in antigenic variation for evasion of the immune system of the host, as has been shown for other mycoplasmas (56). Nothing is known about the fate of the cleaved signal peptides, as to whether they are degraded or recycled.

Transport systems

In light of the scarcity of metabolic pathways and the marked dependence on exogenous nutrients (Table 1, Fig. 5), we expected *M.pneumoniae* to code for many transport systems to compensate for its inability to synthesize essential compounds like amino acids. Three different transport systems, mainly involved in import, were found in *M.pneumoniae*: (i) the ABC transporter system (57) consisting of two ATP-binding, two membrane-spanning and one substrate-binding domain which are frequently present on separate polypeptides, but sometimes also consist of two or three different domains located on the same peptide (D12_orf634 or D12_orf623), (ii) the phosphoenolpyruvate: carbohydrate phosphotransferase system (PTS), (58) and (iii) facilitated diffusion systems with transmembrane proteins functioning as specific carriers. *Mycoplasma pneumoniae* codes for 43 genes involved in the above mentioned transport systems according to the present status of annotation. In addition, there are several proposed proteins with 6 or 12 transmembrane segments which are candidates for membrane-spanning domains of transport systems. The relatively low number of proteins listed in Table 1 indicates that at least some of the systems might not be very substrate specific, e.g. the transport systems for amino acids. Transport systems for histidine, glutamine, an ORF showing significant similarity to a probable aromatic amino acid permease from yeast and an ABC transport system for oligopeptides were identified based on similarity of the ATP-binding domains of ABC transporters.

Surprisingly, we could not identify a transport system for the precursors for RNA and DNA synthesis, namely adenine, guanine, uracil and thymine which are essential components of mycoplasma growth media.

In this context one has to be aware of the ambiguity in the identification of ABC transport proteins on the basis of sequence similarity of the ATP-binding proteins with respect to the predicted substrate to be transported, since database searches indicate numerous candidates with different specificities but with very similar, high score values. All the annotations in this paper were done on the basis of the highest score values. Therefore it might be possible that the predicted specificity disagrees with the *in vivo* activity in *M.pneumoniae*. Additional information from similarities to transmembrane domains or the substrate-binding proteins is only rarely at hand, since, in general, similarities among these domains are not well conserved. Even in positive examples, the score values are relatively low. Sometimes additional circumstantial evidence is derived from an operon-like organisation of the genes coding for ABC transporters, e.g. the unspecified ABC transporter consisting of the proteins P69, P29 and P37 from nucleotide 519 560 to 523 050 (A05_orf542, A05_orf244 and A05_orf380V). A05_orf542 could act as the membrane-spanning domain, A05_orf244 as the ATP-binding domain and A05_orf380V, as a putative lipoprotein which could function as a substrate-binding protein. These proteins were also identified by their significant similarity to the corresponding genes in *M.hyorhinis* (55).

In *M.pneumoniae* the ABC transport system for oligopeptides consists of two different transmembrane [G07_orf376 = amiD (= oppC in *B.subtilis*); G07_orf389a = oppB] and ATP-binding domains (G07_orf851 = oppF , G07_orf423 = oppD). It is also organized in an operon-like arrangement from nucleotide 750 865 to 756 948. In striking contrast to *B.subtilis*, the substrate-binding domain (oppA) is absent in *M.pneumoniae*. Since an oppA homolog is also absent in *M.genitalium* a sequencing or annotation error seems unlikely. It remains to be experimentally determined whether the substrate-binding protein is dispensable or is part of one of the transmembrane or ATP-binding proteins.

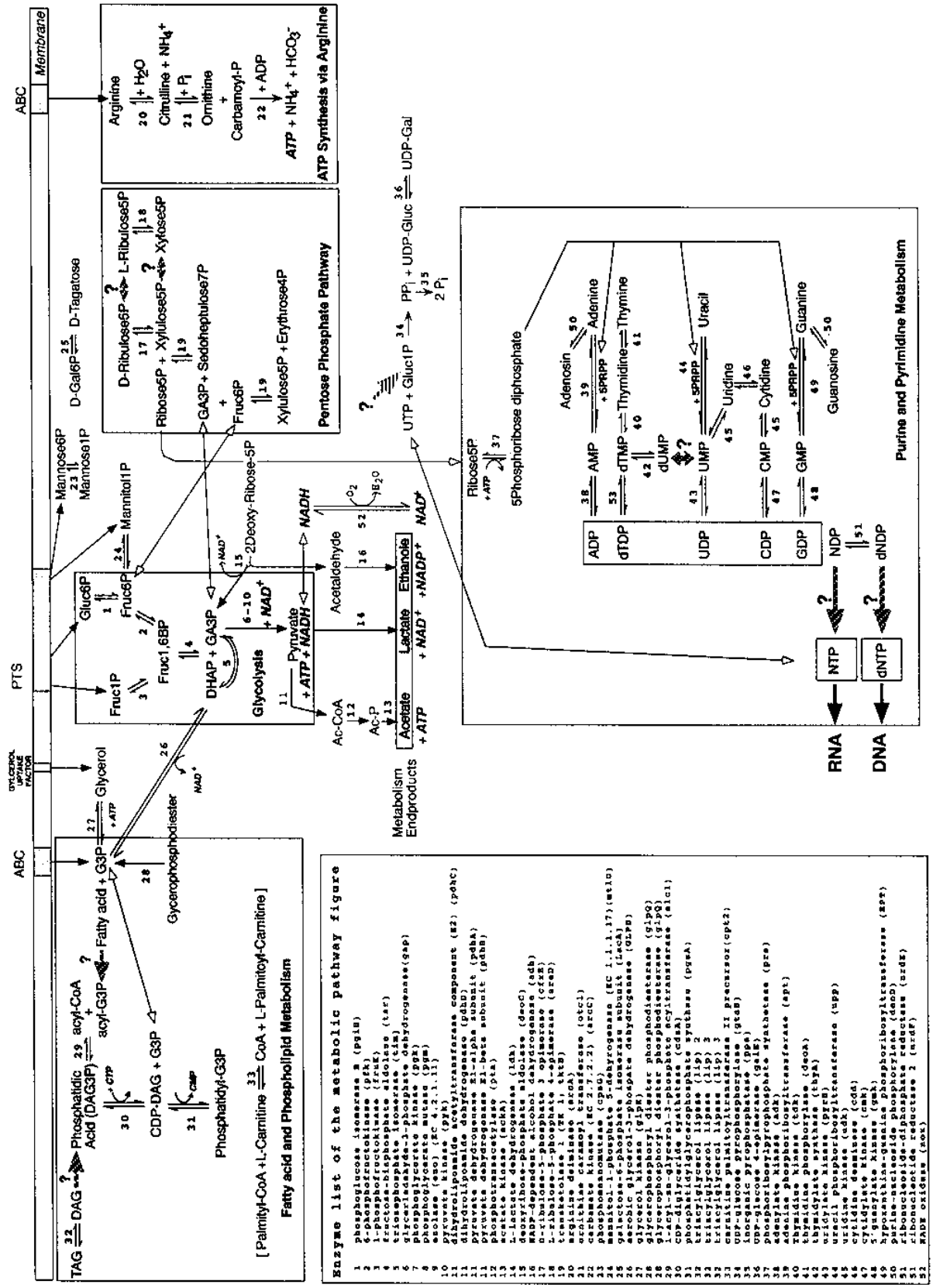


Figure 5. Schematic diagram of the metabolic pathways of *M.pneumoniae* deduced from Table 1. Shaded arrows with question marks indicate missing enzymatic activities.

It is also possible that one or more of the lipoproteins function as substrate-binding proteins.

There is also evidence for bacterial ABC export systems in *M.pneumoniae* (59). For example D12_orf634 (msbA), D12_orf623 (pmd1) and D02_orf660 (lcnDR3) have the conserved ATP binding motif and the membrane-spanning domains on the same polypeptide. In addition D12_orf623 and D12_orf634 show also significant similarities to multidrug resistance proteins of different organisms.

Among the proposed PTS transport systems, we identified one for glucose and one for mannitol. They are similar to the homologous systems from several Gram-positive bacteria, with a EIIA and EIIBC domains on two separate polypeptides for the mannitol transport system and with three domains (EIIABC) of enzyme II in one polypeptide for the glucose transport system.

Besides glucose and mannitol, fructose also seems to be imported by the PTS system. According to our data the fructose-permease II component R02_orf694 (fruA) contains all three domains of enzyme II in one gene (EIIABC). In addition, R02_orf694 and the 1-phosphofructokinase (fruK, R02_orf300) are probably in one operon, but we do not find fruF which is also part of the fructose operon in enteric bacteria (58).

Protein secretion

Both, Gram-positive and Gram-negative bacteria have a well conserved protein translocation system. The components identified which are part of the well characterized *E.coli* system (60) include cytosolic chaperones or regulators [trigger factor, SecB, DnaK, SRP (a ribonucleoprotein composed of 4.5 S RNA and Ffh) and FtsY] which deliver the protein to a membrane receptor (SecA). The receptor is also supposed to function as a motor, pushing the protein across the membrane via specific protein channels (SecY, SecG, SecE, SecD and SecF). The secreted proteins to be transported carry an N-terminal signal peptide which will be removed by a signal peptidase (SPaseI). Two routes of export have been proposed either via SecB and SecA or by SRP. The protein secretion system in *M.pneumoniae* is less complex (Table 1). So far, the trigger factor, DnaK, SRP, FtsY and SecA have been identified. From the channel-forming proteins only SecY is present but SecG, SecF, SecE, SecD and the cytosolic receptor protein SecB are missing. Also absent is the signal peptidase SPaseI although computer-assisted motif prediction programs indicate the presence of corresponding substrates (signal peptides). The simplified protein export system might be a reflection of the fact that *M.pneumoniae* is only surrounded by a cytoplasmic membrane. Another problem concerns refolding of secreted proteins which are normally exported in an unfolded stage. Refolding might be catalyzed by chaperones which have to function on the cell surface (60). This might impose a special problem on the wall-less bacteria in general, since they do not possess a periplasmic space which could prevent proteins from diffusing. To anchor the proposed chaperones on the cell surface as lipoproteins would be a possible way to solve this problem.

Nucleotide synthesis: purine and pyrimidine salvage pathways

Guanine, guanosine, uracil, thymine, thymidine, cytidine, adenine and adenosine may serve as precursors for nucleic acids and nucleotide coenzymes, as determined in nutritional studies of *Mollicutes*. These components can be used for the synthesis of ribonucleotides by the salvage pathway as predicted from the enzymes listed (Table 1, Fig. 5). The ribonucleotides are converted to deoxyribonucleotides by ribonucleoside–diphosphate reductase, an enzyme complex formed by the gene products of nrdE (F10_orf721) and nrdF (F10_orf339). Adenine, guanine and uracil can be metabolized directly to the corresponding nucleoside monophosphates by the enzymes adenine phosophoribosyltransferase (apt, F11_orf133), hypoxanthine-guanine phosphoribosyltransferase (hpt, K05_orf175) and uracil phosphoribosyltransferase (upp, B01_orf178). Uridylate, adenylate and guanylate kinases catalyze the generation of ADP, GDP and UDP. Surprisingly, we could not find the nucleoside diphosphate kinase (ndk), the key enzyme for the conversion from NDP to NTP. This finding is in agreement with data from the genomic sequence analysis of *M.genitalium*.

Another important enzyme, the CTP synthetase which converts UTP to CTP is also missing. Therefore the only route for the synthesis of CTP appears to be from cytidine to CMP by uridine kinase (H03_orf213) and to CDP by cytidylate kinase (P01_orf217). Deoxythymidine monophosphate (dTMP) could be either synthesized by thymidine kinase (tdk, B01_orf191) or by thymidylate synthase (thA, F10_orf328).

It will be of special interest to experimentally identify the enzyme(s) of *M.pneumoniae* which convert NDPs to NTPs, since such an enzymatic activity seems to be essential.

Carbohydrate metabolism and energy conservation

The ability to metabolize glucose and/or arginine and use it for the ATP synthesis is one of the key features in classification of *Mollicutes*. *Mycoplasma pneumoniae* is listed in Bergey's manual of systematic bacteriology as a glucose fermenter but not as an arginine-hydrolyzing species (61). This contrasts with our sequencing results, since the three enzymes involved in the arginine degradation pathway, arginine deiminase (H03_orf438), ornithine carbamoyltransferase (H10_orf273) and carbamate kinase (F10_orf309) are present according to our sequence data. The arginine deiminase gene occurs twice but one copy is inactive due to a raster-mutation resulting in two proposed ORFs (H10_orf198 and H10_orf238) corresponding to the N-terminal and C-terminal halves of a complete deiminase. The change in reading frame was also confirmed by sequencing of directly amplified genomic DNA. All these proposed ORFs are organized in an operon-like arrangement except for the deiminase (H03_orf438) which seems to be expressed as a single gene located far away from the mentioned operon. Included in this operon is a proposed protein (F10_orf565) with 12 predicted transmembrane domains indicative of a putative permease.

Glucose, fructose and mannitol are transported by the PTS system into the cell and further degraded by the Embden–Meyerhof–Parnas (EMP) pathway to pyruvate. All enzymes required for this pathway have been identified. The second pathway for metabolizing glucose, the pentose phosphate pathway, is incomplete in *M.pneumoniae*. We found only the enzymes ribulose-5-phosphate-3-epimerase and transketolase (Fig. 5). Glucose-6-phosphate dehydrogenase (G6Pde), 6-phospho-gluconate dehydrogenase (6PGde), and a transaldolase are missing. These data agree with enzymatic studies showing that G6Pde and 6PGde are absent in mycoplasmas (62).

Pyruvate can be further metabolized by two alternative reactions, either to lactate by lactate dehydrogenase (K05_orf312) or to acetyl-CoA by the pyruvate dehydrogenase complex and further to acetate by the phosphotransacetylase (A05_orf320, pta) and the acetate kinase (G12_orf390, ackA). The pyruvate dehydrogenase complex consists of E1α (F11_orf358a) E1β (F11_orf327), the two subunits of the pyruvate dehydrogenase, the dihydrolipoamide acetyltransferase E2 (F11_orf402) and the dihydrolipoamide dehydrogenase E3 (F11_orf457). The corresponding genes are clustered (nt 549 943–557 431; pcosMPF11); part of this cluster also contains the genes coding for NADH oxidase (nox, F11_orf479) and lipoate protein ligase (lplA, F11orf339). The later enzyme joins lipoic acid in an amide linkage to the ε amino group of a lysine residue of the dihydrolipoamide acetyltransferase.

Membrane phospho- and glycolipid synthesis

In *M.pneumoniae* strain FH the following membrane phospho- and glycolipids have been found: digalactosyldiacylglycerol, trigalactosyldiacylglycerol, glucosylgalactosyldiacylglycerol, phosphatidylglycerol (PG) and diphosphatidylglycerol (DPG) (63). Since *M.pneumoniae* FH and *M.pneumoniae* M129 are very similar we assume that both strains carry essentially the same genes for phospho- and glycolipid-synthesis.

About 10 genes are required for the synthesis of the above-mentioned lipids; but according to our DNA sequence analysis only three of the expected genes could be unambiguously identified. They code (Fig. 5) for the enzymes 1-acylglycerol-3-phosphate acyltransferase (plsC; gene name in *Saccharomyces cerevisiae* is slc1), phosphatidic acid cytidyltransferase (cdsA) and glycerolphosphate phosphatidyltransferase (pgsA). These enzymes are involved in the biochemical pathway for the synthesis of PG and DPG. Missing are the glycerol-3-phosphate acyltransferase (plsB) catalysing the synthesis of 1-acylglycerol-3-phosphate (acyl-G3P) from glycerol-3-phosphate (G3P), the phosphatidylglycerol phosphate phosphatase which converts phosphatidylglycerol-3-phosphate to PG and finally the cardiolipin synthetase (cls) which synthesizes DPG from PG. Interestingly, we find a gene homologous to the plsX gene from *E.coli* which is involved in membrane lipid synthesis in an undefined manner. The glycolipid synthesis could start with phosphatidic acid and would probably require a phosphatidic acid phosphatase and several UDP-glucosyl- or galactosyltransferases. None of these enzymes could be identified by similarity searches in databases.

As expected from biochemical studies no gene involved in fatty acid or cholesterol synthesis was determined in the sequence analysis. These components are incorporated as such from the medium.

An interesting enzyme is the proposed carnitine palmitoyl-transferase encoded by C09_orf600, which might be involved in the modifacation of exogenous phosphatidylcholine (67).

CONCLUSIONS

It is impossible to address each proposed *M.pneumoniae* gene in this paper. We have tried to cover the most important categories of functions and point to genes which should be present, but could not be found by our applied methods. Typical examples are the missing diphosphonucleoside kinase for the conversion of (d)NDPs to (d)NTPs, and the substrate binding domain (oppA) for the oligopeptide ABC transporter. In addition, we could not find any indication for a number of genes/proteins, which should be there based on experimental evidence. *Mycoplasma pneumoniae* has been shown to be motile and to exhibit chemotactic behaviour (64). Motility genes are difficult to identify since the motility in *M.pneumoniae* is independent of pili or flagella and it is not yet known which are potential candidates. Therefore, any progress in this field depends on the isolation of mutants. Furthermore, none of the components of the chemotactic signal pathway, the Che proteins, which are well conserved among bacteria, or any other 'two-component signal transduction system' could be detected. Chemotactic behaviour in *M.pneumoniae* is difficult to study. While it might be possible that these bacteria are chemotaxis negative, only additional experiments will clarify this point.

It has been reported that *M.pneumoniae* produces hydrogen peroxide considered to be a pathogenicity factor (17). Therefore, to protect itself from oxidative stress one would expect to find the standard enzymes dealing with these stress factors like catalase, superoxide dismutase or peroxidase, but we have no similarity based evidence that these enzymes exist in *M.pneumoniae*. Experimental data on this topic are also inconsistent (62).

The results of our sequence analysis explain quite well the kind of changes which have led to the observed reduction of the genome size in *M.pneumoniae* from the presumed genome size of several million base pairs of the ancestral bacteria. The main cause is the loss of complete anabolic (no amino acid synthesis) and metabolic pathways and of genes for the synthesis of complex structures like the bacterial cell wall which requires a large number of genes. In addition, for several processes like DNA repair, DNA recombination, cell division or protein secretion, the number of genes involved is smaller than in the more complex bacteria.

No significant changes were observed in the size of individual genes which resemble more or less their counterparts in *E.coli* or *B.subtilis*. The occasionally observed smaller intergenic regions, like those found in the ATPase operon, do not appear to significantly contribute to the overall genome size reduction.

In contrast with the loss of complete pathways we frequently observed the amplification of complete genes or segments of genes (see sections on lipoprotein families or on the repetitive DNA sequences RepMP2/3, RepMP4 and RepMP5). In these two instances the obvious advantage would be the potential of expressing antigenic variants of surface-exposed proteins.

The various truncated genes which are also present in full length copies e.g. arginine deiminase (H03_orf438 and H03_orf238), DNA primase (H91_orf620 and D12_orf212) and the dihydrofolate reductase (H10_orf506 and F10_orf160) might be relics of recombination events which took place in the course of the process of evolution.

Finally among the many proposed proteins are a few which share the highest similarity over their entire length with a eukaryotic protein. The most prominent examples are the pre-B cell enhancing factor (pbeF, D09_orf451) and the carnitine palmitoyltransferase II precursor (cpt2, C09_orf600). Both might be candidates for examples of horizontal gene transfer, but at the present state of analysis a definitive answer cannot be given.

It will be the main task of future studies to reconcile the experimental evidence and the DNA sequence-based predictions, i.e. to indentify the genes for observed functions and vice versa, and to assign functions to proposed open reading frames with hitherto unknown functions.

One obvious topic is the comparative analysis between the completely sequenced genomes of the closely related species *M.pneumoniae* and *M.genitalium* (9). Since the present paper is already very voluminous we decided to publish this analysis in an additional paper (Himmelreich *et al*., in preparation).

ACKNOWLEDGEMENTS

We thank R. Frank and A. Bosserhoff for the synthesis of oligonucleotides, B. Reiner for her expertise in computer data analysis, Raphael Mosbach for his technical assistance concerning hardware problems, U. Leibfried for technical assistance, I. Schmidt for preparing the manuscript, D. Hofmann and H. Göhlmann for reading of the manuscript and H. Schaller for financial assistance and his encouragement throughout our work. We thank S. Razin, A. Wieslander, K. Dybvig, K. Sitaraman, R. Walker, H. Neimark and R. Miles who read drafts of this publication. Their corrections, critical comments and suggestions helped us very much. This research was supported by a grant from the Deutsche Forschungsgemeinschaft (He 780/5-1–He 780/5-4) and by the Fonds der Chemischen Industrie.

REFERENCES

1 Chanock, R. M., Dienes, L., Eaton, M. D., Edward, D. G., Freundt, E. A., Hayflick, L., Hers, J. F. P., Jensen, K. E., Liu, C., Marmion, B. P., Morton, H. E., Mufson, M. A., Smith, P. F., Somerson, N. L. and Taylor-Robinson, D. (1963) *Science*, **140,** 662.
2 Krause, D. C. (1996) *Mol. Microbiol*., **20,** 247–253.
3 Jacobs, E. (1991) *Rev. Med. Microbiol*., **2,** 83–90.
4 Dybvig, K. (1990) *Annu. Rev. Microbiol*., **44,** 81–104.
5 Morowitz, H. J. (1984) *Isr. J. Med. Sci*., **20,** 750–753.
6 Razin, S. (1992) *FEMS Microbiol Lett*, **100,** 423–431.
7 Bove, J. M. (1993) *Clin. Infect. Dis*., **17 Suppl 1,** 10–31
8 Peterson, S. N., Hu, P. C., Bott, K. F. and Hutchison, C. A. d. (1993) *J. Bacteriol*., **175,** 7918–7930
9 Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M. *et al*. (1995) *Science*, **270,** 397–403.
10 Hilbert, H., Himmelreich, R., Plagens, H. and Herrmann, R. (1996) *Nucleic Acids Res*., **24,** 628–639.
11 Bork, P., Ouzounis, C., Casari, G., Schneider, R., Sander, C., Dolan, M., Gilbert, W. and Gillevet, P. M. (1995) *Mol. Microbiol*., **16,** 955–967
12 Sterky, F., Holmberg, A. and Uhlen, M. (1996) HUGO'96, Heidelberg, Germany.
13 Glass, J. L., Glass, J. S., Lefkowitz, E. J., Chen, E. Y. and Cassel, G. H. (1996) IOM Letters, USA, Vol. 4, pp. 12., Proc. Meet. Int. Org. Mycoplasm., Orlando, Florida.
14 Inamine, J. M., Loechel, S. and Hu, P. C. (1988) *Gene*, **73,** 175–183.
15 Wenzel, R. and Herrmann, R. (1989) *Nucleic Acids Res*., **17,** 7029–7043.
16 Su, C. J., Chavoya, A. and Baseman, J. B. (1988) *Infect Immunol*., **56,** 3157–3161.
17 Almagor, M., Yatziv, S. and Kahane, I. (1983) *Infect. Immunol*., **41,** 251–256.
18 Sanger, F., Nicklen, R. and Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA*, **79,** 5463–5467.
19 Bairoch, A. and Boeckmann, B. (1991) *Nucleic Acids Res*., **19,** 2247–2249.
20 Barker, W. C., George, D. G., Mewes, H.-W., Pfeiffer, F. and Tsugita, A. (1993) *Nucleic Acids Res*., **21,** 3089–3092.
21 Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl Acad. Sci. USA*, **85,** 2444–2448.
22 Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) *J. Mol. Biol*., **215,** 403–410.
23 Bairoch, A. (1992) *Nucleic Acids Res*., **20,** 2013–2018.
24 Inamine, J. M., Ho, K. C., Loechel, S. and Hu, P. C. (1990) *J. Bacteriol*., **172,** 504–506.
25 Nakai, K. and Kanehisa, M. (1991) *Proteins: Struct., Funct. Genet*., **11,** 95–110.
26 Proft, T. and Herrmann, R. (1994) *Mol. Microbiol*., **13,** 337–348.
27 Proft, T., Hilbert, H., Layh Schmitt, G. and Herrmann, R. (1995) *J. Bacteriol*., **177,** 3370–3378.
28 Razin, S. and Jacobs, E. (1992b) *J. Gen. Microbiol*., **138,** 407–422.
29 Ruland, K., Wenzel, R. and Herrmann, R. (1990) *Nucleic Acids Res*., **18,** 6311–6317.
30 Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. *et al*. (1995) *Science*, **269,** 496–512.
31 Riley, M. (1993) *Microbiol. Rev*., **57,** 862–952.
32 Baker, T. A. and Wickner, S. H. (1992) *Annu. Rev. Genet*., **26,** 447–477.
33 Mills, L. B., Stanbridge, E. J., Sedwick, W. D. and Korn, D. (1977) *J. Bacteriol*., **132,** 641–649.
34 Barnes, M. H., Tarantino, P. M., Jr., Spacciapoli, P., Brown, N. C., Yu, H. and Dybvig, K. (1994) *Mol. Microbiol*., **13,** 843–54
35 Koonin, E. V. and Bork, P. (1996) *Trends Biochem. Sci*., **21,** 128–129.
36 Camerini-Otero, R. D. and Hsieh, P. (1995) *Annu. Rev. Genet*., **29,** 509–552.
37 Demple, B. and Harrison, L. (1994) *Annu. Rev. Biochem*., **63,** 915–948.
38 Sancar, A. and Sancar, G. B. (1988) *Annu. Rev. Biochem*., **57,** 29–67.
39 Haldenwang, W. G. (1995) *Microbiol. Rev*., **59,** 1–30
40 Hyman, H. C., Gafny, R., Glaser, G. and Razin, S. (1988) *J. Bacteriol*., **170,** 3262–3268.
41 Moran, C. P. j., Lang, N., LeGrice, S. F. J., Lee, G., Stephens, M., Sonnenshein, A. L., Pero, J. and Losik, R. (1982) *Mol. Gen. Genet*., **186,** 339–346.
42 Das, A. (1993) *Annu. Rev. Biochem*., **62,** 893–930.
43 Hecker, M., Schumann, W. and Voelker, U. (1996) *Mol. Microbiol*., **19,** 417–428.
44 Parkinson, J. S. (1993) *Cell*, **73,** 857–871.
45 Simoneau, P., Li, C. M., Loechel, S., Wenzel, R., Herrmann, R. and Hu, P. C. (1993) *Nucleic Acids Res*, **21,** 4967–4974.
46 Breton, R., Watson, D., Yaguchi, M. and Lapointe, J. (1990) *J. Biol. Chem*., **265,** 18248–18255.
47 Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, **71,** 1342–1346.
48 Shapiro, L. (1993) *Cell*, **73,** 841–855.
49 Proft, T., Hilbert, H., Plagens, H. and Herrmann, R. (1996) *Gene*, **171,** 79–82.
50 Kahane, I., Tucker, S., Leith, D. K., Morrison, P. J. and Baseman, J. B. (1985) *Infect. Immunol*., **50,** 944–946.
51 Su, C. J., Chavoya, A., Dallo, S. F. and Baseman, J. B. (1990) *Infect. Immunol*., **58,** 2669–2674.
52 Ruland, K., Himmelreich, R. and Herrmann, R. (1994) *J. Bacteriol*., **176,** 5202–5209
53 Vicente, M. and Errington, J. (1996) *Mol. Microbiol*., **20,** 1–7.
54 Sankaran, K., Gupta, S. D. and Wu, H. C. (1995) *Methods Enzymol*, **250,** 683–697
55 Gilson, E., Alloing, G., Schmidt, T., Claverys, J. P., Dudler, R. and Hofnung, M. (1988) *EMBO J*., **7,** 3971–3974.
56 Citti, C. and Wise, K. S. (1995) *Mol. Mircobiol*., **18,** 649–660.
57 Higgins, C. F. (1992) *Annu. Rev. Cell Biol*., **8,** 67–113.
58 Postma, P. W., Lengeler, J. W. and Jacobson, G. R. (1993) *Microbiol. Rev*., **57,** 543–594
59 Fath, M. J. and Kolter, R. (1993) *Microbiol. Rev*., **57,** 995–1017.
60 Schatz, G. and Dobberstein, B. (1996) *Science*, **271,** 1519–1526
61 Freundt, E. A. and Razin, S. (1984) In Krieg, N. R. and Holt, J. G. e. (eds), *Bergey's Manual of Systematic Bacteriology*, Vol. 1. Williams and Wilkins, Baltimore, pp. 742–770.
62 Pollack, J. D. (1992) In Maniloff, J., McElhaney, R. N., Finch, L. R. and Baseman, J. B. e. (eds), *Mycoplasmas—Molecular Biology and Pathogenesis*. American Society for Microbiology, Washington, DC, pp. 181–200.
63 Plackett, P., Marmion, B. P., Shaw, E. J. and Lemke, R. M. (1969) *Aust. J. Exp. Biol. Med. Sci*., **47,** 171–195.
64 Kirchhoff, H. (1992) In Maniloff, J., McElhaney, R. N., Finch, L. R. and Baseman, J. B. e. (eds.), *Mycoplasmas—Molecular Biology and Pathogenesis*. Americam Society for Microbiology, Washington, DC, pp. 289–308.
65 Matic, I., Rayssiguier, C. and Radman M. (1995) *Cell*, **80,** 507–515.
66 Atkins, J. F. and Gesteland, R. F. (1996) *Nature*, **379,** 769–771
67 Rottem, S., Adar, L., Gross, Z., Ne'Eman, Z. and Davis, P. J. (1986) *J. Bacteriol*., **167,** 299–304.