Received 7 July; accepted 21 September 1998.


1. Dalbey, R. E., Lively, M. O., Bron, S. & van Dijl, J. M. The chemistry and enzymology of the type 1 signal peptidases. *Protein Sci.* **6,** 1129–1138 (1997).
2. Kuo, D. W. *et al*. *Escherichia coli* leader peptidase: production of an active form lacking a requirement for detergent and development of peptide substrates. *Arch. Biochem. Biophys.* **303,** 274–280 (1993).
3. Tschantz, W. R. *et al*. Characterization of a soluble, catalytically active form of *Escherichia coli* leader peptidase: requirement of detergent or phospholipid for optimal activity. *Biochemistry* **34,** 3935–3941 (1995).
4. Allsop, A. E. *et al*. in *Anti-Infectives, Recent Advances in Chemistry and Structure-Activity Relationships* (eds Bently, P. H. & O'Hanlon, P. J.) 61–72 (R. Soc. Chem., Cambridge, 1997).
5. Black, M. T. & Bruton, G. Inhibitors of bacterial signal peptidases. *Curr. Pharm. Des.* **4,** 133–154 (1998).
6. Date, T. Demonstration by a novel genetic technique that leader peptidase is an essential enzyme in *Escherichia coli*. *J. Bacteriol.* **154,** 76–83 (1983).
7. Whitely, P. & von Heijne, G. The DsbA-DsbB system affects the formation of disulfide bonds in periplasmic but not in intramembraneous protein domains. *FEBS Lett.* **332,** 49–51 (1993).
8. Peat, T. S. *et al*. Structure of the UmuD′ protein and its regulation in response to DNA damage. *Nature* **380,** 727–730 (1996).
9. Paetzel, M. *et al*. Crystallization of a soluble, catalytically active form of *Escherichia coli* leader peptidase. *Proteins Struct. Funct. Genet.* **23,** 122–125 (1995).
10. van Klompenburg, W. *et al*. Phosphatidylethanolamine mediated insertion of the catalytic domain of leader peptidase in membranes. *FEBS Lett.* **431,** 75–79 (1998).
11. Kim, Y. T., Muramatsu, T. & Takahashi, K. Identification of Trp 300 as an important residue for *Escherichia coli* leader peptidase activity. *Eur. J. Biochem.* **234,** 358–362 (1995).
12. Landolt-Marticorena, C., Williams, K. A., Deber, C. M. & Reithmeirer, R. A. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J. Mol. Biol.* **229,** 602–608 (1993).
13. James, M. N. G. in *Proteolysis and Protein Turnover* (eds Bond, J. S. & Barrett, A. J.) 1–8 (Portland, Brookfield, VT, 1994).
14. Strynadka, N. C. J. *et al*. Molecular structure of the acyl-enzyme intermediate in β-lactamase at 1.7 Å resolution. *Nature* **359,** 393–400 (1992).
15. Manard, R. & Storer, A. C. Oxyanion hole interactions in serine and cysteine proteases. *Biol. Chem. Hoppe-Seyler* **373,** 393–400 (1992).
16. Nicolas, A. *et al*. Contribution of cutinase Ser 42 side chain to the stabilization of the oxyanion transition state. *Biochemistry* **35,** 398–410 (1996).
17. Paetzel, M. *et al*. Use of site-directed chemical modification to study an essential lysine in *Escherichia coli* leader peptidase. *J. Biol. Chem.* **272,** 9994–10003 (1997).
18. Paetzel, M. & Dalbey, R. E. Catalytic hydroxyl/amine dyads with serine proteases. *Trends Biochem. Sci.* **22,** 28–31 (1997).
19. von Heijne, G. Signal sequences. The limits of variation. *J. Mol. Biol.* **184,** 99–105 (1985).
20. Izard, J. W. & Kendall, D. A. Signal peptides: exquisitely designed transport promoters. *Mol. Microbiol.* **13,** 765–773 (1994).
21. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497 (1968).
22. Otwinowski, Z. in *DENZO* (eds Sawyer, L., Isaacs, N. & Baily, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
23. Collaborative Computational Project No. 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
24. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kieldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
25. Brunger, A. T. *X-PLOR: A System for X-ray Crystallography and NMR (Version 3.1)* (Yale Univ. Press, New Haven, 1987).
26. Tronrud, D. E. Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crystallogr. A* **48,** 912–916 (1992).
27. Wolfe, P. B., Wickner, W. & Goodman, J. M. Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. *J. Biol. Chem.* **258,** 12073–12080 (1983).
28. Kraulis, P. G. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
29. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and the thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11,** 281–296 (1991).
30. Meritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).


**Acknowledgements.** We thank SmithKlineBeecham Pharmaceuticals for penem inhibitor; R. M. Sweet for use of beamline X12C (NSLS, Brookhaven National Laboratory); G. Petsko for the ethylmercury phosphate; M. N. G. James for access to equipment for characterization of earlier crystal forms of SPase; and S. Mosimann and S. Ness for discussions. This work was supported by the Medical Research Council of Canada, the Canadian Bacterial Diseases Network of Excellence, and British Columbia Medical Research Foundation grants to N.C.J.S. M.P. is funded by an MRC of Canada post-doctoral fellowship, N.C.J.S. by an MRC of Canada scholarship, and R.E.D. by the NIH and the American Heart Association.

Correspondence and requests for materials should be addressed to N.C.J.S. (e-mail: natalie@byron.biochem.ubc.ca).

## errata

# Reconciling the spectrum of Sagittarius A* with a two-temperature plasma model

**Rohan Mahadevan**

*Nature* **394,** 651–653 (1998)

A misleading typographical error was introduced into the second sentence of the bold introductory paragraph of this Letter: the word "infrared" should be "inferred". □

# Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence

**S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead & B. G. Barrell**

*Nature* **393,** 537–544 (1998)

As a result of an error during film output, Table 1 was published with some symbols missing. The correct version can be found at http://www.sanger.ac.uk and is reproduced again here (following pages).

Also, in Fig. 2, we incorrectly labelled Rv0649 as *fadD37* instead of *fabD2*. Two of the genes for mycolyl transferases were inverted: Rv0129c encodes antigen 85C and not 85C′ as stated, whereas Rv3803c codes for the secreted protein MPT51 and not antigen 85C (*Infect. Immun.* **59,** 372–382; 1991); Rv3803c is now designated *fbpD*. We thank Morten Harboe and Harald Wiker for drawing this to our attention.

The sequence of Rv0746 from *M. bovis* BCG-Pasteur presented in Fig. 5b was incorrect and should have shown a 16-codon deletion instead of 29, as indicated here:

```
H37Rv.....GSGAPGGAGGAAGLWGTGGAGGAGGSSAGGGGAGGAGGAGGWLLGDGGAGGIGGAST...
..........:::::::::::::::::::::::          ::::::::::::::::::::
BCG.......GSGAPGGAGGAAGLWGTGGA----------------GGAGGWLLGDGGAGGIGGAST...
```

□

## Table 1. Functional classification of *Mycobacterium tuberculosis* protein-coding genes

| | | |
|---|---|---|
| **I. Small-molecule metabolism** | | |
| *A. Degradation* | | |
| 1. Carbon compounds | | |
| Rv0186 | *bglS* | β-glucosidase |
| Rv2202c | *cbhK* | carbohydrate kinase |
| Rv0727c | *fucA* | L-fuculose phosphate aldolase |
| Rv1731 | *gabD1* | succinate-semialdehyde dehydrogenase |
| Rv0234c | *gabD2* | succinate-semialdehyde dehydrogenase |
| Rv0501 | *galE1* | UDP-glucose 4-epimerase |
| Rv0536 | *galE2* | UDP-glucose 4-epimerase |
| Rv0620 | *galK* | galactokinase |
| Rv0619 | *galT* | galactose-1-phosphate uridylyltransferase C-term |
| Rv0618 | *galT'* | galactose-1-phosphate uridylyltransferase N-term |
| Rv0993 | *galU* | UTP-glucose-1-phosphate uridylyltransferase |
| Rv3696c | *glpK* | ATP:glycerol 3-phosphotransferase |
| Rv3255c | *manA* | mannose-6-phosphate isomerase |
| Rv3441c | *mrsA* | phosphoglucomutase or phosphomannomutase |
| Rv0118c | *oxcA* | oxalyl-CoA decarboxylase |
| Rv3068c | *pgmA* | phosphoglucomutase |
| Rv3257c | *pmmA* | phosphomannomutase |
| Rv3308 | *pmmB* | phosphomannomutase |
| Rv2702 | *ppgK* | polyphosphate glucokinase |
| Rv0408 | *pta* | phosphate acetyltransferase |
| Rv0729 | *xylB* | xylulose kinase |
| Rv1096 | - | carbohydrate degrading enzyme |
| 2. Amino acids and amines | | |
| Rv1905c | *aao* | D-amino acid oxidase |
| Rv2531c | *adi* | ornithine/arginine decarboxylase |
| Rv2780 | *ald* | L-alanine dehydrogenase |
| Rv1538c | *ansA* | L-asparaginase |
| Rv1001 | *arcA* | arginine deiminase |
| Rv0753c | *mmsA* | methylmalmonate semialdehyde dehydrogenase |
| Rv0751c | *mmsB* | methylmalmonate semialdehyde oxidoreductase |
| Rv1187 | *rocA* | pyrroline-5-carboxylate dehydrogenase |
| Rv2322c | *rocD1* | ornithine aminotransferase |
| Rv2321c | *rocD2* | ornithine aminotransferase |
| Rv1848 | *ureA* | urease γ subunit |
| Rv1849 | *ureB* | urease β subunit |
| Rv1850 | *ureC* | urease α subunit |
| Rv1853 | *ureD* | urease accessory protein |
| Rv1851 | *ureF* | urease accessory protein |
| Rv1852 | *ureG* | urease accessory protein |
| Rv2913c | - | probable D-amino acid aminohydrolase |
| Rv3551 | - | possible glutaconate CoA-transferase |
| 3. Fatty acids | | |
| Rv2501c | *accA1* | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv0973c | *accA2* | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv2502c | *accD1* | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv0974c | *accD2* | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv3667 | *acs* | acetyl-CoA synthase |
| Rv3409c | *choD* | cholesterol oxidase |
| Rv0222 | *echA1* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0456c | *echA2* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0632c | *echA3* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0673 | *echA4* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0675 | *echA5* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0905 | *echA6* | enoyl-CoA hydratase/isomerase superfamily (aka *eccH*) |
| Rv0971c | *echA7* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1070c | *echA8* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1071c | *echA9* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1142c | *echA10* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1141c | *echA11* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1472 | *echA12* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1935c | *echA13* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2486 | *echA14* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2679 | *echA15* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2831 | *echA16* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3039c | *echA17* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3373 | *echA18* | enoyl-CoA hydratase/isomerase superfamily, N-term |
| Rv3374 | *echA18'* | enoyl-CoA hydratase/isomerase superfamily, C-term |
| Rv3516 | *echA19* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3550 | *echA20* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3774 | *echA21* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0859 | *fadA* | β oxidation complex, β subunit (acetyl-CoA C-acetyltransferase) |
| Rv0243 | *fadA2* | acetyl-CoA C-acetyltransferase |
| Rv1074c | *fadA3* | acetyl-CoA C-acetyltransferase |
| Rv1323 | *fadA4* | acetyl-CoA C-acetyltransferase (aka *thiL*) |
| Rv3546 | *fadA5* | acetyl-CoA C-acetyltransferase |
| Rv3556c | *fadA6* | acetyl-CoA C-acetyltransferase |
| Rv0860 | *fadB* | β oxidation complex, α subunit (multiple activities) |
| Rv0468 | *fadB2* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1715 | *fadB3* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv3141 | *fadB4* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1912c | *fadB5* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1750c | *fadD1* | acyl-CoA synthase |
| Rv0270 | *fadD2* | acyl-CoA synthase |
| Rv3561 | *fadD3* | acyl-CoA synthase |
| Rv0214 | *fadD4* | acyl-CoA synthase |
| Rv0166 | *fadD5* | acyl-CoA synthase |
| Rv1206 | *fadD6* | acyl-CoA synthase |
| Rv0119 | *fadD7* | acyl-CoA synthase |
| Rv0551c | *fadD8* | acyl-CoA synthase |
| Rv2590 | *fadD9* | acyl-CoA synthase |
| Rv0099 | *fadD10* | acyl-CoA synthase |
| Rv1550 | *fadD11* | acyl-CoA synthase, N-term |
| Rv1549 | *fadD11'* | acyl-CoA synthase, C-term |
| Rv1427c | *fadD12* | acyl-CoA synthase |
| Rv3089 | *fadD13* | acyl-CoA synthase |
| Rv1058 | *fadD14* | acyl-CoA synthase |
| Rv2187 | *fadD15* | acyl-CoA synthase |
| Rv0852 | *fadD16* | acyl-CoA synthase |
| Rv3506 | *fadD17* | acyl-CoA synthase |
| Rv3513c | *fadD18* | acyl-CoA synthase |
| Rv3515c | *fadD19* | acyl-CoA synthase |
| Rv1185c | *fadD21* | acyl-CoA synthase |
| Rv2948c | *fadD22* | acyl-CoA synthase |
| Rv3826 | *fadD23* | acyl-CoA synthase |
| Rv1529 | *fadD24* | acyl-CoA synthase |
| Rv1521 | *fadD25* | acyl-CoA synthase |
| Rv2930 | *fadD26* | acyl-CoA synthase |
| Rv0275c | *fadD27* | acyl-CoA synthase |
| Rv2941 | *fadD28* | acyl-CoA synthase |
| Rv2950c | *fadD29* | acyl-CoA synthase |
| Rv0404 | *fadD30* | acyl-CoA synthase |
| Rv1925 | *fadD31* | acyl-CoA synthase |
| Rv3801c | *fadD32* | acyl-CoA synthase |
| Rv1345 | *fadD33* | acyl-CoA synthase |
| Rv0035 | *fadD34* | acyl-CoA synthase |
| Rv2505c | *fadD35* | acyl-CoA synthase |
| Rv1193 | *fadD36* | acyl-CoA synthase |
| Rv0131c | *fadE1* | acyl-CoA dehydrogenase |
| Rv0154c | *fadE2* | acyl-CoA dehydrogenase |
| Rv0215c | *fadE3* | acyl-CoA dehydrogenase |
| Rv0231 | *fadE4* | acyl-CoA dehydrogenase |
| Rv0244c | *fadE5* | acyl-CoA dehydrogenase |
| Rv0271c | *fadE6* | acyl-CoA dehydrogenase |
| Rv0400c | *fadE7* | acyl-CoA dehydrogenase |
| Rv0672 | *fadE8* | acyl-CoA dehydrogenase (aka *aidB*) |
| Rv0752c | *fadE9* | acyl-CoA dehydrogenase |
| Rv0873 | *fadE10* | acyl-CoA dehydrogenase |
| Rv0972c | *fadE12* | acyl-CoA dehydrogenase |
| Rv0975c | *fadE13* | acyl-CoA dehydrogenase |
| Rv1346 | *fadE14* | acyl-CoA dehydrogenase |
| Rv1467c | *fadE15* | acyl-CoA dehydrogenase |
| Rv1679 | *fadE16* | acyl-CoA dehydrogenase |
| Rv1934c | *fadE17* | acyl-CoA dehydrogenase |
| Rv1933c | *fadE18* | acyl-CoA dehydrogenase |
| Rv2500c | *fadE19* | acyl-CoA dehydrogenase (aka *mmgC*) |
| Rv2724c | *fadE20* | acyl-CoA dehydrogenase |
| Rv2789c | *fadE21* | acyl-CoA dehydrogenase |
| Rv3061c | *fadE22* | acyl-CoA dehydrogenase |
| Rv3140 | *fadE23* | acyl-CoA dehydrogenase |
| Rv3139 | *fadE24* | acyl-CoA dehydrogenase |
| Rv3274c | *fadE25* | acyl-CoA dehydrogenase |
| Rv3504 | *fadE26* | acyl-CoA dehydrogenase |
| Rv3505 | *fadE27* | acyl-CoA dehydrogenase |
| Rv3544c | *fadE28* | acyl-CoA dehydrogenase |
| Rv3543c | *fadE29* | acyl-CoA dehydrogenase |
| Rv3560c | *fadE30* | acyl-CoA dehydrogenase |
| Rv3562 | *fadE31* | acyl-CoA dehydrogenase |
| Rv3563 | *fadE32* | acyl-CoA dehydrogenase |
| Rv3564 | *fadE33* | acyl-CoA dehydrogenase |
| Rv3573c | *fadE34* | acyl-CoA dehydrogenase |
| Rv3797 | *fadE35* | acyl-CoA dehydrogenase |
| Rv3761c | *fadE36* | acyl-CoA dehydrogenase |
| Rv1175c | *fadH* | 2,4-Dienoyl-CoA Reductase |
| Rv0855 | *far* | fatty acyl-CoA racemase |
| Rv1143 | *mcr* | α-methyl acyl-CoA racemase |
| Rv1492 | *mutA* | methylmalonyl-CoA mutase, β subunit |
| Rv1493 | *mutB* | methylmalonyl-CoA mutase, α subunit |
| Rv2504c | *scoA* | 3-oxo acid:CoA transferase, α subunit |
| Rv2503c | *scoB* | 3-oxo acid:CoA transferase, β subunit |
| Rv1136 | - | probable carnitine racemase |
| Rv1683 | - | possible acyl-CoA synthase |
| 4. Phosphorous compounds | | |
| Rv2368c | *phoH* | ATP-binding *pho* regulon component |
| Rv1095 | *phoH2* | PhoH-like protein |
| Rv3628 | *ppa* | probable inorganic pyrophosphatase |
| Rv2984 | *ppk* | polyphosphate kinase |
| *B. Energy metabolism* | | |
| 1. Glycolysis | | |
| Rv1023 | *eno* | enolase |
| Rv0363c | *fba* | fructose bisphosphate aldolase |
| Rv1436 | *gap* | glyceraldehyde 3-phosphate dehydrogenase |
| Rv0489 | *gpm* | phosphoglycerate mutase I |
| Rv3010c | *pfkA* | phosphofructokinase I |
| Rv2029c | *pfkB* | phosphofructokinase II |
| Rv0946c | *pgi* | glucose-6-phosphate isomerase |
| Rv1437 | *pgk* | phosphoglycerate kinase |
| Rv1617 | *pykA* | pyruvate kinase |
| Rv1438 | *tpi* | triosephosphate isomerase |
| Rv2419c | - | putative phosphoglycerate mutase |
| Rv3837c | - | putative phosphoglycerate mutase |
| 2. Pyruvate dehydrogenase | | |
| Rv2241 | *aceE* | pyruvate dehydrogenase E1 component |
| Rv3303c | *lpdA* | dihydrolipoamide dehydrogenase |
| Rv2497c | *pdhA* | pyruvate dehydrogenase E1 component α subunit |
| Rv2496c | *pdhB* | pyruvate dehydrogenase E1 component β subunit |
| Rv2495c | *pdhC* | dihydrolipoamide acetyltransferase |
| Rv0462 | - | probable dihydrolipoamide dehydrogenase |
| 3. TCA cycle | | |
| Rv1475c | *acn* | aconitate hydratase |
| Rv0889c | *citA* | citrate synthase 2 |
| Rv2498c | *citE* | citrate lyase β chain |
| Rv1098c | *fum* | fumarase |
| Rv1131 | *gltA1* | citrate synthase 3 |
| Rv0896 | *gltA2* | citrate synthase 1 |
| Rv3339c | *icd1* | isocitrate dehydrogenase |
| Rv0066c | *icd2* | isocitrate dehydrogenase |
| Rv0794c | *lpdB* | dihydrolipoamide dehydrogenase |
| Rv1240 | *mdh* | malate dehydrogenase |
| Rv2967c | *pca* | pyruvate carboxylase |
| Rv3318 | *sdhA* | succinate dehydrogenase A |
| Rv3319 | *sdhB* | succinate dehydrogenase B |
| Rv3316 | *sdhC* | succinate dehydrogenase C subunit |
| Rv3317 | *sdhD* | succinate dehydrogenase D subunit |
| Rv1248c | *sucA* | 2-oxoglutarate dehydrogenase |
| Rv2215 | *sucB* | dihydrolipoamide succinyltransferase |
| Rv0951 | *sucC* | succinyl-CoA synthase β chain |
| Rv0952 | *sucD* | succinyl-CoA synthase α chain |
| 4. Glyoxylate bypass | | |
| Rv0467 | *aceA* | isocitrate lyase |
| Rv1915 | *aceAa* | isocitrate lyase, α module |
| Rv1916 | *aceAb* | isocitrate lyase, β module |
| Rv1837c | *glcB* | malate synthase |
| Rv3323c | *gphA* | phosphoglycolate phosphatase |
| 5. Pentose phosphate pathway | | |
| Rv1445c | *devB* | glucose-6-phosphate 1-dehydrogenase |
| Rv1844c | *gnd* | 6-phosphogluconate dehydrogenase (Gram –) |
| Rv1122 | *gnd2* | 6-phosphogluconate dehydrogenase (Gram +) |
| Rv1446c | *opcA* | unknown function, may aid G6PDH |

| | | |
|---|---|---|
| Rv2436 | *rbsK* | ribokinase |
| Rv1408 | *rpe* | ribulose-phosphate 3-epimerase |
| Rv2465c | *rpi* | phosphopentose isomerase |
| Rv1448c | *tal* | transaldolase |
| Rv1449c | *tkt* | transketolase |
| Rv1121 | *zwf* | glucose-6-phosphate 1-dehydrogenase |
| Rv1447c | *zwf2* | glucose-6-phosphate 1-dehydrogenase |

6. Respiration

*a. aerobic*

| | | |
|---|---|---|
| Rv0527 | *ccsA* | cytochrome *c*-type biogenesis protein |
| Rv0529 | *ccsB* | cytochrome *c*-type biogenesis protein |
| Rv1451 | *ctaB* | cytochrome *c* oxidase assembly factor |
| Rv2200c | *ctaC* | cytochrome *c* oxidase chain II |
| Rv3043c | *ctaD* | cytochrome *c* oxidase polypeptide I |
| Rv2193 | *ctaE* | cytochrome *c* oxidase polypeptide III |
| Rv1542c | *glbN* | hemoglobin-like, oxygen carrier |
| Rv2470 | *glbO* | hemoglobin-like, oxygen carrier |
| Rv2249c | *glpD1* | glycerol-3-phosphate dehydrogenase |
| Rv3302c | *glpD2* | glycerol-3-phosphate dehydrogenase |
| Rv0694 | *lldD1* | L-lactate dehydrogenase (cytochrome) |
| Rv1872c | *lldD2* | L-lactate dehydrogenase |
| Rv1854c | *ndh* | probable NADH dehydrogenase |
| Rv3145 | *nuoA* | NADH dehydrogenase chain A |
| Rv3146 | *nuoB* | NADH dehydrogenase chain B |
| Rv3147 | *nuoC* | NADH dehydrogenase chain C |
| Rv3148 | *nuoD* | NADH dehydrogenase chain D |
| Rv3149 | *nuoE* | NADH dehydrogenase chain E |
| Rv3150 | *nuoF* | NADH dehydrogenase chain F |
| Rv3151 | *nuoG* | NADH dehydrogenase chain G |
| Rv3152 | *nuoH* | NADH dehydrogenase chain H |
| Rv3153 | *nuoI* | NADH dehydrogenase chain I |
| Rv3154 | *nuoJ* | NADH dehydrogenase chain J |
| Rv3155 | *nuoK* | NADH dehydrogenase chain K |
| Rv3156 | *nuoL* | NADH dehydrogenase chain L |
| Rv3157 | *nuoM* | NADH dehydrogenase chain M |
| Rv3158 | *nuoN* | NADH dehydrogenase chain N |
| Rv2195 | *qcrA* | Rieske iron-sulphur component of *ubiQ-cytB* reductase |
| Rv2196 | *qcrB* | cytochrome β component of *ubiQ-cytB* reductase |
| Rv2194 | *qcrC* | cytochrome *b/c* component of *ubiQ-cytB* reductase |

*b. anaerobic*

| | | |
|---|---|---|
| Rv2392 | *cysH* | 3'-phosphoadenylylsulfate (PAPS) reductase |
| Rv2899c | *fdhD* | affects formate dehydrogenase-N |
| Rv2900c | *fdhF* | molybdopterin-containing oxidoreductase |
| Rv1552 | *frdA* | fumarate reductase flavoprotein subunit |
| Rv1553 | *frdB* | fumarate reductase iron sulphur protein |
| Rv1554 | *frdC* | fumarate reductase 15kD anchor protein |
| Rv1555 | *frdD* | fumarate reductase 13kD anchor protein |
| Rv1161 | *narG* | nitrate reductase α subunit |
| Rv1162 | *narH* | nitrate reductase β chain |
| Rv1164 | *narI* | nitrate reductase γ chain |
| Rv1163 | *narJ* | nitrate reductase δ chain |
| Rv1736c | *narX* | fused nitrate reductase |
| Rv2391 | *nirA* | probable nitrite reductase/sulphite reductase |
| Rv0252 | *nirB* | nitrite reductase flavoprotein |
| Rv0253 | *nirD* | probable nitrite reductase small subunit |

*c. Electron transport*

| | | |
|---|---|---|
| Rv0409 | *ackA* | acetate kinase |
| Rv1623c | *appC* | cytochrome *bd-II* oxidase subunit I |
| Rv1622c | *cydB* | cytochrome *d* ubiquinol oxidase subunit II |
| Rv1620c | *cydC* | ABC transporter |
| Rv1621c | *cydD* | ABC transporter |
| Rv2007c | *fdxA* | ferredoxin |
| Rv3554 | *fdxB* | ferredoxin |
| Rv1177 | *fdxC* | ferredoxin 4Fe-4S |
| Rv3503c | *fdxD* | probable ferredoxin |
| Rv3029c | *fixA* | electron transfer flavoprotein β subunit |
| Rv3028c | *fixB* | electron transfer flavoprotein α subunit |
| Rv3106 | *fprA* | adrenodoxin and NADPH ferredoxin reductase |
| Rv0886 | *fprB* | ferredoxin, ferredoxin-NADP reductase |
| Rv3251c | *rubA* | rubredoxin A |
| Rv3250c | *rubB* | rubredoxin B |

7. Miscellaneous oxidoreductases and oxygenases 171

8. ATP-proton motive force

| | | |
|---|---|---|
| Rv1308 | *atpA* | ATP synthase α chain |
| Rv1304 | *atpB* | ATP synthase a chain |
| Rv1311 | *atpC* | ATP synthase ϵ chain |
| Rv1310 | *atpD* | ATP synthase β chain |
| Rv1305 | *atpE* | ATP synthase c chain |
| Rv1306 | *atpF* | ATP synthase b chain |
| Rv1309 | *atpG* | ATP synthase γ chain |
| Rv1307 | *atpH* | ATP synthase δ chain |

*C. Central intermediary metabolism*

1. General

| | | |
|---|---|---|
| Rv2589 | *gabT* | 4-aminobutyrate aminotransferase |
| Rv3432c | *gadB* | glutamate decarboxylase |
| Rv1832 | *gcvB* | glycine decarboxylase |
| Rv1826 | *gcvH* | glycine cleavage system H protein |
| Rv2211c | *gcvT* | T protein of glycine cleavage system |
| Rv1213 | *glgC* | glucose-1-phosphate adenylyltransferase |
| Rv3842c | *glpQ1* | glycerophosphoryl diester phosphodiesterase |
| Rv0317c | *glpQ2* | glycerophosphoryl diester phosphodiesterase |
| Rv3566c | *nhoA* | N-hydroxyarylamine *o*-acetyltransferase |
| Rv0155 | *pntAA* | pyridine transhydrogenase subunit α1 |
| Rv0156 | *pntAB* | pyridine transhydrogenase subunit α2 |
| Rv0157 | *pntB* | pyridine transhydrogenase subunit β |
| Rv1127c | *ppdK* | similar to pyruvate, phosphate dikinase |

2. Gluconeogenesis

| | | |
|---|---|---|
| Rv0211 | *pckA* | phosphoenolpyruvate carboxykinase |
| Rv0069c | *sdaA* | L-serine dehydratase 1 |

3. Sugar nucleotides

| | | |
|---|---|---|
| Rv1512 | *epiA* | nucleotide sugar epimerase |
| Rv3784 | *epiB* | probable UDP-galactose 4-epimerase |
| Rv1511 | *gmdA* | GDP-mannose 4,6 dehydratase |
| Rv0334 | *rmlA* | glucose-1-phosphate thymidyltransferase |
| Rv3264c | *rmlA2* | glucose-1-phosphate thymidyltransferase |
| Rv3464 | *rmlB* | dTDP-glucose 4,6-dehydratase |
| Rv3634c | *rmlB2* | dTDP-glucose 4,6-dehydratase |
| Rv3468c | *rmlB3* | dTDP-glucose 4,6-dehydratase |
| Rv3465 | *rmlC* | dTDP-4-dehydrorhamnose 3,5-epimerase |
| Rv3266c | *rmlD* | dTDP-4-dehydrorhamnose reductase |
| Rv0322 | *udgA* | UDP-glucose dehydrogenase/GDP-mannose 6-dehydrogenase |
| Rv3265c | *wbbL* | dTDP-rhamnosyl transferase |
| Rv1525 | *wbbL2* | dTDP-rhamnosyl transferase |
| Rv3400 | - | probable β-phosphoglucomutase |

4. Amino sugars

| | | |
|---|---|---|
| Rv3436c | *glmS* | glucosamine-fructose-6-phosphate aminotransferase |

5. Sulphur metabolism

| | | |
|---|---|---|
| Rv0711 | *atsA* | arylsulfatase |
| Rv3299c | *atsB* | proable arylsulfatase |
| Rv0663 | *atsD* | proable arylsulfatase |
| Rv3077 | *atsF* | proable arylsulfatase |
| Rv0296c | *atsG* | proable arylsulfatase |
| Rv3796 | *atsH* | proable arylsulfatase |
| Rv1285 | *cysD* | ATP:sulphurylase subunit 2 |
| Rv1286 | *cysN* | ATP:sulphurylase subunit 1 |
| Rv2131c | *cysQ* | homologue of *M.leprae cysQ* |
| Rv3248c | *sahH* | adenosylhomocysteinase |
| Rv3283 | *sseA* | thiosulfate sulfurtransferase |
| Rv2291 | *sseB* | thiosulfate sulfurtransferase |
| Rv3118 | *sseC* | thiosulfate sulfurtransferase |
| Rv0814c | *sseC2* | thiosulfate sulfurtransferase |
| Rv3762c | - | probable alkyl sulfatase |

*D. Amino acid biosynthesis*

1. Glutamate family

| | | |
|---|---|---|
| Rv1654 | *argB* | acetylglutamate kinase |
| Rv1652 | *argC* | N-acetyl-γ-glutamyl-phosphate reductase |
| Rv1655 | *argD* | acetylornithine aminotransferase |
| Rv1656 | *argF* | ornithine carbamoyltransferase |
| Rv1658 | *argG* | arginosuccinate synthase |
| Rv1659 | *argH* | arginosuccinate lyase |
| Rv1653 | *argJ* | glutamate N-acetyltransferase |
| Rv2220 | *glnA1* | glutamine synthase class I |
| Rv2222c | *glnA2* | glutamine synthase class II |
| Rv1878 | *glnA3* | probable glutamine synthase |
| Rv2860c | *glnA4* | proable glutamine synthase |
| Rv2918c | *glnD* | uridylyltransferase |
| Rv2221c | *glnE* | glutamate-ammonia-ligase adenyltransferase |
| Rv3859c | *gltB* | ferredoxin-dependent glutamate synthase |
| Rv3858c | *gltD* | small subunit of NADH-dependent glutamate synthase |
| Rv3704c | *gshA* | possible γ-glutamylcysteine synthase |
| Rv2427c | *proA* | γ-glutamyl phosphate reductase |
| Rv2439c | *proB* | glutamate 5-kinase |
| Rv0500 | *proC* | pyrroline-5-carboxylate reductase |

2. Aspartate family

| | | |
|---|---|---|
| Rv3708c | *asd* | aspartate semialdehyde dehydrogenase |
| Rv3709c | *ask* | aspartokinase |
| Rv2201 | *asnB* | asparagine synthase B |
| Rv3565 | *aspB* | aspartate aminotransferase |
| Rv0337c | *aspC* | aspartate aminotransferase |
| Rv2753c | *dapA* | dihydrodipicolinate synthase |
| Rv2773c | *dapB* | dihydrodipicolinate reductase |
| Rv1202 | *dapE* | succinyl-diaminopimelate desuccinylase |
| Rv2141c | *dapE2* | ArgE/DapE/Acy1/Cpg2/yscS family |
| Rv2726c | *dapF* | diaminopimelate epimerase |
| Rv1293 | *lysA* | diaminopimelate decarboxylase |
| Rv3341 | *metA* | homoserine *o*-acetyltransferase |
| Rv1079 | *metB* | cystathionine γ-synthase |
| Rv3340 | *metC* | cystathionine β-lyase |
| Rv1133c | *metE* | 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase |
| Rv2124c | *metH* | 5-methyltetrahydrofolate-homocysteine methyltransferase |
| Rv1392 | *metK* | *S*-adenosylmethionine synthase |
| Rv0391 | *metZ* | *o*-succinylhomoserine sulfhydrylase |
| Rv1294 | *thrA* | homoserine dehydrogenase |
| Rv1296 | *thrB* | homoserine kinase |
| Rv1295 | *thrC* | homoserine synthase |

3. Serine family

| | | |
|---|---|---|
| Rv0815c | *cysA2* | thiosulfate sulfurtransferase |
| Rv3117 | *cysA3* | thiosulfate sulfurtransferase |
| Rv2335 | *cysE* | serine acetyltransferase |
| Rv0511 | *cysG* | uroporphyrin-III *c*-methyltransferase |
| Rv2847c | *cysG2* | multifunctional enzyme, siroheme synthase |
| Rv2334 | *cysK* | cysteine synthase A |
| Rv1336 | *cysM* | cysteine synthase B |
| Rv1077 | *cysM2* | cystathionine β-synthase |
| Rv0848 | *cysM3* | putative cysteine synthase |
| Rv1093 | *glyA* | serine hydroxymethyltransferase |
| Rv0070c | *glyA2* | serine hydroxymethyltransferase |
| Rv2996c | *serA* | D-3-phosphoglycerate dehydrogenase |
| Rv0505c | *serB* | probable phosphoserine phosphatase |
| Rv3042c | *serB2* | C-term similar to phosphoserine phosphatase |
| Rv0884c | *serC* | phosphoserine aminotransferase |

4. Aromatic amino acid family

| | | |
|---|---|---|
| Rv3227 | *aroA* | 3-phosphoshikimate 1-carboxyvinyl transferase |
| Rv2538c | *aroB* | 3-dehydroquinate synthase |
| Rv2537c | *aroD* | 3-dehydroquinate dehydratase |
| Rv2552c | *aroE* | shikimate 5-dehydrogenase |
| Rv2540c | *aroF* | chorismate synthase |
| Rv2178c | *aroG* | DAHP synthase |
| Rv2539c | *aroK* | shikimate kinase I |
| Rv3838c | *pheA* | prephenate dehydratase |
| Rv1613 | *trpA* | tryptophan synthase α chain |
| Rv1612 | *trpB* | tryptophan synthase β chain |
| Rv1611 | *trpC* | indole-3-glycerol phosphate synthase |
| Rv2192c | *trpD* | anthranilate phosphoribosyltransferase |
| Rv1609 | *trpE* | anthranilate synthase component I |
| Rv2386c | *trpE2* | anthranilate synthase component I |
| Rv3754 | *tyrA* | prephenate dehydrogenase |

5. Histidine

| | | |
|---|---|---|
| Rv1603 | *hisA* | phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide isomerase |
| Rv1601 | *hisB* | imidazole glycerol-phosphate dehydratase |
| Rv1600 | *hisC* | histidinol-phosphate aminotransferase |
| Rv3772 | *hisC2* | histidinol-phosphate aminotransferase |
| Rv1599 | *hisD* | histidinol dehydrogenase |

| | | |
|---|---|---|
| Rv1605 | *hisF* | imidazole glycerol-phosphate synthase |
| Rv2121c | *hisG* | ATP phosphoribosyltransferase |
| Rv1602 | *hisH* | amidotransferase |
| Rv2122c | *hisI* | phosphoribosyl-AMP cyclohydrolase |
| Rv1606 | *hisI2* | probable phosphoribosyl-AMP 1,6 cyclohydrolase |
| Rv0114 | - | similar to HisB |

6. Pyruvate family

| | | |
|---|---|---|
| Rv3423c | *alr* | alanine racemase |

7. Branched amino acid family

| | | |
|---|---|---|
| Rv1559 | *ilvA* | threonine deaminase |
| Rv3003c | *ilvB* | acetolactate synthase I large subunit |
| Rv3470c | *ilvB2* | acetolactate synthase large subunit |
| Rv3001c | *ilvC* | ketol-acid reductoisomerase |
| Rv0189c | *ilvD* | dihydroxy-acid dehydratase |
| Rv2210c | *ilvE* | branched-chain-amino-acid transaminase |
| Rv1820 | *ilvG* | acetolactate synthase II |
| Rv3002c | *ilvN* | acetolactate synthase I small subunit |
| Rv3509c | *ilvX* | probable acetohydroxyacid synthase I large subunit |
| Rv3710 | *leuA* | α-isopropyl malate synthase |
| Rv2995c | *leuB* | 3-isopropylmalate dehydrogenase |
| Rv2988c | *leuC* | 3-isopropylmalate dehydratase large subunit |
| Rv2987c | *leuD* | 3-isopropylmalate dehydratase small subunit |

*E. Polyamine synthesis*

| | | |
|---|---|---|
| Rv2601 | *speE* | spermidine synthase |

*F. Purines, pyrimidines, nucleosides and nucleotides*

1. Purine ribonucleotide biosynthesis

| | | |
|---|---|---|
| Rv1389 | *gmk* | putative guanylate kinase |
| Rv3396c | *guaA* | GMP synthase |
| Rv1843c | *guaB1* | inosine-5'-monophosphate dehydrogenase |
| Rv3411c | *guaB2* | inosine-5'-monophosphate dehydrogenase |
| Rv3410c | *guaB3* | inosine-5'-monophosphate dehydrogenase |
| Rv1017c | *prsA* | ribose-phosphate pyrophosphokinase |
| Rv0357c | *purA* | adenylosuccinate synthase |
| Rv0777 | *purB* | adenylosuccinate lyase |
| Rv0780 | *purC* | phosphoribosylaminoimidazole-succinocarboxamide synthase |
| Rv0772 | *purD* | phosphoribosylamine-glycine ligase |
| Rv3275c | *purE* | phosphoribosylaminoimidazole carboxylase |
| Rv0808 | *purF* | amidophosphoribosyltransferase- |
| Rv0957 | *purH* | phosphoribosylaminoimidazole-carboxamide formyltransferase |
| Rv3276c | *purK* | phosphoribosylaminoimidazole carboxylase ATPase subunit |
| Rv0803 | *purL* | phosphoribosylformylglycin-amidine synthase II |
| Rv0809 | *purM* | 5'-phosphoribosyl-5-aminoimidazole synthase |
| Rv0956 | *purN* | phosphoribosylglycinamide formyltransferase I |
| Rv0788 | *purQ* | phosphoribosylformylglycin-amidine synthase I |
| Rv0389 | *purT* | phosphoribosylglycinamide formyltransferase II |
| Rv2964 | *purU* | formyltetrahydrofolate deformylase |

2. Pyrimidine ribonucleotide biosynthesis

| | | |
|---|---|---|
| Rv1383 | *carA* | carbamoyl-phosphate synthase subunit |
| Rv1384 | *carB* | carbamoyl-phosphate synthase subunit |
| Rv1380 | *pyrB* | aspartate carbamoyltransferase |
| Rv1381 | *pyrC* | dihydroorotase |
| Rv2139 | *pyrD* | dihydroorotate dehydrogenase |
| Rv1385 | *pyrF* | orotidine 5'-phosphate decarboxylase |
| Rv1699 | *pyrG* | CTP synthase |
| Rv2883c | *pyrH* | uridylate kinase |
| Rv0382c | *umpA* | probable uridine 5'-monophosphate synthase |

3. 2'-deoxyribonucleotide metabolism

| | | |
|---|---|---|
| Rv0321 | *dcd* | deoxycytidine triphosphate deaminase |
| Rv2697c | *dut* | deoxyuridine triphosphatase |
| Rv0233 | *nrdB* | ribonucleoside-diphosphate reductase B2 (eukaryotic-like) |
| Rv3051c | *nrdE* | ribonucleoside diphosphate reductase α chain |
| Rv1981c | *nrdF* | ribonucleotide reductase small subunit |
| Rv3048c | *nrdG* | ribonucleoside-diphosphate small subunit |
| Rv3053c | *nrdH* | glutaredoxin electron transport component of NrdEF system |
| Rv3052c | *nrdI* | NrdI/YgaO/YmaA family |
| Rv3247c | *tmk* | thymidylate kinase |
| Rv2764c | *thyA* | thymidylate synthase |
| Rv0570 | *nrdZ* | ribonucleotide reductase, class II |
| Rv3752c | - | probable cytidine/deoxycytidylate deaminase |

4. Salvage of nucleosides and nucleotides

| | | |
|---|---|---|
| Rv3313c | *add* | probable adenosine deaminase |
| Rv2584c | *apt* | adenine phosphoribosyltransferases |
| Rv3315c | *cdd* | probable cytidine deaminase |
| Rv3314c | *deoA* | thymidine phosphorylase |
| Rv0478 | *deoC* | deoxyribose-phosphate aldolase |
| Rv3307 | *deoD* | probable purine nucleoside phosphorylase |
| Rv3624c | *hpt* | probable hypoxanthine-guanine phosphoribosyltransferase |
| Rv3393 | *iunH* | probable inosine-uridine preferring nucleoside hydrolase |
| Rv0535 | *pnp* | phosphorylase from Pnp/MtaP family 2 |
| Rv3309c | *upp* | uracil phophoribosyltransferase |

5. Miscellaneous nucleoside/nucleotide reactions

| | | |
|---|---|---|
| Rv0733 | *adk* | probable adenylate kinase |
| Rv2364c | *bex* | GTP-binding protein of Era/ThdF family |
| Rv1712 | *cmk* | cytidylate kinase |
| Rv2344c | *dgt* | probable deoxyguanosine triphosphate hydrolase |
| Rv2404c | *lepA* | GTP-binding protein LepA |
| Rv2727c | *miaA* | tRNA δ(2)-isopentenylpyrophosphate transferase |
| Rv2445c | *ndkA* | nucleoside diphosphate kinase |
| Rv2440c | *obg* | Obg GTP-binding protein |
| Rv2583c | *relA* | (p)ppGpp synthase I |

*G. Biosynthesis of cofactors, prosthetic groups and carriers*

1. Biotin

| | | |
|---|---|---|
| Rv1568 | *bioA* | adenosylmethionine-8-amino-7-oxononanoate aminotransferase |
| Rv1589 | *bioB* | biotin synthase |
| Rv1570 | *bioD* | dethiobiotin synthase |
| Rv1569 | *bioF* | 8-amino-7-oxononanoate synthase |
| Rv0032 | *bioF2* | C-terminal similar to *B. subtilis* BioF |
| Rv3279c | *birA* | biotin apo-protein ligase |
| Rv1442 | *bisC* | biotin sulfoxide reductase |
| Rv0089 | - | possible *bioC* biotin synthesis gene |

2. Folic acid

| | | |
|---|---|---|
| Rv2763c | *dfrA* | dihydrofolate reductase |
| Rv2447c | *folC* | folylpolyglutamate synthase |
| Rv3356c | *folD* | methylenetetrahydrofolate dehydrogenase |
| Rv3609c | *folE* | GTP cyclohydrolase I |
| Rv3606c | *folK* | 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase |
| Rv3608c | *folP* | dihydropteroate synthase |
| Rv1207 | *folP2* | dihydropteroate synthase |
| Rv3607c | *folX* | may be involved in folate biosynthesis |
| Rv0013 | *pabA* | *p*-aminobenzoate synthase glutamine amidotransferase |
| Rv1005c | *pabB* | *p*-aminobenzoate synthase |
| Rv0812 | *pabC* | aminodeoxychorismate lyase |

3. Lipoate

| | | |
|---|---|---|
| Rv2218 | *lipA* | lipoate biosynthesis protein A |
| Rv2217 | *lipB* | lipoate biosynthesis protein B |

4. Molybdopterin

| | | |
|---|---|---|
| Rv3109 | *moaA* | molybdenum cofactor biosynthesis, protein A |
| Rv0869c | *moaA2* | molybdenum cofactor biosynthesis, protein A |
| Rv0438c | *moaA3* | molybdenum cofactor biosynthesis, protein A |
| Rv3110 | *moaB* | molybdenum cofactor biosynthesis, protein B |
| Rv0984 | *moaB2* | molybdenum cofactor biosynthesis, protein B |
| Rv3111 | *moaC* | molybdenum cofactor biosynthesis, protein C |
| Rv0864 | *moaC2* | molybdenum cofactor biosynthesis, protein C |
| Rv3324c | *moaC3* | molybdenum cofactor biosynthesis, protein C |
| Rv3112 | *moaD* | molybdopterin converting factor subunit 1 |
| Rv0868c | *moaD2* | molybdopterin converting factor subunit 1 |
| Rv3119 | *moaE* | molybdopterin-converting factor subunit 2 |
| Rv0866 | *moaE2* | molybdopterin-converting factor subunit 2 |
| Rv3322c | *moaE3* | molybdopterin-converting factor subunit 2 |
| Rv0994 | *moeA* | molybdopterin biosynthesis |
| Rv3116 | *moeB* | molybdopterin biosynthesis |
| Rv2338c | *moeW* | molybdopterin biosynthesis |
| Rv1681 | *moeX* | weak similarity to *E. coli* MoaA |
| Rv1355c | *moeY* | weak similarity to *E. coli* MoeB |
| Rv3206c | *moeZ* | probably involved in molybdopterin biosynthesis |
| Rv0865 | *mog* | molybdopterin biosynthesis |

5. Pantothenate

| | | |
|---|---|---|
| Rv1092c | *coaA* | pantothenate kinase |
| Rv2225 | *panB* | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| Rv3602c | *panC* | pantoate-β-alanine ligase |
| Rv3601c | *panD* | aspartate 1-decarboxylase |

6. Pyridoxine

| | | |
|---|---|---|
| Rv2607 | *pdxH* | pyridoxamine 5'-phosphate oxidase |

7. Pyridine nucleotide

| | | |
|---|---|---|
| Rv1594 | *nadA* | quinolinate synthase |
| Rv1595 | *nadB* | L-aspartate oxidase |
| Rv1596 | *nadC* | nicotinate-nucleotide pyrophosphatase |
| Rv0423c | *thiC* | thiamine synthesis, pyrimidine moiety |

8. Thiamine

| | | |
|---|---|---|
| Rv0422c | *thiD* | phosphomethylpyrimidine kinase |
| Rv0414c | *thiE* | thiamine synthesis, thiazole moiety |
| Rv0417 | *thiG* | thiamine synthesis, thiazole moiety |
| Rv2977c | *thiL* | probable thiamine-monophosphate kinase |

9. Riboflavin

| | | |
|---|---|---|
| Rv1940 | *ribA* | GTP cyclohydrolase II |
| Rv1415 | *ribA2* | probable GTP cyclohydrolase II |
| Rv1412 | *ribC* | riboflavin synthase α chain |
| Rv2671 | *ribD* | probable riboflavin deaminase |
| Rv2786c | *ribF* | riboflavin kinase |
| Rv1409 | *ribG* | riboflavin biosynthesis |
| Rv1416 | *ribH* | riboflavin synthase β chain |
| Rv3300c | - | probable deaminase, riboflavin synthesis |

10. Thioredoxin, glutaredoxin and mycothiol

| | | |
|---|---|---|
| Rv0773c | *ggtA* | putative γ-glutamyl transpeptidase |
| Rv2394 | *ggtB* | γ -glutamyltranspeptidase precursor |
| Rv2855 | *gorA* | glutathione reductase homologue |
| Rv0816c | *thiX* | equivalent to *M. leprae* ThiX |
| Rv1470 | *trxA* | thioredoxin |
| Rv1471 | *trxB* | thioredoxin reductase |
| Rv3913 | *trxB2* | thioredoxin reductase |
| Rv3914 | *trxC* | thioredoxin |

11. Menaquinone, PQQ, ubiquinone and other terpenoids

| | | |
|---|---|---|
| Rv2682c | *dxs* | 1-deoxy-D-xylulose 5-phosphate synthase |
| Rv0562 | *grcC1* | heptaprenyl diphosphate synthase II |
| Rv0989c | *grcC2* | heptaprenyl diphosphate synthase II |
| Rv3398c | *idsA* | geranylgeranyl pyrophosphate synthase |
| Rv2173 | *idsA2* | geranylgeranyl pyrophosphate synthase |
| Rv3383c | *idsB* | transfergeranyl, similar geranyl pyrophosphate synthase |
| Rv0534c | *menA* | 4-dihydroxy-2-naphthoate octaprenyltransferase |
| Rv0548c | *menB* | naphthoate synthase |
| Rv0553 | *menC* | *o*-succinylbenzoate-CoA synthase |
| Rv0555 | *menD* | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| Rv0542c | *menE* | *o*-succinylbenzoic acid-CoA ligase |
| Rv3853 | *menG* | *S*-adenosylmethionine: 2-demethylmenaquinone |
| Rv3397c | *phyA* | phytoene synthase |
| Rv0693 | *pqqE* | coenzyme PQQ synthesis protein E |
| Rv0558 | *ubiE* | ubiquinone/menaquinone biosynthesis methyltransferase |

12. Heme and porphyrin

| | | |
|---|---|---|
| Rv0509 | *hemA* | glutamyl-tRNA reductase |
| Rv0512 | *hemB* | δ-aminolevulinic acid dehydratase |
| Rv0510 | *hemC* | porphobilinogen deaminase |
| Rv2678c | *hemE* | uroporphyrinogen decarboxylase |

| | | |
|---|---|---|
| Rv1300 | *hemK* | protoporphyrinogen oxidase |
| Rv0524 | *hemL* | glutamate-1-semialdehyde aminotransferase |
| Rv2388c | *hemN* | oxygen-independent coproporphyrinogen III oxidase |
| Rv2677c | *hemY'* | protoporphyrinogen oxidase |
| Rv1485 | *hemZ* | ferrochelatase |
| 13. Cobalamin | | |
| Rv2849c | *cobA* | cob(I)alamin adenosyltransferase |
| Rv2848c | *cobB* | cobyrinic acid a,c-diamide synthase |
| Rv2231c | *cobC* | aminotransferase |
| Rv2236c | *cobD* | cobinamide synthase |
| Rv2064 | *cobG* | percorrin reductase |
| Rv2065 | *cobH* | precorrin isomerase |
| Rv2066 | *cobI* | CobI-CobJ fusion protein |
| Rv2070c | *cobK* | precorrin reductase |
| Rv2072c | *cobL* | probable methyltransferase |
| Rv2071c | *cobM* | precorrin-3 methylase |
| Rv2062c | *cobN* | cobalt insertion |
| Rv2208 | *cobS* | cobalamin (5'-phosphate) synthase |
| Rv2207 | *cobT* | nicotinate-nucleotide-dimethylbenzimidazole transferase |
| Rv0254c | *cobU* | cobinamide kinase |
| Rv0255c | *cobQ* | cobyric acid synthase |
| Rv3713 | *cobQ2* | possible cobyric acid synthase |
| Rv0306 | - | similar to BluB cobalamin synthesis protein *R. capsulatus* |
| 14. Iron utilization | | |
| Rv1876 | *bfrA* | bacterioferritin |
| Rv3841 | *bfrB* | bacterioferritin |
| Rv3215 | *entC* | probable isochorismate synthase |
| Rv3214 | *entD* | weak similarity to many phosphoglycerate mutases |
| Rv2895c | *viuB* | similar to proteins involved in vibriobactin uptake |
| Rv3525c | - | similar to ferripyochelin binding protein |
| *H. Lipid biosynthesis* | | |
| 1. Synthesis of fatty and mycolic acids | | |
| Rv3285 | *accA3* | acetyl/propionyl CoA carboxylase α subunit |
| Rv0904c | *accD3* | acetyl/propionyl CoA carboxylase β subunit |
| Rv3799c | *accD4* | acetyl/propionyl CoA carboxylase β subunit |
| Rv3280 | *accD5* | acetyl/propionyl CoA carboxylase β subunit |
| Rv2247 | *accD6* | acetyl/propionyl CoA carboxylase β subunit |
| Rv2244 | *acpM* | acyl carrier protein (meromycolate extension) |
| Rv2523c | *acpS* | CoA:apo-[ACP] pantethienephosphotransferase |
| Rv2243 | *fabD* | malonyl CoA-[ACP] transacylase |
| Rv0649 | *fabD2* | malonyl CoA-[ACP] transacylase |
| Rv1483 | *fabG1* | 3-oxoacyl-[ACP] reductase (aka MabA) |
| Rv1350 | *fabG2* | 3-oxoacyl-[ACP] Reductase |
| Rv2002 | *fabG3* | 3-oxoacyl-[ACP] reductase |
| Rv0242c | *fabG4* | 3-oxoacyl-[ACP] reductase |
| Rv2766c | *fabG5* | 3-oxoacyl-[ACP] reductase |
| Rv0533c | *fabH* | β-ketoacyl-ACP synthase III |
| Rv2524c | *fas* | fatty acid synthase |
| Rv1484 | *inhA* | enoyl-[ACP] reductase |
| Rv2245 | *kasA* | β-ketoacyl-ACP synthase (meromycolate extension) |
| Rv2246 | *kasB* | β-ketoacyl-ACP synthase (meromycolate extension) |
| Rv1618 | *tesB1* | thioesterase II |
| Rv2605c | *tesB2* | thioesterase II |
| Rv0033 | - | possible acyl carrier protein |
| Rv1344 | - | possible acyl carrier protein |
| Rv1722 | - | possible biotin carboxylase |
| Rv3221c | - | resembles biotin carboxyl carrier |
| Rv3472 | - | possible acyl carrier protein |
| 2. Modification of fatty and mycolic acids | | |
| Rv3391 | *acrA1* | fatty acyl-CoA reductase |
| Rv3392c | *cmaA1* | cyclopropane mycolic acid synthase 1 |
| Rv0503c | *cmaA2* | cyclopropane mycolic acid synthase 2 |
| Rv0824c | *desA1* | acyl-[ACP] desaturase |
| Rv1094 | *desA2* | acyl-[ACP] desaturase |
| Rv3229c | *desA3* | acyl-[ACP] desaturase |
| Rv0645c | *mmaA1* | methoxymycolic acid synthase 1 |
| Rv0644c | *mmaA2* | methoxymycolic acid synthase 2 |
| Rv0643c | *mmaA3* | methoxymycolic acid synthase 3 |
| Rv0642c | *mmaA4* | methoxymycolic acid synthase 4 |
| Rv0447c | *ufaA1* | unknown fatty acid methyltransferase |
| Rv3538 | *ufaA2* | unknown fatty acid methyltransferase |
| Rv0469 | *umaA1* | unknown mycolic acid methyltransferase |
| Rv0470c | *umaA2* | unknown mycolic acid methyltransferase |
| 3. Acyltransferases, mycolyltransferases and phospholipid synthesis | | |
| Rv2289 | *cdh* | CDP-diacylglycerol phosphatidylhydrolase |
| Rv2881c | *cdsA* | phosphatidate cytidylyltransferase |
| Rv3804c | *fbpA* | antigen 85A, mycolyltransferase |
| Rv1886c | *fbpB* | antigen 85B, mycolyltransferase |
| Rv0129c | *fbpC* | antigen 85C, mycolytransferase |
| Rv3803c | *fbpD* | antigen MPT51, mycolyltransferase |
| Rv0564c | *gpdA1* | glycerol-3-phosphate dehydrogenase |
| Rv2982c | *gpdA2* | glycerol-3-phosphate dehydrogenase |
| Rv2612c | *pgsA* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv1822 | *pgsA2* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv2746c | *pgsA3* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv1551 | *plsB1* | glycerol-3-phosphate acyltransferase |
| Rv2482c | *plsB2* | glycerol-3-phosphate acyltransferase |
| Rv0437c | *psd* | putative phosphatidylserine decarboxylase |
| Rv0436c | *pssA* | CDP-diacylglycerol-serine *o*-phosphatidyltransferase |
| Rv0045c | - | possible dihydrolipoamide acetyltransferase |
| Rv0914c | - | lipid transfer protein |
| Rv1543 | - | probable fatty-acyl CoA reductase |
| Rv1627c | - | lipid carrier protein |
| Rv1814 | - | possible C-5 sterol desaturase |
| Rv1867 | - | similar to acetyl CoA synthase/lipid carriers |
| Rv2261c | - | apolipoprotein N-acyltransferase-a |
| Rv2262c | - | apolipoprotein N-acyltransferase-b |
| Rv3523 | - | lipid carrier protein |
| Rv3720 | - | C-term similar to cyclopropane fatty acid synthases |
| *I. Polyketide and non-ribosomal peptide synthesis* | | |
| Rv2940c | *mas* | mycocerosic acid synthase |
| Rv2384 | *mbtA* | mycobactin/exochelin synthesis (salicylate-AMP ligase) |
| Rv2383c | *mbtB* | mycobactin/exochelin synthesis (serine/threonine ligation) |
| Rv2382c | *mbtC* | mycobactin/exochelin synthesis |
| Rv2381c | *mbtD* | mycobactin/exochelin synthesis (polyketide synthase) |
| Rv2380c | *mbtE* | mycobactin/exochelin synthesis (lysine ligation) |
| Rv2379c | *mbtF* | mycobactin/exochelin synthesis (lysine ligation) |
| Rv2378c | *mbtG* | mycobactin/exochelin synthesis (lysine hydroxylase) |
| Rv2377c | *mbtH* | mycobactin/exochelin synthesis |
| Rv0101 | *nrp* | unknown non-ribosomal peptide synthase |
| Rv1153c | *omt* | PKS *o*-methyltransferase |
| Rv3824c | *papA1* | PKS-associated protein, unknown function |
| Rv3820c | *papA2* | PKS-associated protein, unknown function |
| Rv1182 | *papA3* | PKS-associated protein, unknown function |
| Rv1528c | *papA4* | PKS-associated protein, unknown function |
| Rv2939 | *papA5* | PKS-associated protein, unknown function |
| Rv2946c | *pks1* | polyketide synthase |
| Rv3825c | *pks2* | polyketide synthase |
| Rv1180 | *pks3* | polyketide synthase |
| Rv1181 | *pks4* | polyketide synthase |
| Rv1527c | *pks5* | polyketide synthase |
| Rv0405 | *pks6* | polyketide synthase |
| Rv1661 | *pks7* | polyketide synthase |
| Rv1662 | *pks8* | polyketide synthase |
| Rv1664 | *pks9* | polyketide synthase |
| Rv1660 | *pks10* | polyketide synthase (chalcone synthase-like) |
| Rv1665 | *pks11* | polyketide synthase (chalcone synthase-like) |
| Rv2048c | *pks12* | polyketide synthase (erythronolide synthase-like) |
| Rv3800c | *pks13* | polyketide synthase |
| Rv1342c | *pks14* | polyketide synthase (chalcone synthase-like) |
| Rv2947c | *pks15* | polyketide synthase |
| Rv1013 | *pks16* | polyketide synthase |
| Rv1663 | *pks17* | polyketide synthase |
| Rv1372 | *pks18* | polyketide synthase |
| Rv2931 | *ppsA* | phenolpthiocerol synthesis (*pksB*) |
| Rv2932 | *ppsB* | phenolpthiocerol synthesis (*pksC*) |
| Rv2933 | *ppsC* | phenolpthiocerol synthesis (*pksD*) |
| Rv2934 | *ppsD* | phenolpthiocerol synthesis (*pksE*) |
| Rv2935 | *ppsE* | phenolpthiocerol synthesis (*pksF*) |
| Rv2928 | *tesA* | thioesterase |
| Rv1544 | - | probable ketoacyl reductase |
| *J. Broad regulatory functions* | | |
| 1. Repressors/activators | | |
| Rv1657 | *argR* | arginine repressor |
| Rv1267c | *embR* | regulator of *embAB* genes (AfsR/DndI/RedD family) |
| Rv1909c | *furA* | ferric uptake regulatory protein |
| Rv2359 | *furB* | ferric uptake regulatory protein |
| Rv2919c | *glnB* | nitrogen regulatory protein |
| Rv2711 | *ideR* | iron dependent repressor, IdeR |
| Rv2720 | *lexA* | LexA, SOS repressor protein |
| Rv1479 | *moxR* | transcriptional regulator, MoxR homologue |
| Rv3692 | *moxR2* | transcriptional regulator, MoxR homologue |
| Rv3164c | *moxR3* | transcriptional regulator, MoxR homologue |
| Rv0212c | *nadR* | similar to *E.coli* NadR |
| Rv0117 | *oxyS* | transcriptional regulator (LysR family) |
| Rv1379 | *pyrR* | regulatory protein pyrimidine biosynthesis |
| Rv2788 | *sirR* | iron-dependent transcriptional repressor |
| Rv3082c | *virS* | putative virulence regulating protein (AraC/XylS family) |
| Rv3219 | *whiB1* | WhiB transcriptional activator homologue |
| Rv3260c | *whiB2* | WhiB transcriptional activator homologue |
| Rv3416 | *whiB3* | WhiB transcriptional activator homologue |
| Rv3681c | *whiB4* | WhiB transcriptional activator homologue |
| Rv0023 | - | putative transcriptional regulator |
| Rv0043c | - | transcriptional regulator (GntR family) |
| Rv0067c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0078 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0081 | - | transcriptional regulator (ArsR family) |
| Rv0135c | - | putative transcriptional regulator |
| Rv0144 | - | putative transcriptional regulator |
| Rv0158 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0165c | - | transcriptional regulator (GntR family) |
| Rv0195 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0196 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0232 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0238 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0273c | - | putative transcriptional regulator |
| Rv0302 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0324 | - | putative transcriptional regulator |
| Rv0328 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0348 | - | putative transcriptional regulator |
| Rv0377 | - | transcriptional regulator (LysR family) |
| Rv0386 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0452 | - | putative transcriptional regulator |
| Rv0465c | - | transcriptional regulator (PbsX/Xre family) |
| Rv0472c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0474 | - | transcriptional regulator (PbsX/Xre family) |
| Rv0485 | - | transcriptional regulator (ROK family) |
| Rv0494 | - | transcriptional regulator (GntR family) |
| Rv0552 | - | putative transcriptional regulator |
| Rv0576 | - | putative transcriptional regulator |
| Rv0586 | - | transcriptional regulator (GntR family) |
| Rv0650 | - | transcriptional regulator (ROK family) |
| Rv0653c | - | putative transcriptional regulator |
| Rv0681 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0691c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0737 | - | putative transcriptional regulator |
| Rv0744c | - | putative transcriptional regulator |
| Rv0792c | - | transcriptional regulator (GntR |

| | | |
|---|---|---|
| | | family) |
| Rv0823c | - | transcriptional regulator (NifR3/Smm1 family) |
| Rv0827c | - | transcriptional regulator (ArsR family) |
| Rv0890c | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0891c | - | putative transcriptional regulator |
| Rv0894 | - | putative transcriptional regulator |
| Rv1019 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1049 | - | transcriptional regulator (MarR family) |
| Rv1129c | - | transcriptional regulator (PbsX/Xre family) |
| Rv1151c | - | putative transcriptional regulator |
| Rv1152 | - | transcriptional regulator (GntR family) |
| Rv1167c | - | putative transcriptional regulator |
| Rv1219c | - | putative transcriptional regulator |
| Rv1255c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1332 | - | putative transcriptional regulator |
| Rv1353c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1358 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv1359 | - | putative transcriptional regulator |
| Rv1395 | - | transcriptional regulator (AraC/XylS family) |
| Rv1404 | - | transcriptional regulator (MarR family) |
| Rv1423 | - | putative transcriptional regulator |
| Rv1460 | - | putative transcriptional regulator |
| Rv1474c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1534 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1556 | - | putative transcriptional regulator |
| Rv1674c | - | putative transcriptional regulator |
| Rv1675c | - | putative transcriptional regulator |
| Rv1719 | - | transcriptional regulator (IclR family) |
| Rv1773c | - | transcriptional regulator (IclR family) |
| Rv1776c | - | putative transcriptional regulator |
| Rv1816 | - | putative transcriptional regulator |
| Rv1846c | - | putative transcriptional regulator |
| Rv1931c | - | transcriptional regulator (AraC/XylS family) |
| Rv1956 | - | putative transcriptional regulator |
| Rv1963c | - | putative transcriptional regulator |
| Rv1985c | - | transcriptional regulator (LysR family) |
| Rv1990c | - | putative transcriptional regulator |
| Rv1994c | - | transcriptional regulator (MerR family) |
| Rv2017 | - | putative transcriptional regulator (PbsX/Xre family) |
| Rv2021c | - | putative transcriptional regulator |
| Rv2034 | - | transcriptional regulator (ArsR family) |
| Rv2175c | - | putative transcriptional regulator |
| Rv2250c | - | putative transcriptional regulator |
| Rv2258c | - | putative transcriptional regulator |
| Rv2282c | - | transcriptional regulator (LysR family) |
| Rv2308 | - | putative transcriptional regulator |
| Rv2324 | - | transcriptional regulator (Lrp/AsnC family) |
| Rv2358 | - | transcriptional regulator (ArsR family) |
| Rv2488c | - | transcriptional regulator (LuxR/UhpA family) |
| Rv2506 | - | transcriptional regulator (TetR/AcrR family) |
| Rv2621c | - | putative transcriptional regulator |
| Rv2640c | - | transcriptional regulator (ArsR family) |
| Rv2642 | - | transcriptional regulator (ArsR family) |
| Rv2669 | - | putative transcriptional regulator |
| Rv2745c | - | putative transcriptional regulator |
| Rv2779c | - | transcriptional regulator (Lrp/AsnC family) |
| Rv2887 | - | transcriptional regulator (MarR family) |
| Rv2912c | - | transcriptional regulator (TetR/AcrR family) |
| Rv2989 | - | transcriptional regulator (IclR family) |
| Rv3050c | - | putative transcriptional regulator |
| Rv3055 | - | putative transcriptional regulator |
| Rv3058c | - | putative transcriptional regulator |
| Rv3060c | - | transcriptional regulator (GntR family) |
| Rv3066 | - | putative transcriptional regulator |
| Rv3095 | - | putative transcriptional regulator |
| Rv3124 | - | transcriptional regulator (AfsR/DndI/RedD family) |
| Rv3160c | - | putative transcriptional regulator |
| Rv3167c | - | putative transcriptional regulator |
| Rv3173c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3183 | - | putative transcriptional regulator |
| Rv3208 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3249c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3291c | - | transcriptional regulator (Lrp/AsnC family) |
| Rv3295 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3334 | - | transcriptional regulator (MerR family) |
| Rv3405c | - | putative transcriptional regulator |
| Rv3522 | - | putative transcriptional regulator |
| Rv3557c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3574 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3575c | - | transcriptional regulator (LacI family) |
| Rv3583c | - | putative transcriptional regulator |
| Rv3676 | - | transcriptional regulator (Crp/Fnr family) |
| Rv3678c | - | transcriptional regulator (LysR family) |
| Rv3736 | - | transcriptional regulator (AraC/XylS family) |
| Rv3744 | - | transcriptional regulator (ArsR family) |
| Rv3830c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3833 | - | transcriptional regulator (AraC/XylS family) |
| Rv3840 | - | putative transcriptional regulator |
| Rv3855 | - | putative transcriptional regulator |

2. Two component systems

| | | |
|---|---|---|
| Rv1028c | *kdpD* | sensor histidine kinase |
| Rv1027c | *kdpE* | two-component response regulator |
| Rv3246c | *mtrA* | two-component response regulator |
| Rv3245c | *mtrB* | sensor histidine kinase |
| Rv0844c | *narL* | two-component response regulator |
| Rv0757 | *phoP* | two-component response regulator |
| Rv0758 | *phoR* | sensor histidine kinase |
| Rv0491 | *regX3* | two-component response regulator |
| Rv0490 | *senX3* | sensor histidine kinase |
| Rv0602c | *tcrA* | two-component response regulator |
| Rv0260c | - | two-component response regulator |
| Rv0600c | - | sensor histidine kinase |
| Rv0601c | - | sensor histidine kinase |
| Rv0818 | - | two-component response regulator |
| Rv0845 | - | sensor histidine kinase |
| Rv0902c | - | sensor histidine kinase |
| Rv0903c | - | two-component response regulator |
| Rv0981 | - | two-component response regulator |
| Rv0982 | - | sensor histidine kinase |
| Rv1032c | - | sensor histidine kinase |
| Rv1033c | - | two-component response regulator |
| Rv1626 | - | two-component response regulator |
| Rv2027c | - | sensor histidine kinase |
| Rv2884 | - | two-component response regulator |
| Rv3132c | - | sensor histidine kinase |
| Rv3133c | - | two-component response regulator |
| Rv3143 | - | putative sensory transduction protein |
| Rv3220c | - | sensor histidine kinase |
| Rv3764c | - | sensor histidine kinase |
| Rv3765c | - | two-component response regulator |

3. Serine-threonine protein kinases and phosphoprotein phosphatases

| | | |
|---|---|---|
| Rv0015c | *pknA* | serine-threonine protein kinase |
| Rv0014c | *pknB* | serine-threonine protein kinase |
| Rv0931c | *pknD* | serine-threonine protein kinase |
| Rv1743 | *pknE* | serine-threonine protein kinase |
| Rv1746 | *pknF* | serine-threonine protein kinase |
| Rv0410c | *pknG* | serine-threonine protein kinase |
| Rv1266c | *pknH* | serine-threonine protein kinase |
| Rv2914c | *pknI* | serine-threonine protein kinase |
| Rv2088 | *pknJ* | serine-threonine protein kinase |
| Rv3080c | *pknK* | serine-threonine protein kinase |
| Rv2176 | *pknL* | serine-threonine protein kinase, truncated |
| Rv0018c | *ppp* | putative phosphoprotein phosphatase |
| Rv2234 | *ptpA* | low molecular weight protein-tyrosine-phosphatase |
| Rv0153c | - | putative protein-tyrosine-phosphatase |

**II. Macromolecule metabolism**

*A. Synthesis and modification of macromolecules*

1. Ribosomal protein synthesis and modification

| | | |
|---|---|---|
| Rv3420c | *rimI* | ribosomal protein S18 acetyl transferase |
| Rv0995 | *rimJ* | acetylation of 30S S5 subunit |
| Rv0641 | *rplA* | 50S ribosomal protein L1 |
| Rv0704 | *rplB* | 50S ribosomal protein L2 |
| Rv0701 | *rplC* | 50S ribosomal protein L3 |
| Rv0702 | *rplD* | 50S ribosomal protein L4 |
| Rv0716 | *rplE* | 50S ribosomal protein L5 |
| Rv0719 | *rplF* | 50S ribosomal protein L6 |
| Rv0056 | *rplI* | 50S ribosomal protein L9 |
| Rv0651 | *rplJ* | 50S ribosomal protein L10 |
| Rv0640 | *rplK* | 50S ribosomal protein L11 |
| Rv0652 | *rplL* | 50S ribosomal protein L7/L12 |
| Rv3443c | *rplM* | 50S ribosomal protein L13 |
| Rv0714 | *rplN* | 50S ribosomal protein L14 |
| Rv0723 | *rplO* | 50S ribosomal protein L15 |
| Rv0708 | *rplP* | 50S ribosomal protein L16 |
| Rv3456c | *rplQ* | 50S ribosomal protein L17 |
| Rv0720 | *rplR* | 50S ribosomal protein L18 |
| Rv2904c | *rplS* | 50S ribosomal protein L19 |
| Rv1643 | *rplT* | 50S ribosomal protein L20 |
| Rv2442c | *rplU* | 50S ribosomal protein L21 |
| Rv0706 | *rplV* | 50S ribosomal protein L22 |
| Rv0703 | *rplW* | 50S ribosomal protein L23 |
| Rv0715 | *rplX* | 50S ribosomal protein L24 |
| Rv1015c | *rplY* | 50S ribosomal protein L25 |
| Rv2441c | *rpmA* | 50S ribosomal protein L27 |
| Rv0105c | *rpmB* | 50S ribosomal protein L28 |
| Rv2058c | *rpmB2* | 50S ribosomal protein L28 |
| Rv0709 | *rpmC* | 50S ribosomal protein L29 |
| Rv0722 | *rpmD* | 50S ribosomal protein L30 |
| Rv1298 | *rpmE* | 50S ribosomal protein L31 |
| Rv2057c | *rpmG* | 50S ribosomal protein L33 |
| Rv3924c | *rpmH* | 50S ribosomal protein L34 |
| Rv1642 | *rpmI* | 50S ribosomal protein L35 |
| Rv3461c | *rpmJ* | 50S ribosomal protein L36 |
| Rv1630 | *rpsA* | 30S ribosomal protein S1 |
| Rv2890c | *rpsB* | 30S ribosomal protein S2 |
| Rv0707 | *rpsC* | 30S ribosomal protein S3 |
| Rv3458c | *rpsD* | 30S ribosomal protein S4 |
| Rv0721 | *rpsE* | 30S ribosomal protein S5 |
| Rv0053 | *rpsF* | 30S ribosomal protein S6 |
| Rv0683 | *rpsG* | 30S ribosomal protein S7 |
| Rv0718 | *rpsH* | 30S ribosomal protein S8 |
| Rv3442c | *rpsI* | 30S ribosomal protein S9 |
| Rv0700 | *rpsJ* | 30S ribosomal protein S10 |
| Rv3459c | *rpsK* | 30S ribosomal protein S11 |
| Rv0682 | *rpsL* | 30S ribosomal protein S12 |
| Rv3460c | *rpsM* | 30S ribosomal protein S13 |
| Rv0717 | *rpsN* | 30S ribosomal protein S14 |
| Rv2056c | *rpsN2* | 30S ribosomal protein S14 |
| Rv2785c | *rpsO* | 30S ribosomal protein S15 |
| Rv2909c | *rpsP* | 30S ribosomal protein S16 |
| Rv0710 | *rpsQ* | 30S ribosomal protein S17 |
| Rv0055 | *rpsR* | 30S ribosomal protein S18 |
| Rv2055c | *rpsR2* | 30S ribosomal protein S18 |
| Rv0705 | *rpsS* | 30S ribosomal protein S19 |
| Rv2412 | *rpsT* | 30S ribosomal protein S20 |
| Rv3241c | - | member of S30AE ribosomal protein family |

2. Ribosome modification and maturation

| | | |
|---|---|---|
| Rv1010 | *ksgA* | 16S rRNA dimethyltransferase |
| Rv2838c | *rbfA* | ribosome-binding factor A |
| Rv2907c | *rimM* | 16S rRNA processing protein |

3. Aminoacyl tRNA synthases and their modification

| | | |
|---|---|---|
| Rv2555c | *alaS* | alanyl-tRNA synthase |
| Rv1292 | *argS* | arginyl-tRNA synthase |
| Rv2572c | *aspS* | aspartyl-tRNA synthase |
| Rv3580c | *cysS* | cysteinyl-tRNA synthase |
| Rv2130c | *cysS2* | cysteinyl-tRNA synthase |
| Rv1406 | *fmt* | methionyl-tRNA formyltransferase |
| Rv3011c | *gatA* | glu-tRNA-gln amidotransferase, subunit B |
| Rv3009c | *gatB* | glu-tRNA-gln amidotransferase, subunit A |
| Rv3012c | *gatC* | glu-tRNA-gln amidotransferase, subunit C |
| Rv2992c | *gltS* | glutamyl-tRNA synthase |
| Rv2357c | *glyS* | glycyl-tRNA synthase |
| Rv2580c | *hisS* | histidyl-tRNA synthase |
| Rv1536 | *ileS* | isoleucyl-tRNA synthase |
| Rv0041 | *leuS* | leucyl-tRNA synthase |
| Rv3598c | *lysS* | lysyl-tRNA synthase |
| Rv1640c | *lysX* | C-term lysyl-tRNA synthase |
| Rv1007c | *metS* | methionyl-tRNA synthase |
| Rv1649 | *pheS* | phenylalanyl-tRNA synthase α subunit |

| | | |
|---|---|---|
| Rv1650 | *pheT* | phenylalanyl-tRNA synthase β subunit |
| Rv2845c | *proS* | prolyl-tRNA synthase |
| Rv3834c | *serS* | seryl-tRNA synthase |
| Rv2614c | *thrS* | threonyl-tRNA synthase |
| Rv2906c | *trmD* | tRNA (guanine-N1)-methyltransferase |
| Rv3336c | *trpS* | tryptophanyl tRNA synthase |
| Rv1689 | *tyrS* | tyrosyl-tRNA synthase |
| Rv2448c | *valS* | valyl-tRNA synthase |

4. Nucleoproteins

| | | |
|---|---|---|
| Rv1407 | *fmu* | similar to Fmu protein |
| Rv3852 | *hns* | HU-histone protein |
| Rv2986c | *hupB* | DNA-binding protein II |
| Rv1388 | *mIHF* | integration host factor |

5. DNA replication, repair, recombination and restriction/modification

| | | |
|---|---|---|
| Rv1317c | *alkA* | DNA-3-methyladenine glycosidase II |
| Rv2836c | *dinF* | DNA-damage-inducible protein F |
| Rv1329c | *dinG* | probable ATP-dependent helicase |
| Rv3056 | *dinP* | DNA-damage-inducible protein |
| Rv1537 | *dinX* | probable DNA-damage-inducible protein |
| Rv0001 | *dnaA* | chromosomal replication initiator protein |
| Rv0058 | *dnaB* | DNA helicase (contains intein) |
| Rv1547 | *dnaE1* | DNA polymerase III, α subunit |
| Rv3370c | *dnaE2* | DNA polymerase III α chain |
| Rv2343c | *dnaG* | DNA primase |
| Rv0002 | *dnaN* | DNA polymerase III, β subunit |
| Rv3711c | *dnaQ* | DNA polymerase III ε chain |
| Rv3721c | *dnaZX* | DNA polymerase III, γ (dnaZ) and τ (dnaX) |
| Rv2924c | *fpg* | formamidopyrimidine-DNA glycosylase |
| Rv0006 | *gyrA* | DNA gyrase subunit A |
| Rv0005 | *gyrB* | DNA gyrase subunit B |
| Rv2092c | *helY* | probable helicase, Ski2 subfamily |
| Rv2101 | *helZ* | probable helicase, Snf2/Rad54 family |
| Rv2756c | *hsdM* | type I restriction/modification system DNA methylase |
| Rv2755c | *hsdS'* | type I restriction/modification system specificity determinant |
| Rv3296 | *lhr* | ATP-dependent helicase |
| Rv3014c | *ligA* | DNA ligase |
| Rv3062 | *ligB* | DNA ligase |
| Rv3731 | *ligC* | probable DNA ligase |
| Rv1020 | *mfd* | transcription-repair coupling factor |
| Rv2528c | *mrr* | restriction system protein |
| Rv2985 | *mutT1* | MutT homologue |
| Rv1160 | *mutT2* | MutT homologue |
| Rv0413 | *mutT3* | MutT homologue |
| Rv3589 | *mutY* | probable DNA glycosylase |
| Rv3297 | *nei* | probable endonuclease VIII |
| Rv3674c | *nth* | probable endonuclease III |
| Rv1316c | *ogt* | methylated-DNA-protein-cysteine methyltransferase |
| Rv1629 | *polA* | DNA polymerase I |
| Rv1402 | *priA* | putative primosomal protein n' (replication factor Y) |
| Rv3585 | *radA* | probable DNA repair RadA homologue |
| Rv2737c | *recA* | recombinase (contains intein) |
| Rv0630c | *recB* | exodeoxyribonuclease V |
| Rv0631c | *recC* | exodeoxyribonuclease V |
| Rv0629c | *recD* | exodeoxyribonuclease V |
| Rv0003 | *recF* | DNA replication and SOS induction |
| Rv2973c | *recG* | ATP-dependent DNA helicase |
| Rv1696 | *recN* | recombination and DNA repair |
| Rv3715c | *recR* | RecBC-Independent process of DNA repair |
| Rv2736c | *recX* | regulatory protein for RecA |
| Rv2593c | *ruvA* | Holliday junction binding protein, DNA helicase |
| Rv2592c | *ruvB* | Holliday junction binding protein |
| Rv2594c | *ruvC* | Holliday junction resolvase, endodeoxyribonuclease |
| Rv0054 | *ssb* | single strand binding protein |
| Rv1210 | *tagA* | DNA-3-methyladenine glycosidase I |
| Rv3646c | *topA* | DNA topoisomerase |
| Rv2976c | *ung* | uracil-DNA glycosylase |
| Rv1638 | *uvrA* | excinuclease ABC subunit A |
| Rv1633 | *uvrB* | excinuclease ABC subunit B |
| Rv1420 | *uvrC* | excinuclease ABC subunit C |
| Rv0949 | *uvrD* | DNA-dependent ATPase I and helicase II |
| Rv3198c | *uvrD2* | putative UvrD |
| Rv0427c | *xthA* | exodeoxyribonuclease III |
| Rv0071 | - | group II intron maturase |
| Rv0861c | - | probable DNA helicase |
| Rv0944 | - | possible formamidopyrimidine-DNA glycosylase |
| Rv1688 | - | probable 3-methylpurine DNA glycosylase |
| Rv2090 | - | partially similar to DNA polymerase I |
| Rv2191 | - | similar to both PolC and UvrC proteins |
| Rv2464c | - | probable DNA glycosylase, endonuclease VIII |
| Rv3201c | - | probable ATP-dependent DNA helicase |
| Rv3202c | - | similar to UvrD proteins |
| Rv3263 | - | probable DNA methylase |
| Rv3644c | - | similar in N-term to DNA polymerase III |

6. Protein translation and modification

| | | |
|---|---|---|
| Rv0429c | *def* | polypeptide deformylase |
| Rv2534c | *efp* | elongation factor P |
| Rv2882c | *frr* | ribosome recycling factor |
| Rv0684 | *fusA* | elongation factor G |
| Rv0120c | *fusA2* | elongation factor G |
| Rv1080c | *greA* | transcription elongation factor G |
| Rv3462c | *infA* | initiation factor IF-1 |
| Rv2839c | *infB* | initiation factor IF-2 |
| Rv1641 | *infC* | initiation factor IF-3 |
| Rv0009 | *ppiA* | peptidyl-prolyl *cis-trans* isomerase |
| Rv2582 | *ppiB* | peptidyl-prolyl *cis-trans* isomerase |
| Rv1299 | *prfA* | peptide chain release factor 1 |
| Rv3105c | *prfB* | peptide chain release factor 2 |
| Rv2889c | *tsf* | elongation factor EF-Ts |
| Rv0685 | *tuf* | elongation factor EF-Tu |

7. RNA synthesis, RNA modification and DNA transcription

| | | |
|---|---|---|
| Rv1253 | *deaD* | ATP-dependent DNA/RNA helicase |
| Rv2783c | *gpsI* | pppGpp synthase and polyribonucleotide phosphorylase |
| Rv2841c | *nusA* | transcription termination factor |
| Rv2533c | *nusB* | N-utilization substance protein B |
| Rv0639 | *nusG* | transcription antitermination protein |
| Rv3907c | *pcnA* | polynucleotide polymerase |
| Rv3232c | *pvdS* | alternative sigma factor for siderophore production |
| Rv3211 | *rhlE* | probable ATP-dependent RNA helicase |
| Rv1297 | *rho* | transcription termination factor rho |
| Rv3457c | *rpoA* | α subunit of RNA polymerase |
| Rv0667 | *rpoB* | β subunit of RNA polymerase |
| Rv0668 | *rpoC* | β' subunit of RNA polymerase |
| Rv1364c | *rsbU* | SigB regulation protein |
| Rv3287c | *rsbW* | anti-sigma B factor |
| Rv2703 | *sigA* | RNA polymerase sigma factor (aka MysA, RpoV) |
| Rv2710 | *sigB* | RNA polymerase sigma factor (aka MysB) |
| Rv2069 | *sigC* | ECF subfamily sigma subunit |
| Rv3414c | *sigD* | ECF subfamily sigma subunit |
| Rv1221 | *sigE* | ECF subfamily sigma subunit |
| Rv3286c | *sigF* | ECF subfamily sigma subunit |
| Rv0182c | *sigG* | sigma-70 factors ECF subfamily |
| Rv3223c | *sigH* | ECF subfamily sigma subunit |
| Rv1189 | *sigI* | ECF family sigma factor |
| Rv3328c | *sigJ* | similar to SigI, ECF family |
| Rv0445c | *sigK* | ECF-type sigma factor |
| Rv0735 | *sigL* | sigma-70 factors ECF subfamily |
| Rv3911 | *sigM* | probable sigma factor, similar to SigE |
| Rv3366 | *spoU* | probable rRNA methylase |
| Rv3455c | *truA* | probable pseudouridylate synthase |
| Rv2793c | *truB* | tRNA pseudouridine 55 synthase |
| Rv1644 | *tsnR* | putative 23S rRNA methyltransferase |
| Rv3649 | - | ATP-dependent DNA/RNA helicase |

8. Polysaccharides (cytoplasmic)

| | | |
|---|---|---|
| Rv1326c | *glgB* | 1,4-α-glucan branching enzyme |
| Rv1328 | *glgP* | probable glycogen phosphorylase |
| Rv1564c | *glgX* | probable glycogen debranching enzyme |
| Rv1563c | *glgY* | putative α-amylase |
| Rv1562c | *glgZ* | maltooligosyltrehalose trehalohydrolase |
| Rv0126 | - | probable glycosyl hydrolase |
| Rv1781c | - | probable 4-α-glucanotransferase |
| Rv2471 | - | probable maltase α-glucosidase |

*B. Degradation of macromolecules*

1. RNA

| | | |
|---|---|---|
| Rv1014c | *pth* | peptidyl-tRNA hydrolase |
| Rv2925c | *rnc* | RNAse III |
| Rv2444c | *rne* | similar at C-term to ribonuclease E |
| Rv2902c | *rnhB* | ribonuclease HII |
| Rv3923c | *rnpA* | ribonuclease P protein component |
| Rv1340 | *rphA* | ribonuclease PH |

2. DNA

| | | |
|---|---|---|
| Rv0670 | *end* | endonuclease IV (apurinase) |
| Rv1108c | *xseA* | exonuclease VII large subunit |
| Rv1107c | *xseB* | exonuclease VII small subunit |

3. Proteins, peptides and glycopeptides

| | | |
|---|---|---|
| Rv3305c | *amiA* | probable aminohydrolase |
| Rv3306c | *amiB* | probable aminohydrolase |
| Rv3596c | *clpC* | ATP-dependent Clp protease |
| Rv2461c | *clpP* | ATP-dependent Clp protease proteolytic subunit |
| Rv2460c | *clpP2* | ATP-dependent Clp protease proteolytic subunit |
| Rv2457c | *clpX* | ATP-dependent Clp protease ATP-binding subunit ClpX |
| Rv2667 | *clpX'* | similar to ClpC from *M. leprae* but shorter |
| Rv3419c | *gcp* | glycoprotease |
| Rv2725c | *hflX* | GTP-binding protein |
| Rv1223 | *htrA* | serine protease |
| Rv2861c | *mapA1* | methionine aminopeptidase |
| Rv0734 | *mapA2* | probable methionine aminopeptidase |
| Rv0319 | *pcp* | pyrrolidone-carboxylate peptidase |
| Rv0125 | *pepA* | probable serine protease |
| Rv2213 | *pepB* | aminopeptidase A/I |
| Rv0800 | *pepC* | aminopeptidase I |
| Rv2467 | *pepD* | probable aminopeptidase |
| Rv2089c | *pepE* | cytoplasmic peptidase |
| Rv2535c | *pepQ* | cytoplasmic peptidase |
| Rv2782c | *pepR* | protease/peptidase, M16 family (insulinase) |
| Rv2109c | *prcA* | proteasome α-type subunit 1 |
| Rv2110c | *prcB* | proteasome β-type subunit 2 |
| Rv0782 | *ptrBa* | protease II, α subunit |
| Rv0781 | *ptrBb* | protease II, β subunit |
| Rv0724 | *sppA* | protease IV, signal peptide peptidase |
| Rv0198c | - | probable zinc metalloprotease |
| Rv0457c | - | probable peptidase |
| Rv0840c | - | probable proline iminopeptidase |
| Rv0983 | - | probable serine protease |
| Rv1977 | - | probable zinc metallopeptidase |
| Rv3668c | - | probable alkaline serine protease |
| Rv3671c | - | probable serine protease |
| Rv3883c | - | probable secreted protease |
| Rv3886c | - | protease |

4. Polysaccharides, lipopolysaccharides and phospholipids

| | | |
|---|---|---|
| Rv0062 | *celA* | cellulase/endoglucanase |
| Rv3915 | *cwlM* | hydrolase |
| Rv0315 | - | probable β-1,3-glucanase |
| Rv1090 | - | probable inactivated cellulase/endoglucanase |
| Rv1327c | - | probable glycosyl hydrolase, α-amylase family |
| Rv1333 | - | probable hydrolase |
| Rv3463 | - | probable neuraminidase |
| Rv3717 | - | possible N-acetylmuramoyl-L-alanine amidase |

5. Esterases and lipases

| | | |
|---|---|---|
| Rv0220 | *lipC* | probable esterase |
| Rv1923 | *lipD* | probable esterase |
| Rv3775 | *lipE* | probable hydrolase |
| Rv3487c | *lipF* | probable esterase |
| Rv0646c | *lipG* | probable hydrolase |
| Rv1399c | *lipH* | probable lipase |
| Rv1400c | *lipI* | probable lipase |
| Rv1900c | *lipJ* | probable esterase |
| Rv2385 | *lipK* | probable acetyl-hydrolase |
| Rv1497 | *lipL* | esterase |
| Rv2284 | *lipM* | probable esterase |
| Rv2970c | *lipN* | probable lipase/esterase |
| Rv1426c | *lipO* | probable esterase |
| Rv2463 | *lipP* | probable esterase |
| Rv2485c | *lipQ* | probable carboxlyesterase |
| Rv3084 | *lipR* | probable acetyl-hydrolase |
| Rv3176c | *lipS* | probable esterase/lipase |
| Rv2045c | *lipT* | probable carboxylesterase |
| Rv1076 | *lipU* | probable esterase |
| Rv3203 | *lipV* | probable lipase |
| Rv0217c | *lipW* | probable esterase |
| Rv2351c | *plcA* | phospholipase C precursor |
| Rv2350c | *plcB* | phospholipase C precursor |
| Rv2349c | *plcC* | phospholipase C precursor |
| Rv1755c | *plcD* | partial CDS for phospholipase C |
| Rv1104 | - | probable esterase pseudogene |
| Rv1105 | - | probable esterase pseudogene |

6. Aromatic hydrocarbons

| | | |
|---|---|---|
| Rv3469c | *mhpE* | probable 4-hydroxy-2-oxovalerate aldolase |
| Rv0316 | - | probable muconolactone isomerase |
| Rv0771 | - | probable 4-carboxymuconolactone decarboxylase |
| Rv0939 | - | probable dehydrase |
| Rv1723 | - | 6-aminohexanoate-dimer hydro- |

| | | |
|---|---|---|
| | | lase |
| Rv2715 | - | 2-hydroxymuconic semialdehyde hydrolase |
| Rv3530c | - | probable *cis*-diol dehydrogenase |
| Rv3534c | - | 4-hydroxy-2-oxovalerate aldolase |
| Rv3536c | - | aromatic hydrocarbon degradation |

*C. Cell envelope*

1. Lipoproteins (*lppA-lpr0*) 65

2. Surface polysaccharides, lipopolysaccharides, proteins and antigens

| | | |
|---|---|---|
| Rv0806c | *cpsY* | probable UDP-glucose-4-epimerase |
| Rv3811 | *csp* | secreted protein |
| Rv1677 | *dsbF* | highly similar to C-term Mpt53 |
| Rv3794 | *embA* | involved in arabinogalactan synthesis |
| Rv3795 | *embB* | involved in arabinogalactan synthesis |
| Rv3793 | *embC* | involved in arabinogalactan synthesis |
| Rv3875 | *esat6* | early secretory antigen target |
| Rv0112 | *gca* | probable GDP-mannose dehydratase |
| Rv0113 | *gmhA* | phosphoheptose isomerase |
| Rv2965c | *kdtB* | lipopolysaccharide core biosynthesis protein |
| Rv2878c | *mpt53* | secreted protein Mpt53 |
| Rv1980c | *mpt64* | secreted immunogenic protein Mpb64/Mpt64 |
| Rv2875 | *mpt70* | major secreted immunogenic protein Mpt70 precursor |
| Rv2873 | *mpt83* | surface lipoprotein Mpt83 |
| Rv0899 | *ompA* | member of OmpA family |
| Rv3810 | *pirG* | cell surface protein precursor (Erp protein) |
| Rv3782 | *rfbE* | similar to rhamnosyl transferase |
| Rv1302 | *rfe* | undecaprenyl-phosphate α-N-acetylglucosaminyltransferase |
| Rv2145c | *wag31* | antigen 84 (aka wag31) |
| Rv0431 | - | tuberculin related peptide (AT103) |
| Rv0954 | - | cell envelope antigen |
| Rv1514c | - | involved in polysaccharide synthesis |
| Rv1518 | - | involved in exopolysaccharide synthesis |
| Rv1758 | - | partial cutinase |
| Rv1910c | - | probable secreted protein |
| Rv1919c | - | weak similarity to pollen antigens |
| Rv1984c | - | probable secreted protein |
| Rv1987 | - | probable secreted protein |
| Rv2223c | - | probable exported protease |
| Rv2224c | - | probable exported protease |
| Rv2301 | - | probable cutinase |
| Rv2345 | - | precursor of probable membrane protein |
| Rv2672 | - | putative exported protease |
| Rv3019c | - | similar to Esat6 |
| Rv3036c | - | probable secreted protein |
| Rv3449 | - | probable precursor of serine protease |
| Rv3451 | - | probable cutinase |
| Rv3452 | - | probable cutinase precursor |
| Rv3724 | - | probable cutinase precursor |

3. Murein sacculus and peptidoglycan

| | | |
|---|---|---|
| Rv2911 | *dacB* | penicillin binding protein |
| Rv2981c | *ddlA* | D-alanine-D-alanine ligase A |
| Rv3809c | *glf* | UDP-galactopyranose mutase |
| Rv1018c | *glmU* | UDP-N-acetylglucosamine pyrophosphorylase |
| Rv3382c | *lytB1* | LytB protein homologue |
| Rv1110 | *lytB2* | very similar to LytB |
| Rv1315 | *murA* | UDP-N-acetylglucosamine-1-carboxyvinyltransferase |
| Rv0482 | *murB* | UDP-N-acetylenolpyruvoylglucosamine reductase |
| Rv2152c | *murC* | UDP-N-acetyl-muramate-alanine ligase |
| Rv2155c | *murD* | UDP-N-acetylmuramoylalanine-D-glutamate ligase |
| Rv2158c | *murE* | *meso*-diaminopimelate-adding enzyme |
| Rv2157c | *murF* | D-alanine:D-alanine-adding enzyme |
| Rv2153c | *murG* | transferase in peptidoglycan synthesis |
| Rv1338 | *murI* | glutamate racemase |
| Rv2156c | *murX* | phospho-N-acetylmuramoyl-petapeptide transferase |
| Rv3332 | *nagA* | N-acetylglucosamine-6-P-deacetylase |
| Rv0016c | *pbpA* | penicillin-binding protein |
| Rv2163c | *pbpB* | penicillin-binding protein 2 |
| Rv0050 | *ponA1* | penicillin-bonding protein |
| Rv3682 | *ponA2* | class A penicillin binding protein |
| Rv0017c | *rodA* | FtsW/RodA/SpovE family |
| Rv0907 | - | probable penicillin binding protein |
| Rv1367c | - | probable penicillin binding protein |
| Rv1730c | - | probable penicillin binding protein |
| Rv1922 | - | probable penicillin binding protein |
| Rv2864c | - | probable penicillin binding protein |
| Rv3330 | - | probable penicillin binding protein |
| Rv3627c | - | probable penicillin binding protein |

4. Conserved membrane proteins

| | | |
|---|---|---|
| Rv0402c | *mmpL1* | conserved large membrane protein |
| Rv0507 | *mmpL2* | conserved large membrane protein |
| Rv0206c | *mmpL3* | conserved large membrane protein |
| Rv0450c | *mmpL4* | conserved large membrane protein |
| Rv0676c | *mmpL5* | conserved large membrane protein |
| Rv1557 | *mmpL6* | conserved large membrane protein |
| Rv2942 | *mmpL7* | conserved large membrane protein |
| Rv3823c | *mmpL8* | conserved large membrane protein |
| Rv2339 | *mmpL9* | conserved large membrane protein |
| Rv1183 | *mmpL10* | conserved large membrane protein |
| Rv0202c | *mmpL11* | conserved large membrane protein |
| Rv1522c | *mmpL12* | conserved large membrane protein |
| Rv0403c | *mmpS1* | conserved small membrane protein |
| Rv0506 | *mmpS2* | conserved small membrane protein |
| Rv2198c | *mmpS3* | conserved small membrane protein |
| Rv0451c | *mmpS4* | conserved small membrane protein |
| Rv0677c | *mmpS5* | conserved small membrane protein |

5. Other membrane proteins 211

**III. Cell processes**

*A. Transport/binding proteins*

1. Amino acids

| | | |
|---|---|---|
| Rv2127 | *ansP* | L-asparagine permease |
| Rv0346c | *aroP2* | probable aromatic amino acid permease |
| Rv0917 | *betP* | glycine betaine transport |
| Rv1704c | *cycA* | transport of D-alanine, D-serine and glycine |
| Rv3666c | *dppA* | probable peptide transport system permease |
| Rv3665c | *dppB* | probable peptide transport system permease |
| Rv3664c | *dppC* | probable peptide transport system permease |
| Rv3663c | *dppD* | probable ABC-transporter |
| Rv0522 | *gabP* | probable 4-amino butyrate transporter |
| Rv0411c | *glnH* | putative glutamine binding protein |
| Rv2564 | *glnQ* | probable ATP-binding transport protein |
| Rv1280c | *oppA* | probable oligopeptide transport protein |
| Rv1283c | *oppB* | oligopeptide transport protein |
| Rv1282c | *oppC* | oligopeptide transport system permease |
| Rv1281c | *oppD* | probable peptide transport protein |
| Rv2320c | *rocE* | arginine/ornithine transporter |
| Rv3253c | - | probable cationic amino acid transport |
| Rv3454 | - | possible proline permease |

2. Cations

| | | |
|---|---|---|
| Rv2920c | *amt* | putative ammonium transporter |
| Rv1607 | *chaA* | putative calcium/proton antiporter |
| Rv1239c | *corA* | probable magnesium and cobalt transport protein |
| Rv0092 | *ctpA* | cation-transporting ATPase |
| Rv0103c | *ctpB* | cation transport ATPase |
| Rv3270 | *ctpC* | cation transport ATPase |
| Rv1469 | *ctpD* | probable cadmium-transporting ATPase |
| Rv0908 | *ctpE* | probable cation transport ATPase |
| Rv1997 | *ctpF* | probable cation transport ATPase |
| Rv1992c | *ctpG* | probable cation transport ATPase |
| Rv0425c | *ctpH* | C-terminal region putative cation-transporting ATPase |
| Rv0107c | *ctpI* | probable magnesium transport ATPase |
| Rv0969 | *ctpV* | cation transport ATPase |
| Rv3044 | *fecB* | putative FeIII-dicitrate transporter |
| Rv0265c | *fecB2* | iron transport protein FeIII dicitrate transporter |
| Rv1029 | *kdpA* | potassium-transporting ATPase A chain |
| Rv1030 | *kdpB* | potassium-transporting ATPase B chain |
| Rv1031 | *kdpC* | potassium-transporting ATPase C chain |
| Rv3236c | *kefB* | probable glutathione-regulated potassium-efflux protein |
| Rv2877c | *merT* | possible mercury resistance transport system |
| Rv1811 | *mgtC* | probable magnesium transport ATPase protein C |
| Rv0362 | *mgtE* | putative magnesium ion transporter |
| Rv2856 | *nicT* | probable nickel transport protein |
| Rv0924c | *nramp* | transmembrane protein belonging to Nramp family |
| Rv2691 | *trkA* | probable potassium uptake protein |
| Rv2692 | *trkB* | probable potassium uptake protein |
| Rv2287 | *yjcE* | probable Na+/H+ exchanger |
| Rv2723 | - | probable membrane protein, tellurium resistance |
| Rv3162c | - | probable membrane protein |
| Rv3237c | - | possible potassium channel protein |
| Rv3743c | - | probable cation-transporting ATPase |

3. Carbohydrates, organic acids and alcohols

| | | |
|---|---|---|
| Rv2443 | *dctA* | C4-dicarboxylate transport protein |
| Rv3476c | *kgtP* | sugar transport protein |
| Rv1902c | *nanT* | probable sialic acid transporter |
| Rv1236 | *sugA* | membrane protein probably involved in sugar transport |
| Rv1237 | *sugB* | sugar transport protein |
| Rv1238 | *sugC* | ABC transporter component of sugar uptake system |
| Rv3331 | *sugI* | probable sugar transport protein |
| Rv2835c | *ugpA* | *sn*-glycerol-3-phosphate permease |
| Rv2833c | *ugpB* | *sn*-glycerol-3-phosphate-binding periplasmic lipoprotein |
| Rv2832c | *ugpC* | *sn*-glycerol-3-phosphate transport ATP-binding protein |
| Rv2834c | *ugpE* | *sn*-glycerol-3-phosphate transport system protein |
| Rv2316 | *uspA* | sugar transport protein |
| Rv2318 | *uspC* | sugar transport protein |
| Rv2317 | *uspE* | sugar transport protein |
| Rv1200 | - | probable sugar transporter |
| Rv2038c | - | probable ABC sugar transporter |
| Rv2039c | - | probable sugar transporter |
| Rv2040c | - | probable sugar transporter |
| Rv2041c | - | probable sugar transporter |

4. Anions

| | | |
|---|---|---|
| Rv2684 | *arsA* | probable arsenical pump |
| Rv2685 | *arsB* | probable arsenical pump |
| Rv3578 | *arsB2* | probable arsenical pump |
| Rv2643 | *arsC* | probable arsenical pump |
| Rv2397c | *cysA* | sulphate transport ATP-binding protein |
| Rv2399c | *cysT* | sulphate transport system permease protein |
| Rv2398c | *cysW* | sulphate transport system permease protein |
| Rv1857 | *modA* | molybdate binding protein |
| Rv1858 | *modB* | transport system permease, molybdate uptake |
| Rv1859 | *modC* | molybdate uptake ABC-transporter |
| Rv1860 | *modD* | precursor of Apa (45/47 kD secreted protein) |
| Rv2329c | *narK1* | probable nitrite extrusion protein |
| Rv1737c | *narK2* | nitrite extrusion protein |
| Rv0261c | *narK3* | nitrite extrusion protein |
| Rv0267 | *narU* | similar to nitrite extrusion protein 2 |
| Rv0934 | *phoS1* | PstS component of phosphate uptake |
| Rv0928 | *phoS2* | PstS component of phosphate uptake |
| Rv0820 | *phoT* | phosphate transport system ABC transporter |
| Rv3301c | *phoY1* | phosphate transport system regulator |
| Rv0821c | *phoY2* | phosphate transport system regulator |
| Rv0545c | *pitA* | low-affinity inorganic phosphate transporter |
| Rv2281 | *pitB* | phosphate permease |
| Rv0930 | *pstA1* | PstA component of phosphate uptake |
| Rv0936 | *pstA2* | PstA component of phosphate uptake |
| Rv0933 | *pstB* | ABC transport component of phosphate uptake |
| Rv0935 | *pstC* | PstC component of phosphate uptake |
| Rv0929 | *pstC2* | membrane-bound component of |

| | | |
|---|---|---|
| | | phosphate transport system |
| Rv0932c | *pstS* | PstS component of phosphate uptake |
| Rv2400c | *subI* | sulphate binding precursor |
| Rv0143c | - | probable chloride channel |
| Rv1707 | - | probable sulphate permease |
| Rv1739c | - | possible sulphate transporter |
| Rv3679 | - | possible anion transporter |
| Rv3680 | - | probable anion transporter |
| 5. Fatty acid transport | | |
| Rv2790c | *ltp1* | non-specific lipid transport protein |
| Rv3540c | *ltp2* | non-specific lipid transport protein |
| 6. Efflux proteins | | |
| Rv2936 | *drrA* | similar daunorubicin resistance ABC-transporter |
| Rv2937 | *drrB* | similar daunorubicin resistance transmembrane protein |
| Rv2938 | *drrC* | similar daunorubicin resistance transmembrane protein |
| Rv2846c | *efpA* | putative efflux protein |
| Rv3065 | *emrE* | resistance to ethidium bromide |
| Rv0783c | - | multidrug resistance protein |
| Rv0849 | - | possible quinolone efflux pump |
| Rv1145 | - | probable drug transporter |
| Rv1146 | - | probable drug transporter |
| Rv1250 | - | probable drug efflux protein |
| Rv1258c | - | probable multidrug resistance pump |
| Rv1410c | - | probable drug efflux protein |
| Rv1634 | - | probable drug efflux protein |
| Rv1819c | - | probable multidrug resistance pump |
| Rv2136c | - | putative bacitracin resistance protein |
| Rv2209 | - | probable drug efflux protein |
| Rv2333c | - | probable tetracenomycin C resistance protein |
| Rv2994 | - | probable fluoroquinolone efflux protein |
| Rv1877 | - | probable drug efflux protein |
| Rv2459 | - | probable drug efflux protein |
| *B. Chaperones/Heat shock* | | |
| Rv0384c | *clpB* | heat shock protein |
| Rv0352 | *dnaJ* | acts with GrpE to stimulate DnaK ATPase |
| Rv2373c | *dnaJ2* | DnaJ homologue |
| Rv0350 | *dnaK* | 70 kD heat shock protein, chromosome replication |
| Rv3417c | *groEL1* | 60 kD chaperonin 1 |
| Rv0440 | *groEL2* | 60 kD chaperonin 2 |
| Rv3418c | *groES* | 10 kD chaperone |
| Rv0351 | *grpE* | stimulates DnaK ATPase activity |
| Rv2374c | *hrcA* | heat-inducible transcription repressor |
| Rv0251c | *hsp* | possible heat shock protein |
| Rv0353 | *hspR* | heat shock regulator |
| Rv2031c | *hspX* | 14kD antigen, heat shock protein Hsp20 family |
| Rv2299c | *htpG* | heat shock protein Hsp90 family |
| Rv0563 | *htpX* | probable (transmembrane) heat shock protein |
| Rv2701c | *suhB* | putative extragenic suppressor protein |
| Rv3269 | - | probable heat shock protein |
| *C. Cell division* | | |
| Rv3641c | *fic* | possible cell division protein |
| Rv3102c | *ftsE* | membrane protein |
| Rv3610c | *ftsH* | inner membrane protein, chaperone |
| Rv2748c | *ftsK* | chromosome partitioning |
| Rv2151c | *ftsQ* | ingrowth of wall at septum |
| Rv2154c | *ftsW* | membrane protein (shape determination) |
| Rv3101c | *ftsX* | membrane protein |
| Rv2921c | *ftsY* | cell division protein FtsY |
| Rv2150c | *ftsZ* | circumferential ring, GTPase |
| Rv3919c | *gid* | glucose inhibited division protein B |
| Rv3625c | *mesJ* | probable cell cycle protein |
| Rv3917c | *parA* | chromosome partitioning; DNA - binding |
| Rv3918c | *parB* | possibly involved in chromosome partitioning |
| Rv2922c | *smc* | member of Smc1/Cut3/Cut14 family |
| Rv0012 | - | possible cell division protein |
| Rv0435c | - | ATPase of AAA-family |
| Rv2115c | - | ATPase of AAA-family |
| Rv3213c | - | possible role in chromosome segregation |
| Rv1708 | - | possible role in chromosome partitioning |
| *D. Protein and peptide secretion* | | |
| Rv2916c | *ffh* | signal recognition particle protein |
| Rv2903c | *lepB* | signal peptidase I |
| Rv1614 | *lgt* | prolipoprotein diacylglyceryl transferase |
| Rv1539 | *lspA* | lipoprotein signal peptidase |
| Rv0379 | *sec* | probable transport protein SecE/Sec61- γ family |
| Rv3240c | *secA* | SecA, preprotein translocase subunit |
| Rv1821 | *secA2* | SecA, preprotein translocase subunit |
| Rv2587c | *secD* | protein-export membrane protein |
| Rv0638 | *secE* | SecE preprotein translocase |
| Rv2586c | *secF* | protein-export membrane protein |
| Rv1440 | *secG* | protein-export membrane protein SecG |
| Rv0732 | *secY* | SecY subunit of preprotein translocase |
| Rv2462c | *tig* | chaperone protein, similar to trigger factor |
| Rv2813 | - | probable general secretion pathway protein |
| *E. Adaptations and atypical conditions* | | |
| Rv1901 | *cinA* | competence damage protein |
| Rv3648c | *cspA* | cold shock protein, transcriptional regulator |
| Rv0871 | *cspB* | probable cold shock protein |
| Rv3063 | *cstA* | starvation-induced stress response protein |
| Rv3490 | *otsA* | probable α,α-trehalose-phosphate synthase |
| Rv2006 | *otsB* | trehalose-6-phosphate phosphatase |
| Rv3372 | *otsB2* | trehalose-6-phosphate phosphatase |
| Rv3758c | *proV* | osmoprotection ABC transporter |
| Rv3757c | *proW* | transport system permease |
| Rv3759c | *proX* | similar to osmoprotection proteins |
| Rv3756c | *proZ* | transport system permease |
| Rv1026 | - | probable pppGpp-5'phosphohydrolase |
| *F. Detoxification* | | |
| Rv2428 | *ahpC* | alkyl hydroperoxide reductase |
| Rv2429 | *ahpD* | member of AhpC/TSA family |
| Rv2238c | *ahpE* | member of AhpC/TSA family |
| Rv2521 | *bcp* | bacterioferritin comigratory protein |
| Rv1608c | *bcpB* | probable bacterioferritin comigratory protein |
| Rv3473c | *bpoA* | probable non-heme bromoperoxidase |
| Rv1123c | *bpoB* | probable non-heme bromoperoxidase |
| Rv0554 | *bpoC* | probable non-heme bromoperoxidase |
| Rv3617 | *ephA* | probable epoxide hydrolase |
| Rv1938 | *ephB* | probable epoxide hydrolase |
| Rv1124 | *ephC* | probable epoxide hydrolase |
| Rv2214c | *ephD* | probable epoxide hydrolase |
| Rv3670 | *ephE* | probable epoxide hydrolase |
| Rv0134 | *ephF* | probable epoxide hydrolase |
| Rv3171c | *hpx* | probable non-heme haloperoxidase |
| Rv1908c | *katG* | catalase-peroxidase |
| Rv3846 | *sodA* | superoxide dismutase |
| Rv0432 | *sodC* | superoxide dismutase precursor - (Cu-Zn) |
| Rv1932 | *tpx* | thiol peroxidase |
| Rv0634c | - | putative glyoxylase II |
| Rv2581c | - | putative glyoxylase II |
| Rv3177 | - | probable non-heme haloperoxidase |
| IV. Other | | |
| *A. Virulence* | | |
| Rv0169 | *mce1* | cell invasion protein |
| Rv0589 | *mce2* | cell invasion protein |
| Rv1966 | *mce3* | cell invasion protein |
| Rv3499c | *mce4* | cell invasion protein |
| Rv3100c | *smpB* | probable small protein b |
| Rv1694 | *tlyA* | cytotoxin/hemolysin homologue |
| Rv0024 | - | putative p60 homologue |
| Rv0167 | - | part of *mce1* operon |
| Rv0168 | - | part of *mce1* operon |
| Rv0170 | - | part of *mce1* operon |
| Rv0171 | - | part of *mce1* operon |
| Rv0172 | - | part of *mce1* operon |
| Rv0174 | - | part of *mce1* operon |
| Rv0587 | - | part of *mce2* operon |
| Rv0588 | - | part of *mce2* operon |
| Rv0590 | - | part of *mce2* operon |
| Rv0591 | - | part of *mce2* operon |
| Rv0592 | - | part of *mce2* operon |
| Rv0594 | - | part of *mce2* operon |
| Rv1085c | - | possible hemolysin |
| Rv1477 | - | putative exported p60 protein homologue |
| Rv1478 | - | putative exported p60 protein homologue |
| Rv1566c | - | putative exported p60 protein homologue |
| Rv1964 | - | part of *mce3* operon |
| Rv1965 | - | part of *mce3* operon |
| Rv1967 | - | part of *mce3* operon |
| Rv1968 | - | part of *mce3* operon |
| Rv1969 | - | part of *mce3* operon |
| Rv1971 | - | part of *mce3* operon |
| Rv2190c | - | putative p60 homologue |
| Rv3494c | - | part of *mce4* operon |
| Rv3496c | - | part of *mce4* operon |
| Rv3497c | - | part of *mce4* operon |
| Rv3498c | - | part of *mce4* operon |
| Rv3500c | - | part of *mce4* operon |
| Rv3501c | - | part of *mce4* operon |
| Rv3896c | - | putative p60 homologue |
| Rv3922c | - | possible hemolysin |

*B. IS elements, Repeated sequences, and Phage*

| | |
|---|---|
| 1. IS elements | |
| IS*6110* | 16 copies |
| IS*1081* | 6 copies |
| Others | 34 copies |
| 2. REP *13E12* family | 7 copies |

| | | |
|---|---|---|
| 3. Phage-related functions | | |
| Rv2894c | *xerC* | integrase/recombinase |
| Rv1701 | *xerD* | integrase/recombinase |
| Rv1054 | - | integrase-a |
| Rv1055 | - | integrase-b |
| Rv1573 | - | phiRV1 phage related protein |
| Rv1574 | - | phiRV1 phage related protein |
| Rv1575 | - | phiRV1 phage related protein |
| Rv1576c | - | phiRV1 phage related protein |
| Rv1577c | - | phiRV1 possible prohead protease |
| Rv1578c | - | phiRV1 phage related protein |
| Rv1579c | - | phiRV1 phage related protein |
| Rv1580c | - | phiRV1 phage related protein |
| Rv1581c | - | phiRV1 phage related protein |
| Rv1582c | - | phiRV1 phage related protein |
| Rv1583c | - | phiRV1 phage related protein |
| Rv1584c | - | phiRV1 phage related protein |
| Rv1585c | - | phiRV1 phage related protein |
| Rv1586c | - | phiRV1 integrase |
| Rv2309c | - | integrase |
| Rv2310 | - | excisionase |
| Rv2646 | - | phiRV2 integrase |
| Rv2647 | - | phiRV2 phage related protein |
| Rv2650c | - | phiRV2 phage related protein |
| Rv2651c | - | phiRV2 prohead protease |
| Rv2652c | - | phiRV2 phage related protein |
| Rv2653c | - | phiRV2 phage related protein |
| Rv2654c | - | phiRV2 phage related protein |
| Rv2655c | - | phiRV2 phage related protein |
| Rv2656c | - | phiRV2 phage related protein |
| Rv2657c | - | similar to gp36 of mycobacteriophage L5 |
| Rv2658c | - | phiRV2 phage related protein |
| Rv2659c | - | phiRV2 integrase |
| Rv2830c | - | similar to phage P1 *phd* gene |
| Rv3750c | - | excisionase |
| Rv3751 | - | putative integrase |

*C. PE and PPE families*

| | |
|---|---|
| 1. PE family | |
| PE subfamily | 38 members |
| PE_PGRS subfamily | 61 members |
| 2. PPE family | 68 members |

| | | |
|---|---|---|
| *D. Antibiotic production and resistance* | | |
| Rv2068c | *blaC* | class A β-lactamase |
| Rv3290c | *lat* | lysine-ε aminotransferase |
| Rv2043c | *pncA* | pyrazinamide resistance/sensitivity |
| Rv0133 | - | possible puromycin N-acetyltransferase |
| Rv0262c | - | aminoglycoside 2'-N-acetyltransferase |
| Rv0802c | - | acetyltransferase |
| Rv1082 | - | similar to *S. lincolnensis lmbE* |
| Rv1170 | - | similar to *S. lincolnensis lmbE* |
| Rv1347c | - | possible aminoglycoside 6'-N-acetyltransferase |
| Rv2036 | - | similar to lincomycin production genes |
| Rv2303c | - | similar to *S. griseus* macrotetrolide resistance protein |
| Rv3225c | - | probable aminoglycoside 3'-phosphotransferases |
| Rv3700c | - | probable acetyltransferase |
| Rv3817 | - | probable aminoglycoside 3'-phosphotransferase |

| | |
|---|---|
| *E. Bacteriocin-like proteins* | 3 |
| *F. Cytochrome P450 enzymes* | 22 |
| *G. Coenzyme F420-dependent enzymes* | 3 |
| *H. Miscellaneous transferases* | 61 |
| *I. Miscellaneous phosphatases, lyases, and hydrolases* | 18 |
| *J. Cyclases* | 6 |
| *K. Chelatases* | 2 |
| *V. Conserved hypotheticals* | 912 |
| *VI. Unknowns* | 606 |
| TOTAL | 3924 |

# Reconciling the spectrum of Sagittarius A* with a two-temperature plasma model

**Rohan Mahadevan**

*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

**The radio source Sagittarius A* is thought to be powered by gas accreting onto a supermassive black hole at the centre of our Galaxy[1,2]. Using the high infrared accretion rates[3], however, standard accretion models[4] are unable to explain the observed low luminosity and spectral energy distribution[5–8], which has led to the consideration of a new model: advection-dominated accretion flows[9–12]. In an advection-dominated flow, most of the accretion energy is stored as thermal energy in the gas which is then lost as the gas falls into the black hole. This model requires the protons to have a much higher temperature than the electrons, and the gas therefore has a two-temperature structure[10,13,14]. Although this model explains the low total luminosity[15–18] and much of the spectral energy distribution (from millimetre wavelengths to hard X-rays), it has been difficult to reconcile with low-frequency radio observations. Here we show that a neglected emission process associated with the protons naturally explains the radio observations without any 'fine tuning' of the model parameters. This result simultaneously supports the two-temperature model of the gas and suggests that an advection-dominated accretion flow onto a black hole of $2.5 \times 10^6$ solar masses provides an accurate description of Sagittarius A*.**

Figure 1 shows the most up-to-date observations of the Galactic Centre[18]. The spectrum rises at radio and submillimetre frequencies $\nu \approx 10^9$–$10^{12}$ Hz, where most of the emission occurs, and has a sharp drop in the infrared. The X-ray observations consist of a possible detection at soft X-ray energies, and firm upper limits in the hard X-rays. The X-ray error-box corresponds to uncertainties in the observed photon index which lies between 1.0 and 2.0 (ref. 18). At very high energies, the EGRET satellite has observed $\gamma$-ray emission from the Galactic Centre region[8]. But owing to the low angular resolution of the measurements, $\sim 1°$, the observations should perhaps be considered as upper limits.

The spectrum from a two-temperature advection-dominated accretion flow (ADAF) is determined by the cooling properties of the protons and electrons in the flow. The protons are at virial temperatures at all radii (proton temperature $T_p \approx 10^{12}$ K close to the black hole) and cool by creating neutral pions[19], while the electrons have much lower temperatures ($T_e \approx 10^{9.5}$ K) and cool by various optically thin processes, such as synchrotron, inverse Compton and bremsstrahlung radiation[11,20].

Figure 1a shows the spectrum from the ADAF model of Sgr A* in ref. 18. This spectrum fits the submillimetre to hard X-ray spectrum quite well, but fails to explain the non-uniform radio spectrum. The radio luminosity, $L_\nu$, is well represented by $L_\nu \propto \nu^{0.2}$ up to $\nu \approx 43$ GHz, which subsequently rises to $L_\nu \propto \nu^{0.8}$ for $\nu \gtrsim 86$ GHz (ref. 21). ADAF models of Sgr A* have always been unable to account for this break, and are substantially underluminous at frequencies below ~86 GHz. This poses a serious problem.

The observed excess of radio emission (beyond what the model predicts) has usually been attributed to a weak jet of material that might emerge from the ADAF; jets are known to be strong radio sources. High-resolution radio observations, however, have ruled this out[22–24], which severely constrains any outflow models. In this case, a rather *ad hoc* electron-temperature profile might be needed to account for the excess radio emission[18], which probably does not correspond to physical conditions. More importantly, recent high-resolution measurements constrain the actual size of the emitting region[5,22]. These observations require large brightness temperatures (in excess of $10^{10}$ K) to explain the observed flux at 43 GHz and 86 GHz. In an ADAF, however, the electron temperature is always well below $10^{10}$ K at all radii[11], and therefore cannot account for these high temperatures.

This apparent problem is solved by considering another emission process associated with the protons. In addition to producing neutral pions, energetic proton collisions can also create charged pions, which subsequently decay into positrons and electrons (referred to here as $e^{\pm}$). This had been neglected in earlier work because these particles do not produce significant amounts of $\gamma$-ray

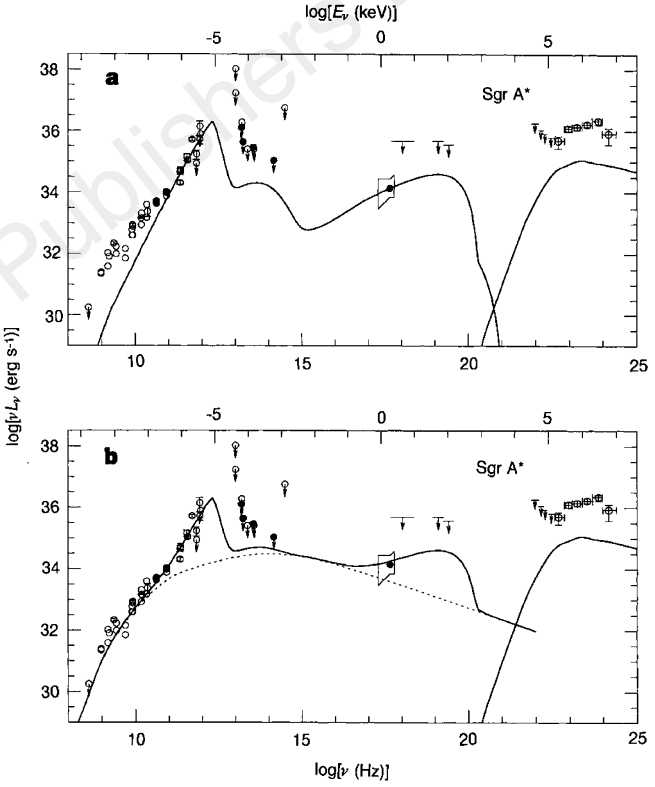



**Figure 1** A comparison of the predicted emission from an ADAF model of the Galactic Centre with the observations. **a**, The spectrum of Sgr A*; the horizontal axis is the logarithm of the frequency and the vertical axis is the logarithm of the energy at that frequency. The measured fluxes were converted to luminosities assuming a distance of 8.5 kpc to the Galactic Centre. The data are the most up-to-date compilation of observations taken from ref. 18. The arrows represent upper limits, and the 'box' at frequency $\sim 10^{17}$ Hz represents the uncertainty in the observed photon index. The solid line is the spectrum from the baseline ADAF model of Sgr A* used in ref. 18. The ADAF parameters are $\alpha = 0.3$, $\beta = 0.5$, $M = 2.5 \times 10^6 M_\odot$, and $\dot{M} = 7.2 \times 10^{-6} M_\odot\,\mathrm{yr}^{-1}$, where $\alpha$ is the viscosity parameter[41], $\beta$ determines the strength of the magnetic field, and is defined so that $(1-\beta)$ is the ratio of magnetic to total pressure, $M$ is the dynamically measured mass of Sgr A*[42,43], and $\dot{M}$ is the mass accretion rate. For frequencies $\lesssim 10^{20}$ Hz, the spectrum is determined by the individual optically thin cooling processes of $\sim 10^{9.5}$ K thermal electrons, while for $\nu \gtrsim 10^{20}$ Hz the spectrum is solely due to the decay of neutral pions. The discrepancy between the model and the observations above $\nu \sim 10^{20}$ Hz is not considered serious, as it is unclear at present whether the ~1° beam of EGRET is detecting a point source or some diffuse emission. These observations should therefore be considered as upper limits rather than detections of a central source. **b**, The solid line represents the total spectrum from the ADAF around Sgr A*, which includes the present results. The parameters used are identical to those in **a**. The dotted line represents only the synchrotron emission from the positrons and electrons.

emission[19].

The high-energy $e^{\pm}$, however, can interact with the magnetic fields in the ADAF to produce synchrotron emission from radio to hard X-ray energies. Because the pions, and therefore the $e^{\pm}$, are created by proton–proton collisions, the energy spectra of the protons and $e^{\pm}$ are related. This allows a direct investigation of the assumption that the protons have a different average temperature from the electrons, and at the same time determines if the $e^{\pm}$ are created in sufficient number, and with the right energy, to produce the observed radio emission.

Here we assume that the energy spectrum of the protons is represented by a power-law distribution, $N(E_p) \propto E_p^{-s}$ with index $s$, where $N(E_p)$ represents the number of protons with energy $E_p$. The index is generally between 2 and 4, and we set it to $s = 2.75$, at the cosmic-ray value, suggesting that a similar acceleration mechanism might be at work in ADAFs[19]. The results are insensitive to the exact value of $s$ (ref. 19).

The rate of production and energy spectrum of the $e^{\pm}$, $R(E)$ is determined by the frequency of proton collisions as well as their energy spectrum. For the assumed power-law proton distribution, the energy distribution of the $e^{\pm}$ is shown in Fig. 2. The spectrum rises at low energies, turns over at $E \approx 35$ MeV, and, as expected, extends as a power-law, $E^{-s}$, with the same energy dependence as the parent proton distribution[25]. Because the created charged pion has a mass of ~140 MeV and decays into four particles, one of which is an electron or positron, we expect that on average the $e^{\pm}$ should carry away one-quarter of the total energy available (that is, $\sim 140/4 = 35$ MeV)[26]. This is an expected turnover which is characteristic of $e^{\pm}$ production, and is shown in Fig. 2.

Determining the synchrotron emissivity from the $e^{\pm}$ requires a knowledge of their steady-state energy distribution $N(E)$. At a given energy $E$, the colliding protons produce $R(E)$ electrons and positrons. However, because the $e^{\pm}$ cool by synchrotron radiation, they lose their energy very efficiently, and the steady-state distribution is therefore determined by the competing effects of the creation and depletion of particles. This requires the net flux of particles between two energies to be equal to their rate of injection, $d[N(E)\dot{E}_S(E)]/dE = R(E)$, where $\dot{E}_S(E)$ is the total synchrotron cooling rate as a function of energy[27].

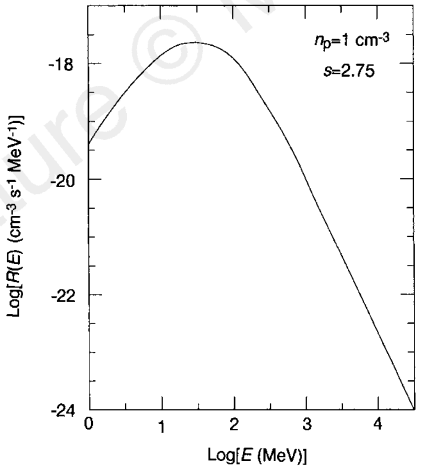



**Figure 2** The energy spectrum, $R(E)$, of positrons and electrons that are created by colliding power-law protons with energy index $s = 2.75$. The vertical axis is the logarithm of number of positron and electrons created per unit volume, per second, per energy interval, and the horizontal axis is the logarithm of the energy. The scale on the vertical axis corresponds to a number density of protons equal to unity. For a number density $N$, the vertical axis must be multiplied by $N^2$. The particles that are responsible for most of the emission are determined by the energy at which the function $E^2R(E)$ peaks, which occurs in the range $100\,\text{MeV} < E < 500\,\text{MeV}$. The shape of the spectrum depends only on the physics of particle collisions and decays[25,26], and at high energies has the spectral shape $R(E) \propto E^{-s}$ (ref. 25). The spectrum therefore contains information about the parent proton distribution, as well as determining the shape of the resulting synchrotron spectrum. It therefore acts as a link between the form of the proton energy distribution and the observed synchrotron spectrum.

Using the steady-state distribution $N(E)$, the $e^{\pm}$ synchrotron spectrum, from the ADAF around Sgr A*, is shown by the dotted line in Fig. 1b. The spectrum rises at low frequencies, turns over, and extends as a power-law at high frequencies. The spectral break at $\nu \approx 10^{15}$ Hz is a direct consequence of the turn over in the $e^{\pm}$ energy spectrum shown in Fig. 2. At high frequencies, the spectrum is optically thin and has a spectral dependence, $L_\nu \propto \nu^{-s/2}$. The spectral slope therefore depends on the proton index $s$, which is a direct consequence of the $e^{\pm}$ having a steady-state distribution $N(E) \propto E^{-(s+1)}$ (ref. 27). At lower frequencies, the expected optically thin spectral dependence is $L_\nu \propto \nu^{-0.5}$ which corresponds to $N(E) \propto E^{-2}$ (ref. 27). However, in an ADAF, the emission at these low frequencies is self-absorbed by the plasma, and the resultant spectrum shown therefore has a different spectral dependence.

The solid curve in Fig. 1b represents the total radiation from the ADAF which includes this spectrum. At high frequencies $\gtrsim 10^{13}$ Hz, the synchrotron emission contributes to, but does not significantly change, the total luminosity. In particular, the agreement with the X-ray flux is not affected, and the additional infrared flux is still well below the stringent upper limits.

At lower energies the result is striking. The emission reproduces the required spectral break at ~86 GHz, is able to account for the 'excess' radio emission below this frequency, and diminishes sufficiently quickly at lower frequencies to agree with the radio upper limit at 400 MHz. As the emission at each radio frequency in Fig. 1a corresponds to a black-body spectrum at a given radius[11,20], the total spectrum shown by the solid line in Fig. 1b indicates that ADAFs produce more emission at a given frequency than the local black-body spectrum. The excess emission is from the high-energy electrons radiating at larger radii. This resolves the problem with the low energy radio emission: no outflow model is needed to account for the observed emission, and the high brightness temperatures inferred[5,22] are easily accounted for by the non-thermal origin of the emission.

The quite good agreement with the radio observations suggests that the emission observed is most probably from the hot protons in the ADAF. But before drawing any conclusions, we examine the essential ingredients required to explain the radio spectrum. Assuming that the dynamics of the flow are determined, reproducing the radio spectrum requires high-energy electrons (or $e^{\pm}$) with energies ~100 MeV at all radii. In an ADAF, this requirement is naturally satisfied. Assuming that viscosity primarily heats the protons into a power-law distribution at all radii, the production of high-energy $e^{\pm}$ with the same energy is completely determined by only the nuclear physics of particle collisions and decays[25,26]. In particular, the shape of the $e^{\pm}$ spectrum (compare Fig. 2) is fixed throughout the flow. We note that the number of $e^{\pm}$ produced is also the right amount; this is a natural consequence of the proton collision time being longer than the accretion time. Whereas shorter collision times would produce excessive amounts of $e^{\pm}$ which would result in too much radio emission, much longer collision times would result in too little radio emission.

The agreement of the theory with the observations depends on two basic assumptions of ADAFs that have always been debated: (1) the existence of a two-temperature plasma, and (2) that viscosity preferentially heats the protons. We have quite good observational evidence that the first assumption is probably true. This is because the radio to hard X-ray spectrum is determined by emission processes associated with both the protons and electrons, at their respective temperatures. If the temperatures of protons and electrons were the same, or were markedly different from their calculated values, the resulting spectrum would be completely different and fail to explain any of the observations.

The second assumption is supported by the present results, and can be discussed in terms of $\delta$, which is the fraction of viscous energy that heats the electrons. The baseline model in ref. 18 set $\delta \approx 0.001$, and showed that for $\delta > 0.01$, too much radiation is produced, and the electron spectrum does not agree with the observations. Here we have a radiation mechanism that accounts for the other fraction $(1 - \delta)$ that heats the protons, and have shown that the agreement with the low-energy radio spectrum requires the amount of energy transferred to the electrons to be small. This shows that the average energy of the protons is probably virial.

Although past work has attempted to answer both these questions theoretically[28–33], the results here provide indirect observational evidence that these assumptions are probably valid. Further, theoretical models which reach contrary conclusions are probably based on assumptions that are not valid in ADAF[18,34]. The present results could therefore be used as tools to aid future theoretical work in resolving these complex questions in plasma physics.

The present results have assumed that all the viscous energy is deposited into a power-law proton distribution, which might seem improbable. However, if half the viscous energy were transferred into a power-law distribution, and half into a thermal distribution, the number of $e^{\pm}$ created reduces only by a factor $\sim$2 (ref. 19), and the results presented here do not change significantly. Therefore, although the agreement with the radio flux requires a power-law proton distribution, it does not require all of the viscous energy to be deposited into the power-law protons.

It is interesting that the good agreement with observations comes from a model in which both the viscous hydrodynamics and the radiative processes have been included self-consistently. Previous models that have attempted to explain the observed spectrum have been phenomenological[35–37], or made simplifying assumptions, such as ignoring the angular momentum of the accreting gas[3,38,39], or, as noted previously[18,40], have errors in the synchrotron calculation which renders the resulting spectrum suspect[3,39]. The ADAF models therefore provides us with a unique self-consistent framework which enables accurate prediction of spectra from accreting black holes.

We stress that there is no fine tuning in the present results. Whereas previous work on ADAFs has not included the $e^{\pm}$ synchrotron radiation, the results presented here show that this process is essential to explaining the observed non-uniform radio spectrum. The model used is identical to that presented in ref. 18, and we have simply taken into account an additional physical process and emission mechanism in the two-temperature ADAF. It is remarkably that, using the same parameters as in ref. 18, an emission mechanism associated with the protons is able to naturally reproduce the entire radio spectrum including the observed spectral break at $\sim$86 GHz. The agreement of the theory with the observations encourages us to take the natural explanation, and conclude that Sgr A* is in fact a $2 \times 10^6$ solar-mass black hole that is accreting by way of a two-temperature ADAF. □

Received 23 March; accepted 19 May 1998.


1. Mezger, P. G., Duschl, W. J. & Zylka, R. The Galactic Center: a laboratory for AGN? *Annu. Rev. Astron. Astrophys.* **7,** 289–388 (1996).
2. Genzel, R., Hollenbach, D. & Townes, C. H. The nucleus of our Galaxy. *Rep. Prog. Phys.* **57,** 417–479 (1994).
3. Melia, F. An accreting black hole model for Sagittarius A*. *Astrophys. J.* **387,** L25–L28 (1992).
4. Frank, J., King, A. & Raine, D. *Accretion Power in Astrophysics* (Cambridge Univ. Press, 1992).
5. Rogers, A. E. E. *et al.* Small-scale structure and position of Sagittarius A* from VLBI at 3 millimeter wavelength. *Astrophys. J.* **434,** L59–L62 (1994).
6. Menten, K. M., Reid, M. J., Eckart, A. & Genzel, R. The position of Sagittarius A*: accurate alignment of the radio and infrared reference frames at the Galactic Center. *Astrophys. J.* **475,** L111–L114 (1997).
7. Predehl, P. & Trümper, J. ROSAT observation of the Sgr A region. *Astron. Astrophys.* **290,** L29–L32 (1994).
8. Merck, M. *et al.* Study of the spectral characteristics of unidentified galactic EGRET sources. Are they pulsar-like? *Astron. Astrophys.* Suppl. **120,** 465–469 (1996).
9. Ichimaru, S. Bimodal behavior of accretion disks—Theory and application to Cygnus X-1 transitions. *Astrophys. J.* **214,** 840–855 (1977).
10. Rees, M. J., Begelman, M. C., Blandford, R. D. & Phinney, E. S. Ion supported tori and the origin of radio jets. *Nature* **295,** 17–21 (1982).
11. Narayan, R. & Yi, I. Advection-dominated accretion: underfed black holes and neutron stars. *Astrophys. J.* **452,** 710–735 (1995).
12. Abramowicz, M. A., Chen, X., Kato, S., Lasota, J.-P. & Regev, O. Thermal equilibria of accretion disks. *Astrophys. J.* **438,** L37–L39 (1995).
13. Shapiro, S. L., Lightman, A. P. & Eardley, D. M. A two-temperature accretion disk model for Cygnus X-1—Structure and spectrum. *Astrophys. J.* **204,** 187–199 (1976).
14. Phinney, E. S. in *Plasma Astrophysics* (eds Guyenne, T. D. & Levy, G.) 337–341 (SP-161, ESA, Paris, 1981).
15. Rees, M. J. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 166–176 (Am. Inst. Phys., New York, 1982).
16. Narayan, R., Yi, I. & Mahadevan, R. Explaining the spectrum of Sagittarius A* with a model of an accreting black hole. *Nature* **374,** 623–625 (1995).
17. Manmoto, T., Mineshige, S. & Kusunose, M. Spectrum of optically thin advection-dominated accretion flow around a black hole: application to Sagittarius A*. *Astrophys. J.* **489,** 791–803 (1997).
18. Narayan, R., Mahadevan, R., Grindlay, J. E., Popham, R. G. & Gammie, C. Advection-dominated accretion model of Sagittarius A*: evidence for a black hole at the Galactic Center. *Astrophys. J.* **492,** 554–568 (1998).
19. Mahadevan, R., Narayan, R. & Krolik, J. Gamma-ray emission from advection-dominated accretion flows around black holes: application to the Galactic Center. *Astrophys. J.* **486,** 268–275 (1977).
20. Mahadevan, R. Scaling laws for advection-dominated flows: applications to low-luminosity galactic nuclei. *Astrophys. J.* **477,** 585–601 (1997).
21. Falcke, H. *et al.* The simultaneous spectrum of Sgr A* from λ20 cm to λ1 mm and the nature of the mm-excess. *Astrophys. J.* (in the press).
22. Backer, D. C. *et al.* Upper limit of 3.3 astronomical units to the diameter of the Galactic Center radio source Sgr A*. *Science* **262,** 1414–1416 (1993).
23. Marcaide, J. M. *et al.* Position and morphology of the compact non-thermal radio source at the galactic center. *Astron. Astrophys.* **258,** 295–301 (1992).
24. Alberdi, A. *et al.* VLBA image of Sgr A* at λ = 1.35 cm. Astron. Astrophys. 277, L1–L4 (1993).
25. Ginzburg, V. L. & Syrovatskii, S. I. *The Origin of Cosmic Rays* (Macmillan, New York, 1964).
26. Dermer, C. D. Binary collision rates of relativistic thermal plasmas. II. Spectra. *Astrophys. J.* **307,** 47–59 (1986).
27. Rybicki, G. & Lightman, A. *Radiative Processes in Astrophysics* (Wiley, New York, 1979).
28. Begelman, M. C. & Chiueh, T. Thermal coupling of ions and electrons by collective effects in two-temperature accretion flows. *Astrophys. J.* **332,** 872–890 (1988).
29. Quataert, E. Particle heating by Alfvenic turbulence in hot accretion flows. *Astrophys. J.* (in the press); also as preprint astro-ph/9710127 (1998).
30. Gruzinov, A. Radiative efficiency of collisionless accretion. *Astrophys. J.* (in the press); also as preprint astro-ph/9710132 (1998).
31. Blackman, E. Fermi energization in magnetized astrophysical flows. *Phys. Rev. Lett.* (in the press); also as preprint astro-ph/9710137 (1998).
32. Bisnovatyi-Kogan, G. S. & Lovelace, R. V. E. Influence of ohmic heating on advection-dominated accretion flows. *Astrophys. J.* **486,** L43–L46 (1997).
33. Quataert, E. & Gruzinov, A. Turbulence and particle heating in advection-dominated accretion flows. *Astrophys. J.* (submitted); also as preprint astro-ph/9803112 (1998).
34. Narayan, R., Mahadevan, R. & Quataert, E. Advection dominated accretion around black holes in *The Theory of Black Hole Accretion Discs* (eds Abramovicz, M. A., Bjornsson, G. & Pringle, J. E.) (Cambridge Univ. Press, in the press).
35. Duschl, W. J. & Lesch, H. The spectrum of Sgr A* and its variability. *Astron. Astrophys.* **286,** 431–436 (1994).
36. Falcke, H. in *Unsolved Problems in the Milky Way* (eds Blitz, L. & Teuben, P. J.) 163–170 (IAU Symp. No. 169, Kluwer, Dordrecht, 1996).
37. Beckert, T. & Duschl, W. J. Synchrotron radiation from quasi-monoenergetic electrons. Modeling the spectrum of Sgr A*. *Astron. Astrophys.* **328,** 95–106 (1997).
38. Mastichiadis, A. & Ozernoy, L. M. X-ray and gamma-ray emission of Sagittarius A* as a wind-accreting black hole. *Astron. Astrophys.* **426,** 599–603 (1994).
39. Melia, F. An accreting black hole model for Sagittarius A*. 2: A detailed study. *Astrophys. J.* **426,** 577–585 (1994).
40. Mahadevan, R., Narayan, R. & Yi, I. Harmony in electrons: cyclotron and synchrotron emission by thermal electrons in a magnetic field. *Astrophys. J.* **465,** 327–337 (1996).
41. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24,** 337–355 (1973).
42. Haller, J. *et al.* Stellar kinematics and the black hole in the Galactic Center. *Astrophys. J.* **456,** 194–205 (1996); Erratum. *Astrophys. J.* **468,** 955 (1996).
43. Eckart, A. & Genzel, R. Stellar proper motions in the central 0.1 pc of the Galaxy. *Mon. Not. R. Astron. Soc.* **284,** 576–598 (1997).


**Acknowledgements.** I thank R. Narayan, E. Blackman, A. Fabian, C. Gammie, Z. Haiman, J. Herrnstein, J. Krolik, A. Mody, M. Rees and E. Quataert for discussions and comments.

Correspondence and requests for materials should be addressed to the author (e-mail: rohan@ast.cam.ac.uk).

# Long-lived giant cells detected at the surface of the Sun

**J. G. Beck*, T. L. Duvall Jr† & P. H. Scherrer***

** W. W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, California 94305-4085, USA*

*† Laboratory for Astronomy and Solar Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA*

**Giant convective cells have been predicted[1] to exist in the Sun. Such cells should span the entire zone unstable to convective motions—now known to cover the outer 29 per cent of the Sun's radius[2]—and could be dredging up the magnetic flux that is thought to be the source of solar activity (sunspots). Several studies[3–5] have failed to detect these giant cells, although there**

# Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence

**S. T. Cole*, R. Brosch*, J. Parkhill, T. Garnier*, C. Churcher, D. Harris, S. V. Gordon*, K. Eiglmeier*, S. Gas*, C. E. Barry III†, F. Tekaia‡, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh§, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead & B. G. Barrell**

*Sanger Centre, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK*
** Unité de Génétique Moléculaire Bactérienne, and ‡ Unité de Génétique Moléculaire des Levures, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France*
*† Tuberculosis Research Unit, Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840, USA*
*§ Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark*

**Countless millions of people have died from tuberculosis, a chronic infectious disease caused by the tubercle bacillus. The complete genome sequence of the best-characterized strain of *Mycobacterium tuberculosis*, H37Rv, has been determined and analysed in order to improve our understanding of the biology of this slow-growing pathogen and to help the conception of new prophylactic and therapeutic interventions. The genome comprises 4,411,529 base pairs, contains around 4,000 genes, and has a very high guanine + cytosine content that is reflected in the biased amino-acid content of the proteins. *M. tuberculosis* differs radically from other bacteria in that a very large portion of its coding capacity is devoted to the production of enzymes involved in lipogenesis and lipolysis, and to two new families of glycine-rich proteins with a repetitive structure that may represent a source of antigenic variation.**

Despite the availability of effective short-course chemotherapy (DOTS) and the Bacille Calmette-Guérin (BCG) vaccine, the tubercle bacillus continues to claim more lives than any other single infectious agent[1]. Recent years have seen increased incidence of tuberculosis in both developing and industrialized countries, the widespread emergence of drug-resistant strains and a deadly synergy with the human immunodeficiency virus (HIV). In 1993, the gravity of the situation led the World Health Organisation (WHO) to declare tuberculosis a global emergency in an attempt to heighten public and political awareness. Radical measures are needed now to prevent the grim predictions of the WHO becoming reality. The combination of genomics and bioinformatics has the potential to generate the information and knowledge that will enable the conception and development of new therapies and interventions needed to treat this airborne disease and to elucidate the unusual biology of its aetiological agent, *Mycobacterium tuberculosis*.

The characteristic features of the tubercle bacillus include its slow growth, dormancy, complex cell envelope, intracellular pathogenesis and genetic homogeneity[2]. The generation time of *M. tuberculosis*, in synthetic medium or infected animals, is typically ~24 hours. This contributes to the chronic nature of the disease, imposes lengthy treatment regimens and represents a formidable obstacle for researchers. The state of dormancy in which the bacillus remains quiescent within infected tissue may reflect metabolic shutdown resulting from the action of a cell-mediated immune response that can contain but not eradicate the infection. As immunity wanes, through ageing or immune suppression, the dormant bacteria reactivate, causing an outbreak of disease often many decades after the initial infection[3]. The molecular basis of dormancy and reactivation remains obscure but is expected to be genetically programmed and to involve intracellular signalling pathways.

The cell envelope of *M. tuberculosis*, a Gram-positive bacterium with a G + C-rich genome, contains an additional layer beyond the peptidoglycan that is exceptionally rich in unusual lipids, glycolipids and polysaccharides[4,5]. Novel biosynthetic pathways generate cell-wall components such as mycolic acids, mycocerosic acid, phenolthiocerol, lipoarabinomannan and arabinogalactan, and several of these may contribute to mycobacterial longevity, trigger inflammatory host reactions and act in pathogenesis. Little is known about the mechanisms involved in life within the macrophage, or the extent and nature of the virulence factors produced by the bacillus and their contribution to disease.

It is thought that the progenitor of the *M. tuberculosis* complex, comprising *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, *M. africanum* and *M. microti*, arose from a soil bacterium and that the human bacillus may have been derived from the bovine form following the domestication of cattle. The complex lacks interstrain genetic diversity, and nucleotide changes are very rare[6]. This is important in terms of immunity and vaccine development as most of the proteins will be identical in all strains and therefore antigenic drift will be restricted. On the basis of the systematic sequence analysis of 26 loci in a large number of independent isolates[6], it was concluded that the genome of *M. tuberculosis* is either unusually inert or that the organism is relatively young in evolutionary terms.

Since its isolation in 1905, the H37Rv strain of *M. tuberculosis* has found extensive, worldwide application in biomedical research because it has retained full virulence in animal models of tuberculosis, unlike some clinical isolates; it is also susceptible to drugs and amenable to genetic manipulation. An integrated map of the 4.4 megabase (Mb) circular chromosome of this slow-growing pathogen had been established previously and ordered libraries of cosmids and bacterial artificial chromosomes (BACs) were available[7,8].

## Organization and sequence of the genome

**Sequence analysis.** To obtain the contiguous genome sequence, a combined approach was used that involved the systematic sequence analysis of selected large-insert clones (cosmids and BACs) as well as

random small-insert clones from a whole-genome shotgun library. This culminated in a composite sequence of 4,411,529 base pairs (bp) (Figs 1, 2), with a G + C content of 65.6%. This represents the second-largest bacterial genome sequence currently available (after that of *Escherichia coli*)[9]. The initiation codon for the *dnaA* gene, a hallmark for the origin of replication, *oriC*, was chosen as the start point for numbering. The genome is rich in repetitive DNA, particularly insertion sequences, and in new multigene families and duplicated housekeeping genes. The G + C content is relatively constant throughout the genome (Fig. 1) indicating that horizontally transferred pathogenicity islands of atypical base composition are probably absent. Several regions showing higher than average G + C content (Fig. 1) were detected; these correspond to sequences belonging to a large gene family that includes the polymorphic G + C-rich sequences (PGRSs).

**Genes for stable RNA.** Fifty genes coding for functional RNA molecules were found. These molecules were the three species produced by the unique ribosomal RNA operon, the 10Sa RNA involved in degradation of proteins encoded by abnormal messenger RNA, the RNA component of RNase P, and 45 transfer RNAs. No 4.5S RNA could be detected. The *rrn* operon is situated unusually as it occurs about 1,500 kilobases (kb) from the putative *oriC*; most eubacteria have one or more *rrn* operons near to *oriC* to exploit the gene-dosage effect obtained during replication[10]. This arrangement may be related to the slow growth of *M. tuberculosis*. The genes encoding tRNAs that recognize 43 of the 61 possible sense codons were distributed throughout the genome and, with one exception, none of these uses A in the first position of the anticodon, indicating that extensive wobble occurs during translation. This is consistent with the high G + C content of the genome and the consequent bias in codon usage. Three genes encoding tRNAs for methionine were found; one of these genes (*metV*) is situated in a region that may correspond to the terminus of replication (Figs 1, 2). As *metV* is linked to defective genes for integrase and excisionase, perhaps it was once part of a phage or similar mobile genetic element.

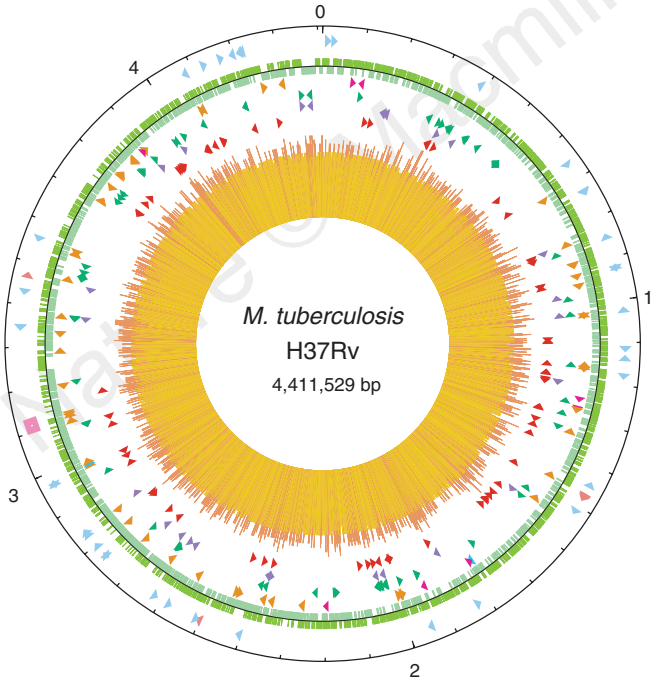



**Figure 1** Circular map of the chromosome of *M. tuberculosis* H37Rv. The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes (tRNAs are blue, others are pink) and the direct repeat region (pink cube); the second ring inwards shows the coding sequence by strand (clockwise, dark green; anticlockwise, light green); the third ring depicts repetitive DNA (insertion sequences, orange; 13E12 REP family, dark pink; prophage, blue); the fourth ring shows the positions of the PPE family members (green); the fifth ring shows the PE family members (purple, excluding PGRS); and the sixth ring shows the positions of the PGRS sequences (dark red). The histogram (centre) represents G + C content, with <65% G + C in yellow, and >65% G + C in red. The figure was generated with software from DNASTAR.

**Insertion sequences and prophages.** Sixteen copies of the promiscuous insertion sequence IS*6110* and six copies of the more stable element IS*1081* reside within the genome of H37Rv[8]. One copy of IS*1081* is truncated. Scrutiny of the genomic sequence led to the identification of a further 32 different insertion sequence elements, most of which have not been described previously, and of the 13E12 family of repetitive sequences which exhibit some of the characteristics of mobile genetic elements (Fig. 1). The newly discovered insertion sequences belong mainly to the IS*3* and IS*256* families, although six of them define a new group. There is extensive similarity between IS*1561* and IS*1552* with insertion sequence elements found in *Nocardia* and *Rhodococcus* spp., suggesting that they may be widely disseminated among the actinomycetes.

Most of the insertion sequences in *M. tuberculosis* H37Rv appear to have inserted in intergenic or non-coding regions, often near tRNA genes (Fig. 1). Many are clustered, suggesting the existence of insertional hot-spots that prevent genes from being inactivated, as has been described for *Rhizobium*[11]. The chromosomal distribution of the insertion sequences is informative as there appears to have been a selection against insertions in the quadrant encompassing *oriC* and an overrepresentation in the direct repeat region that contains the prototype IS*6110*. This bias was also observed experimentally in a transposon mutagenesis study[12].

At least two prophages have been detected in the genome sequence and their presence may explain why *M. tuberculosis* shows persistent low-level lysis in culture. Prophages phiRv1 and phiRv2 are both ~10 kb in length and are similarly organized, and some of their gene products show marked similarity to those encoded by certain bacteriophages from *Streptomyces* and saprophytic mycobacteria. The site of insertion of phiRv1 is intriguing as it corresponds to part of a repetitive sequence of the 13E12 family that itself appears to have integrated into the biotin operon. Some strains of *M. tuberculosis* have been described as requiring biotin as a growth supplement, indicating either that phiRv1 has a polar effect on expression of the distal *bio* genes or that aberrant excision, leading to mutation, may occur. During the serial attenuation of *M. bovis* that led to the vaccine strain *M. bovis* BCG, the phiRv1 prophage was lost[13]. In a systematic study of the genomic diversity of prophages and insertion sequences (S.V.G. *et al.*, manuscript in preparation), only IS*1532* exhibited significant variability, indicating that most of the prophages and insertion sequences are currently stable. However, from these combined observations, one can conclude that horizontal transfer of genetic material into the free-living ancestor of the *M. tuberculosis* complex probably occurred in nature before the tubercle bacillus adopted its specialized intracellular niche.

**Figure 2** Linear map of the chromosome of *M. tuberculosis* H37Rv showing the position and orientation of known genes and coding sequences (CDS). We used the following functional categories (adapted from ref. 20): lipid metabolism (black); intermediary metabolism and respiration (yellow); information pathways (pink); regulatory proteins (sky blue); conserved hypothetical proteins (orange); proteins of unknown function (light green); insertion sequences and phage-related functions (blue); stable RNAs (purple); cell wall and cell processes (dark green); PE and PPE protein families (magenta); virulence, detoxification and adaptation (white). For additional information about gene functions, refer to http://www.sanger.ac.uk. ▶

**Genes encoding proteins.** 3,924 open reading frames were identified in the genome (see Methods), accounting for ~91% of the potential coding capacity (Figs 1, 2). A few of these genes appear to have in-frame stop codons or frameshift mutations (irrespective of the source of the DNA sequenced) and may either use frameshifting during translation or correspond to pseudogenes. Consistent with the high G + C content of the genome, GTG initiation codons (35%) are used more frequently than in *Bacillus subtilis* (9%) and *E. coli* (14%), although ATG (61%) is the most common translational start. There are a few examples of atypical initiation codons, the most notable being the ATC used by *infC*, which begins with ATT in both *B. subtilis* and *E. coli*[9,14]. There is a slight bias in the orientation of the genes (Fig. 1) with respect to the direction of replication as ~59% are transcribed with the same polarity as replication, compared with 75% in *B. subtilis*. In other bacteria, genes transcribed in the same direction as the replication forks are believed to be expressed more efficiently[9,14]. Again, the more even distribution in gene polarity seen in *M. tuberculosis* may reflect the slow growth and infrequent replication cycles. Three genes (*dnaB*, *recA* and Rv1461) have been invaded by sequences encoding inteins (protein introns) and in all three cases their counterparts in *M. leprae* also contain inteins, but at different sites[15] (S.T.C. *et al.*, unpublished observations).

**Protein function, composition and duplication.** By using various database comparisons, we attributed precise functions to ~40% of the predicted proteins and found some information or similarity for another 44%. The remaining 16% resembled no known proteins and may account for specific mycobacterial functions. Examination of the amino-acid composition of the *M. tuberculosis* proteome by correspondence analysis[16], and comparison with that of other microorganisms whose genome sequences are available, revealed a statistically significant preference for the amino acids Ala, Gly, Pro, Arg and Trp, which are all encoded by G + C-rich codons, and a comparative reduction in the use of amino acids encoded by A + T-rich codons such as Asn, Ile, Lys, Phe and Tyr (Fig. 3). This approach also identified two groups of proteins rich in Asn or Gly that belong to new families, PE and PPE (see below). The fraction of the proteome that has arisen through gene duplication is similar to that seen in *E. coli* or *B. subtilis* (~51%; refs 9, 14), except that the level of sequence conservation is considerably higher, indicating that there may be extensive redundancy or differential production of the corresponding polypeptides. The apparent lack of divergence following gene duplication is consistent with the hypothesis that *M. tuberculosis* is of recent descent[6].

## General metabolism, regulation and drug resistance

**Metabolic pathways.** From the genome sequence, it is clear that the tubercle bacillus has the potential to synthesize all the essential amino acids, vitamins and enzyme co-factors, although some of the pathways involved may differ from those found in other bacteria. *M. tuberculosis* can metabolize a variety of carbohydrates, hydrocarbons, alcohols, ketones and carboxylic acids[2,17]. It is apparent from genome inspection that, in addition to many functions involved in lipid metabolism, the enzymes necessary for glycolysis, the pentose phosphate pathway, and the tricarboxylic acid and glyoxylate cycles are all present. A large number (~200) of oxidoreductases, oxygenases and dehydrogenases is predicted, as well as many oxygenases containing cytochrome P450, that are similar to fungal proteins involved in sterol degradation. Under aerobic growth conditions, ATP will be generated by oxidative phosphorylation from electron transport chains involving a ubiquinone cytochrome *b* reductase complex and cytochrome *c* oxidase. Components of several anaerobic phosphorylative electron transport chains are also present, including genes for nitrate reductase (*narGHJI*), fumarate reductase (*frdABCD*) and possibly nitrite reductase (*nirBD*), as well as a new reductase (*narX*) that results from a rearrangement of a homologue of the *narGHJI* operon. Two genes encoding haemoglobin-like proteins, which may protect against oxidative stress or be involved in oxygen capture, were found. The ability of the bacillus to adapt its metabolism to environmental change is significant as it not only has to compete with the lung for oxygen but must also adapt to the microaerophilic/anaerobic environment at the heart of the burgeoning granuloma.

**Regulation and signal transduction.** Given the complexity of the environmental and metabolic choices facing *M. tuberculosis*, an extensive regulatory repertoire was expected. Thirteen putative sigma factors govern gene expression at the level of transcription initiation, and more than 100 regulatory proteins are predicted (Table 1). Unlike *B. subtilis* and *E. coli*, in which there are >30 copies of different two-component regulatory systems[14], *M. tuberculosis* has only 11 complete pairs of sensor histidine kinases and response regulators, and a few isolated kinase and regulatory genes. This relative paucity in environmental signal transduction pathways is probably offset by the presence of a family of eukaryotic-like serine/threonine protein kinases (STPKs), which function as part of a phosphorelay system[18]. The STPKs probably have two domains: the well-conserved kinase domain at the amino terminus is predicted to be connected by a transmembrane segment to the carboxy-terminal region that may respond to specific stimuli. Several of the predicted envelope lipoproteins, such as that encoded by *lppR* (Rv2403), show extensive similarity to this putative receptor domain of STPKs, suggesting possible interplay. The STPKs probably function in signal transduction pathways and may govern important cellular decisions such as dormancy and cell division, and although their partners are unknown, candidate genes for phosphoprotein phosphatases have been identified.

**Drug resistance.** *M. tuberculosis* is naturally resistant to many antibiotics, making treatment difficult[19]. This resistance is due mainly to the highly hydrophobic cell envelope acting as a permeability barrier[4], but many potential resistance determinants are also encoded in the genome. These include hydrolytic or drug-modifying enzymes such as β-lactamases and aminoglycoside acetyl transferases, and many potential drug–efflux systems, such as 14 members of the major facilitator family and numerous ABC transporters. Knowledge of these putative resistance mechanisms will promote better use of existing drugs and facilitate the conception of new therapies.

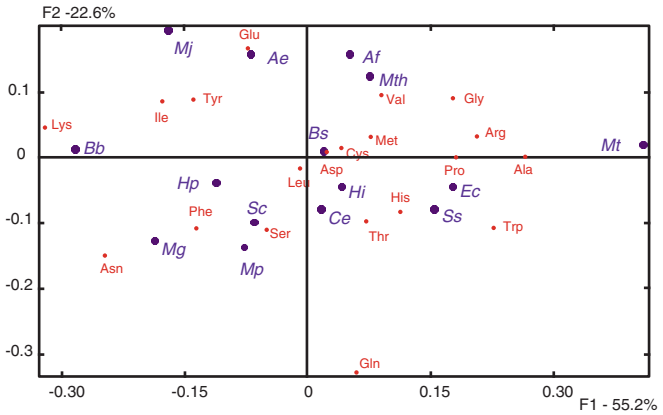
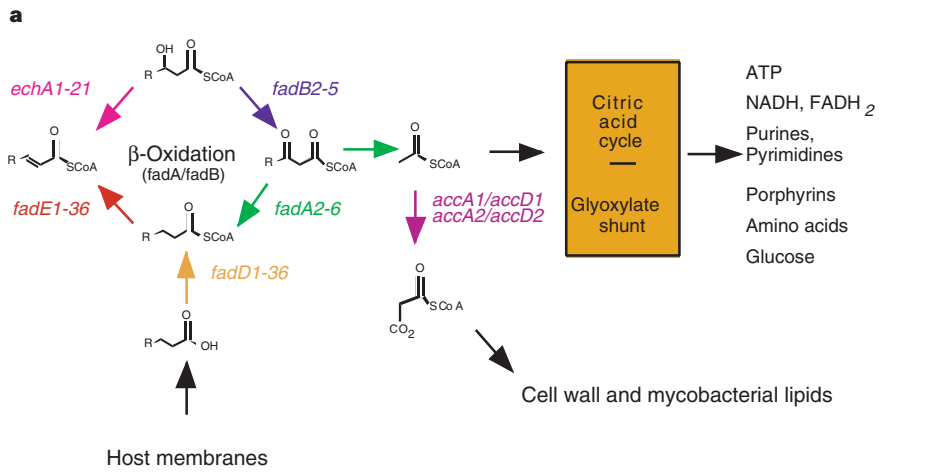
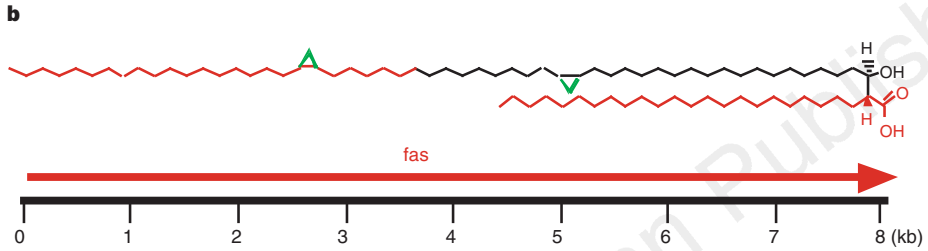
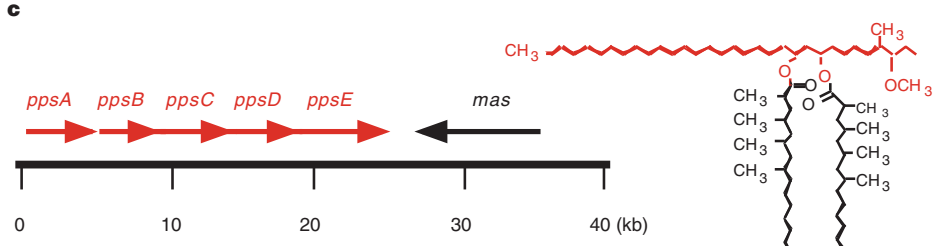



**Figure 3** Correspondence analysis of the proteomes from extensively sequenced organisms as a function of amino-acid composition. Note the extreme position of *M. tuberculosis* and the shift in amino-acid preference reflecting increasing G + C content from left to right. Abbreviations used: *Ae*, *Aquifex aeolicus*; *Af*, *Archaeoglobus fulgidis*; *Bb*, *Borrelia burgdorfei*; *Bs*, *B. subtilis*; *Ce*, *Caenorhabditis elegans*; *Ec*, *E. coli*; *Hi*, *Haemophilus influenzae*; *Hp*, *Helicobacter pylori*; *Mg*, *Mycoplasma genitalium*; *Mj*, *Methanococcus jannaschi*; *Mp*, *Mycoplasma pneumoniae*; *Mt*, *M. tuberculosis*; *Mth*, *Methanobacterium thermoautotrophicum*; *Sc*, *Saccharomyces cerevisiae*; *Ss*, *Synechocystis* sp. strain PCC6803. F1 and F2, first and second factorial axes[16].




| Gene | Function |
|---|---|
| *fas* | Fatty acid synthase, produces $C_{16}$-$C_{26}$ acyl-CoA esters |
| *fabD* | Malonyl-CoA:AcpM acyltransferase, AcpM loading |
| *acpM* | Acyl carrier protein, meromycolate precursor transport |
| *kasA* | Ketoacyl acyl carrier protein synthase, chain elongation |
| *kasB* | Ketoacyl acyl carrier protein synthase, chain elongation |
| *accD* | Acetyl-CoA carboxylase, malonyl-CoA synthesis |
| *mabA* | 3-oxo-acyl-acyl carrier protein reductase, reduces KasA/B product |
| *inhA* | Enoyl-acyl carrier protein reductase |
| *cmaA1* | cyclopropane mycolic acid synthase 1, distal-position specific |
| *cmaA2* | cyclopropane mycolic acid synthase 2, proximal-position specific |




| Gene | Function |
|---|---|
| *mas* | Four rounds extension of $C_{16,18}$ using Me-malonyl CoA |
| *ppsA* | Extension of $C_{18}$ with malony CoA, partial reduction |
| *ppsB* | Extension with malony CoA, partial reduction |
| *ppsC* | Extension with malony CoA, complete reduction |
| *ppsD* | Extension with Me-malony CoA, partial reduction |
| *ppsE* | Extension with malony CoA, partial reduction, decarboxylation |

**Figure 4** Lipid metabolism. **a**, Degradation of host-cell lipids is vital in the intracellular life of *M. tuberculosis*. Host-cell membranes provide precursors for many metabolic processes, as well as potential precursors of mycobacterial cell-wall constituents, through the actions of a broad family of β-oxidative enzymes encoded by multiple copies in the genome. These enzymes produce acetyl CoA, which can be converted into many different metabolites and fuel for the bacteria through the actions of the enzymes of the citric acid cycle and the glyoxylate shunt of this cycle. **b**, The genes that synthesize mycolic acids, the dominant lipid component of the mycobacterial cell wall, include the type I fatty acid synthase (*fas*) and a unique type II system which relies on extension of a precursor bound to an acyl carrier protein to form full-length (~80-carbon) mycolic acids. The *cma* genes are responsible for cyclopropanation. **c**, The genes that produce phthiocerol dimycocerosate form a large operon and represent type I (*mas*) and type II (the *pps* operon) polyketide synthase systems. Functions are colour coordinated.

## Lipid metabolism

Very few organisms produce such a diverse array of lipophilic molecules as *M. tuberculosis*. These molecules range from simple fatty acids such as palmitate and tuberculostearate, through isoprenoids, to very-long-chain, highly complex molecules such as mycolic acids and the phenolphthiocerol alcohols that esterify with mycocerosic acid to form the scaffold for attachment of the mycosides. Mycobacteria contain examples of every known lipid and polyketide biosynthetic system, including enzymes usually found in mammals and plants as well as the common bacterial systems. The biosynthetic capacity is overshadowed by the even more remarkable radiation of degradative, fatty acid oxidation systems and, in total, there are ~250 distinct enzymes involved in fatty acid metabolism in *M. tuberculosis* compared with only 50 in *E. coli*[20].

**Fatty acid degradation.** *In vivo*-grown mycobacteria have been suggested to be largely lipolytic, rather than lipogenic, because of the variety and quantity of lipids available within mammalian cells and the tubercle[2] (Fig. 4a). The abundance of genes encoding components of fatty acid oxidation systems found by our genomic approach supports this proposition, as there are 36 acyl-CoA synthases and a family of 36 related enzymes that could catalyse the first step in fatty acid degradation. There are 21 homologous enzymes belonging to the enoyl-CoA hydratase/isomerase superfamily of enzymes, which rehydrate the nascent product of the acyl-CoA dehydrogenase. The four enzymes that convert the 3-hydroxy fatty acid into a 3-keto fatty acid appear less numerous, mainly because they are difficult to distinguish from other members of the short-chain alcohol dehydrogenase family on the basis of primary sequence. The five enzymes that complete the cycle by thiolysis of the β-ketoester, the acetyl-CoA C-acetyltransferases, do indeed appear to be a more limited family. In addition to this extensive set of dissociated degradative enzymes, the genome also encodes the canonical FadA/FadB β-oxidation complex (Rv0859 and Rv0860). Accessory activities are present for the metabolism of odd-chain and multiply unsaturated fatty acids.

**Fatty acid biosynthesis.** At least two discrete types of enzyme system, fatty acid synthase (FAS) I and FAS II, are involved in fatty acid biosynthesis in mycobacteria (Fig. 4b). FAS I (Rv2524, *fas*) is a single polypeptide with multiple catalytic activities that generates several shorter CoA esters from acetyl-CoA primers[5] and probably creates precursors for elongation by all of the other fatty acid and polyketide systems. FAS II consists of dissociable enzyme components which act on a substrate bound to an acyl-carrier protein (ACP). FAS II is incapable of *de novo* fatty acid synthesis but instead elongates palmitoyl-ACP to fatty acids ranging from 24 to 56 carbons in length[17,21]. Several different components of FAS II may be targets for the important tuberculosis drug isoniazid, including the enoyl-ACP reductase InhA[22], the ketoacyl-ACP synthase KasA and the ACP AcpM[21]. Analysis of the genome shows that there are only three potential ketoacyl synthases: KasA and KasB are highly related, and their genes cluster with *acpM*, whereas KasC is a more distant homologue of a ketoacyl synthase III system. The number of ketoacyl synthase and ACP genes indicates that there is a single FAS II system. Its genetic organization, with two clustered ketoacyl synthases, resembles that of type II aromatic polyketide biosynthetic gene clusters, such as those for actinorhodin, tetracycline and tetracenomycin in *Streptomyces* species[23]. InhA seems to be the sole enoyl-ACP reductase and its gene is co-transcribed with a *fabG* homologue, which encodes 3-oxoacyl-ACP reductase. Both of these proteins are probably important in the biosynthesis of mycolic acids.

Fatty acids are synthesized from malonyl-CoA and precursors are generated by the enzymatic carboxylation of acetyl (or propionyl)-CoA by a biotin-dependent carboxylase (Fig. 4b). From study of the genome we predict that there are three complete carboxylase systems, each consisting of an α- and a β-subunit, as well as three β-subunits without an α-counterpart. As a group, all of the carboxylases seem to be more related to the mammalian homologues than to the corresponding bacterial enzymes. Two of these carboxylase systems (*accA1*, *accD1* and *accA2*, *accD2*) are probably involved in degradation of odd-numbered fatty acids, as they are adjacent to genes for other known degradative enzymes. They may convert propionyl-CoA to succinyl-CoA, which can then be incorporated into the tricarboxylic acid cycle. The synthetic carboxylases (*accA3*, *accD3*, *accD4*, *accD5* and *accD6*) are more difficult to understand. The three extra β-subunits might direct carboxylation to the appropriate precursor or may simply increase the total amount of carboxylated precursor available if this step were rate-limiting.

Synthesis of the paraffinic backbone of fatty and mycolic acids in the cell is followed by extensive postsynthetic modifications and unsaturations, particularly in the case of the mycolic acids[24,25]. Unsaturation is catalysed either by a FabA-like β-hydroxyacyl-ACP dehydrase, acting with a specific ketoacyl synthase, or by an aerobic terminal mixed function desaturase that uses both molecular oxygen and NADPH. Inspection of the genome revealed no obvious candidates for the FabA-like activity. However, three potential aerobic desaturases (encoded by *desA1*, *desA2* and *desA3*) were evident that show little similarity to related vertebrate or yeast enzymes (which act on CoA esters) but instead resemble plant desaturases (which use ACP esters). Consequently, the genomic data indicate that unsaturation of the meromycolate chain may occur while the acyl group is bound to AcpM.

Much of the subsequent structural diversity in mycolic acids is

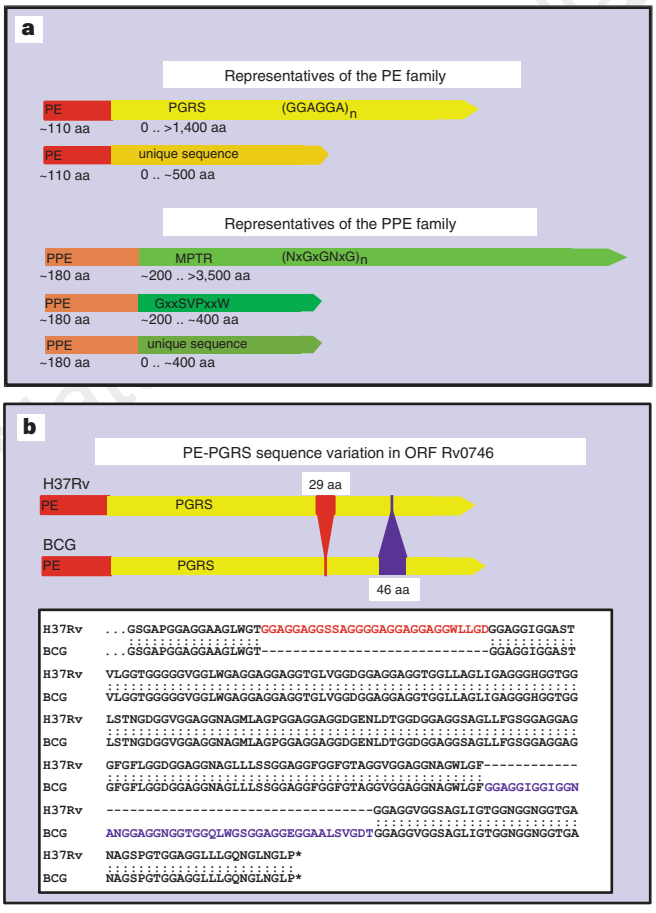



**Figure 5** The PE and PPE protein families. **a**, Classification of the PE and PPE protein families. **b**, Sequence variation between *M. tuberculosis* H37Rv and *M. bovis* BCG-Pasteur in the PE-PGRS encoded by open reading frame (ORF) Rv0746.

generated by a family of *S*-adenosyl-L-methionine-dependent enzymes, which use the unsaturated meromycolic acid as a substrate to generate *cis* and *trans* cyclopropanes and other mycolates. Six members of this family have been identified and characterized[25] and two clustered, convergently transcribed new genes are evident in the genome (*umaA1* and *umaA2*). From the functions of the known family members and the structures of mycolic acids in *M. tuberculosis*, it is tempting to speculate that these new enzymes may introduce the *trans* cyclopropanes into the meromycolate precursor. In addition to these two methyltransferases, there are two other unrelated lipid methyltransferases (Ufa1 and Ufa2) that share homology with cyclopropane fatty acid synthase of *E. coli*[25]. Although cyclopropanation seems to be a relatively common modification of mycolic acids, cyclopropanation of plasma-membrane constituents has not been described in mycobacteria. Tuberculostearic acid is produced by methylation of oleic acid, and may be synthesized by one of these two enzymes.

Condensation of the fully functionalized and preformed meromycolate chain with a 26-carbon α-branch generates full-length mycolic acids that must be transported to their final location for attachment to the cell-wall arabinogalactan. The transfer and subsequent transesterification is mediated by three well-known immunogenic proteins of the antigen 85 complex[26]. The genome encodes a fourth member of this complex, antigen 85C′ (*fbpC2*, Rv0129), which is highly related to antigen 85C. Further studies are needed to show whether the protein possesses mycolytransferase activity and to clarify the reason behind the apparent redundancy.

**Polyketide synthesis.** Mycobacteria synthesize polyketides by several different mechanisms. A modular type I system, similar to that involved in erythromycin biosynthesis[23], is encoded by a very large operon, *ppsABCDE*, and functions in the production of phenolphthiocerol[5]. The absence of a second type I polyketide synthase suggests that the related lipids phthiocerol A and B, phthiodiolone A and phthiotriol may all be synthesized by the same system, either from alternative primers or by differential postsynthetic modification. It is physiologically significant that the *pps* gene cluster occurs immediately upstream of *mas*, which encodes the multifunctional enzyme mycocerosic acid synthase (MAS), as their products phthiocerol and mycocerosic acid esterify to form the very abundant cell-wall-associated molecule phthiocerol dimycocerosate (Fig. 4c).

Members of another large group of polyketide synthase enzymes are similar to MAS, which also generates the multiply methyl-branched fatty acid components of mycosides and phthiocerol dimycocerosate, abundant cell-wall-associated molecules[5]. Although some of these polyketide synthases may extend type I FAS CoA primers to produce other long-chain methyl-branched fatty acids such as mycolipenic, mycolipodienic and mycolipanolic acids or the phthioceranic and hydroxyphthioceranic acids, or may even show functional overlap[5], there are many more of these enzymes than there are known metabolites. Thus there may be new lipid and polyketide metabolites that are expressed only under certain conditions, such as during infection and disease.

A fourth class of polyketide synthases is related to the plant enzyme superfamily that includes chalcone and stilbene synthase[23]. These polyketide synthases are phylogenetically divergent from all other polyketide and fatty acid synthases and generate unreduced polyketides that are typically associated with anthocyanin pigments and flavonoids. The function of these systems, which are often linked to apparent type I modules, is unknown. An example is the gene cluster spanning *pks10*, *pks7*, *pks8* and *pks9*, which includes two of the chalcone-synthase-like enzymes and two modules of an apparent type I system. The unknown metabolites produced by these enzymes are interesting because of the potent biological activities of some polyketides such as the immunosuppressor rapamycin.

**Siderophores.** Peptides that are not ribosomally synthesized are made by a process that is mechanistically analogous to polyketide synthesis[23,27]. These peptides include the structurally related iron-scavenging siderophores, the mycobactins and the exochelins[2,28], which are derived from salicylate by the addition of serine (or threonine), two lysines and various fatty acids and possible polyketide segments. The *mbt* operon, encoding one apparent salicylate-activating protein, three amino-acid ligases, and a single module of a type I polyketide synthase, may be responsible for the biosynthesis of the mycobacterial siderophores. The presence of only one non-ribosomal peptide-synthesis system indicates that this pathway may generate both siderophores and that subsequent modification of a single ε-amino group of one lysine residue may account for the different physical properties and function of the siderophores[28].

## Immunological aspects and pathogenicity

Given the scale of the global tuberculosis burden, vaccination is not only a priority but remains the only realistic public health intervention that is likely to affect both the incidence and the prevalence of the disease[29]. Several areas of vaccine development are promising, including DNA vaccination, use of secreted or surface-exposed proteins as immunogens, recombinant forms of BCG and rational attenuation of *M. tuberculosis*[29]. All of these avenues of research will benefit from the genome sequence as its availability will stimulate more focused approaches. Genes encoding ~90 lipoproteins were identified, some of which are enzymes or components of transport systems, and a similar number of genes encoding preproteins (with type I signal peptides) that are probably exported by the Sec-dependent pathway. *M. tuberculosis* seems to have two copies of *secA*. The potent T-cell antigen Esat-6 (ref. 30), which is probably secreted in a Sec-independent manner, is encoded by a member of a multigene family. Examination of the genetic context reveals several similarly organized operons that include genes encoding large ATP-hydrolysing membrane proteins that might act as transporters. One of the surprises of the genome project was the discovery of two extensive families of novel glycine-rich proteins, which may be of immunological significance as they are predicted to be abundant and potentially polymorphic antigens.

**The PE and PPE multigene families.** About 10% of the coding capacity of the genome is devoted to two large unrelated families of acidic, glycine-rich proteins, the PE and PPE families, whose genes are clustered (Figs 1, 2) and are often based on multiple copies of the polymorphic repetitive sequences referred to as PGRSs, and major polymorphic tandem repeats (MPTRs), respectively[31,32]. The names PE and PPE derive from the motifs Pro–Glu (PE) and Pro–Pro–Glu (PPE) found near the N terminus in most cases[33]. The 99 members of the PE protein family all have a highly conserved N-terminal domain of ~110 amino-acid residues that is predicted to have a globular structure, followed by a C-terminal segment that varies in size, sequence and repeat copy number (Fig. 5). Phylogenetic analysis separated the PE family into several subfamilies. The largest of these is the highly repetitive PGRS class, which contains 61 members; members of the other subfamilies, share very limited sequence similarity in their C-terminal domains (Fig. 5). The predicted molecular weights of the PE proteins vary considerably as a few members contain only the N-terminal domain, whereas most have C-terminal extensions ranging in size from 100 to 1,400 residues. The PGRS proteins have a high glycine content (up to 50%), which is the result of multiple tandem repetitions of Gly–Gly–Ala or Gly–Gly–Asn motifs, or variations thereof.

The 68 members of the PPE protein family (Fig. 5) also have a conserved N-terminal domain that comprises ~180 amino-acid residues, followed by C-terminal segments that vary markedly in sequence and length. These proteins fall into at least three groups, one of which constitutes the MPTR class characterized by the presence of multiple, tandem copies of the motif Asn–X–Gly–X–Gly–Asn–X–Gly. The second subgroup contains a characteristic, well-conserved motif around position 350, whereas the third contains

proteins that are unrelated except for the presence of the common 180-residue PPE domain.

The subcellular location of the PE and PPE proteins is unknown and in only one case, that of a lipase (Rv3097), has a function been demonstrated. On examination of the protein database from the extensively sequenced *M. leprae*[15], no PGRS- or MPTR-related polypeptides were detected but a few proteins belonging to the non-MPTR subgroup of the PPE family were found. These proteins include one of the major antigens recognized by leprosy patients, the serine-rich antigen[34]. Although it is too early to attribute biological functions to the PE and PPE families, it is tempting to speculate that they could be of immunological importance. Two interesting possibilities spring to mind. First, they could represent the principal source of antigenic variation in what is otherwise a genetically and antigenically homogeneous bacterium. Second, these glycine-rich proteins might interfere with immune responses by inhibiting antigen processing.

Several observations and results support the possibility of antigenic variation associated with both the PE and the PPE family proteins. The PGRS member Rv1759 is a fibronectin-binding protein of relative molecular mass 55,000 (ref. 35) that elicits a variable antibody response, indicating either that individuals mount different immune responses or that this PGRS protein may vary between strains of *M. tuberculosis*. The latter possibility is supported by restriction fragment length polymorphisms for various PGRS and MPTR sequences in clinical isolates[33]. Direct support for genetic variation within both the PE and the PPE families was obtained by comparative DNA sequence analysis (Fig. 5). The gene for the PE–PGRS protein Rv0746 of BCG differs from that in H37Rv by the deletion of 29 codons and the insertion of 46 codons. Similar variation was seen in the gene for the PPE protein Rv0442 (data not shown). As these differences were all associated with repetitive sequences they could have resulted from intergenic or intragenic recombinational events or, more probably, from strand slippage during replication[32]. These mechanisms are known to generate antigenic variability in other bacterial pathogens[36].

There are several parallels between the PGRS proteins and the Epstein–Barr virus nuclear antigens (EBNAs). Members of both polypeptide families are glycine-rich, contain extensive Gly–Ala repeats, and exhibit variation in the length of the repeat region between different isolates. The Gly–Ala repeat region of EBNA1 functions as a *cis*-acting inhibitor of the ubiquitin/proteasome antigen-processing pathway that generates peptides presented in the context of major histocompatibility complex (MHC) class I molecules[37,38]. MHC class I knockout mice are very susceptible to *M. tuberculosis*, underlining the importance of a cytotoxic T-cell response in protection against disease[3,39]. Given the many potential effects of the PPE and PE proteins, it is important that further studies are performed to understand their activity. If extensive antigenic variability or reduced antigen presentation were indeed found, this would be significant for vaccine design and for understanding protective immunity in tuberculosis, and might even explain the varied responses seen in different BCG vaccination programmes[40].

**Pathogenicity.** Despite intensive research efforts, there is little information about the molecular basis of mycobacterial virulence[41]. However, this situation should now change as the genome sequence will accelerate the study of pathogenesis as never before, because other bacterial factors that may contribute to virulence are becoming apparent. Before the completion of the genome sequence, only three virulence factors had been described[41]: catalase-peroxidase, which protects against reactive oxygen species produced by the phagocyte; *mce*, which encodes macrophage-colonizing factor[42]; and a sigma factor gene, *sigA* (aka *rpoV*), mutations in which can lead to attenuation[41]. In addition to these single-gene virulence factors, the mycobacterial cell wall[4] is also important in pathology, but the complex nature of its biosynthesis makes it difficult to identify critical genes whose inactivation would lead to attenuation.

On inspection of the genome sequence, it was apparent that four copies of *mce* were present and that these were all situated in operons, comprising eight genes, organized in exactly the same manner. In each case, the genes preceding *mce* code for integral membrane proteins, whereas *mce* and the following five genes are all predicted to encode proteins with signal sequences or hydrophobic stretches at the N terminus. These sets of proteins, about which little is known, may well be secreted or surface-exposed; this is consistent with the proposed role of Mce in invasion of host cells[42]. Furthermore, a homologue of *smpB*, which has been implicated in intracellular survival of *Salmonella typhimurium*, has also been identified[43]. Among the other secreted proteins identified from the genome sequence that could act as virulence factors are a series of phospholipases C, lipases and esterases, which might attack cellular or vacuolar membranes, as well as several proteases. One of these phospholipases acts as a contact-dependent haemolysin (N. Stoker, personal communication). The presence of storage proteins in the bacillus, such as the haemoglobin-like oxygen captors described above, points to its ability to stockpile essential growth factors, allowing it to persist in the nutrient-limited environment of the phagosome. In this regard, the ferritin-like proteins, encoded by *bfrA* and *bfrB*, may be important in intracellular survival as the capacity to acquire enough iron in the vacuole is very limited. □

## Methods

**Sequence analysis.** Initially, ~3.2 Mb of sequence was generated from cosmids[8] and the remainder was obtained from selected BAC clones[7] and 45,000 whole-genome shotgun clones. Sheared fragments (1.4–2.0 kb) from cosmids and BACs were cloned into M13 vectors, whereas genomic DNA was cloned in pUC18 to obtain both forward and reverse reads. The PGRS genes were grossly underrepresented in pUC18 but better covered in the BAC and cosmid M13 libraries. We used small-insert libraries[44] to sequence regions prone to compression or deletion and, in some cases, obtained sequences from products of the polymerase chain reaction or directly from BACs[7]. All shotgun sequencing was performed with standard dye terminators to minimize compression problems, whereas finishing reactions used dRhodamine or BigDye terminators (http://www.sanger.ac.uk). Problem areas were verified by using dye primers. Thirty differences were found between the genomic shotgun sequences and the cosmids; twenty of which were due to sequencing errors and ten to mutations in cosmids (1 error per 320 kb). Less than 0.1% of the sequence was from areas of single-clone coverage, and <0.2% was from one strand with only one sequencing chemistry.

**Informatics.** Sequence assembly involved PHRAP, GAP4 (ref. 45) and a customized perl script that merges sequences from different libraries and generates segments that can be processed by several finishers simultaneously. Sequence analysis and annotation was managed by DIANA (B.G.B. *et al.*, unpublished). Genes encoding proteins were identified by TB-parse[46] using a hidden Markov model trained on known *M. tuberculosis* coding and non-coding regions and translation-initiation signals, with corroboration by positional base preference. Interrogation of the EMBL, TREMBL, SwissProt, PROSITE[47] and in-house databases involved BLASTN, BLASTX[48], DOTTER (http://www.sanger.ac.uk) and FASTA[49]. tRNA genes were located and identified using tRNAscan and tRNAscan-SE[50]. The complete sequence, a list of annotated cosmids and linking regions can be found on our website (http://www. sanger.ac.uk) and in MycDB (http://www.pasteur.fr/mycdb/).

Received 15 April; accepted 8 May 1998.

1. Snider, D. E. Jr, Raviglione, M. & Kochi, A. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 2–11 (Am. Soc. Microbiol., Washington DC, 1994).
2. Wheeler, P. R. & Ratledge, C. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 353–385 (Am. Soc. Microbiol., Washington DC, 1994).
3. Chan, J. & Kaufmann, S. H. E. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 271–284 (Am. Soc. Microbiol., Washington DC, 1994).
4. Brennan, P. J. & Draper, P. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 271–284 (Am. Soc. Microbiol., Washington DC, 1994).
5. Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry

and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24,** 263–270 (1997).
6. Sreevatsan, S. *et al*. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* **94,** 9869–9874 (1997).
7. Brosch, R. *et al*. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing and comparative genomics. *Infect. Immun.* **66,** 2221–2229 (1998).
8. Philipp, W. J. *et al*. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc. Natl Acad. Sci. USA* **93,** 3132–3137 (1996).
9. Blattner, F. R. *et al*. The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1462 (1997).
10. Cole, S. T. & Saint-Girons, I. Bacterial genomics. *FEMS Microbiol. Rev.* **14,** 139–160 (1994).
11. Freiberg, C. *et al*. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387,** 394–401 (1997).
12. Bardarov, S. *et al*. Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **94,** 10961–10966 (1997).
13. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* **178,** 1274–1282 (1996).
14. Kunst, F. *et al*. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–256 (1997).
15. Smith, D. R. *et al*. Multiplex sequencing of 1.5 Mb of the *Mycobacterium leprae* genome. *Genome Res.* **7,** 802–819 (1997).
16. Greenacre, M. *Theory and Application of Correspondence Analysis* (Academic, London, 1984).
17. Ratledge, C. R. in *The Biology of the Mycobacteria* (eds Ratledge, C. & Stanford, J.) 53–94 (Academic, San Diego, 1982).
18. Av-Gay, Y. & Davies, J. Components of eukaryotic-like protein signaling pathways in *Mycobacterium tuberculosis*. *Microb. Comp. Genomics* **2,** 63–73 (1997).
19. Cole, S. T. & Telenti, A. Drug resistance in *Mycobacterium tuberculosis*. *Eur. Resp. Rev.* **8,** 701S–713S (1995).
20. Riley, M. & Labedan, B. in Escherichia coli *and* Salmonella (ed. Neidhardt, F. C.) 2118–2202 (ASM, Washington, 1996).
21. Mdluli, K. *et al*. Inhibition of a *Mycobacterium tuberculosis* β-ketoacyl ACP synthase by isoniazid. *Science* **280,** 1607–1610 (1998).
22. Banerjee, A. *et al*. *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* **263,** 227–230 (1994).
23. Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97,** 2465–2497 (1997).
24. Minnikin, D. E. in *The Biology of the Mycobacteria* (eds Ratledge, C. & Stanford, J.) 95–184 (Academic, London, 1982).
25. Barry, C. E. III *et al*. Mycolic acids: structure, biosynthesis, and phsyiological functions. *Prog. Lipid Res.* (in the press).
26. Belisle, J. T. *et al*. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276,** 1420–1422 (1997).
27. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97,** 2651–2673 (1997).
28. Gobin, J. *et al*. Iron acquisition by *Mycobacterium tuberculosis*: isolation and characterization of a family of iron-binding exochelins. *Proc. Natl Acad. Sci. USA* **92,** 5189–5193 (1995).
29. Young, D. B. & Fruth, U. in *New Generation Vaccines* (eds Levine, M., Woodrow, G., Kaper, J. & Cobon, G. S.) 631–645 (Marcel Dekker, New York, 1997).
30. Sorensen, A. L., Nagai, S., Houen, G., Andersen, P. & Anderson, A. B. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* **63,** 1710–1717 (1995).
31. Hermans, P. W. M., van Soolingen, D. & van Embden, J. D. A. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium gordonae*. *J. Bacteriol.* **174,** 4157–4165 (1992).
32. Poulet, S. & Cole, S. T. Characterisation of the polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch. Microbiol.* **163,** 87–95 (1995).
33. Cole, S. T. & Barrell, B. G. in *Genetics and Tuberculosis* (eds Chadwick, D. J. & Cardew, G., *Novartis Foundation Symp. 217*) 160–172 (Wiley, Chichester, 1998).
34. Vega-Lopez, F. *et al*. Sequence and immunological characterization of a serine-rich antigen from *Mycobacterium leprae*. *Infect. Immun.* **61,** 2145–2153 (1993).
35. Abou-Zeid, C. *et al*. Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect. Immun.* **59,** 2712–2718 (1991).
36. Robertson, B. D. & Meyer, T. F. Genetic variation in pathogenic bacteria. *Trends Genet.* **8,** 422–427 (1992).
37. Levitskaya, J. *et al*. Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* **375,** 685–688 (1995).
38. Levitskaya, J., Sharipo, A., Leonchiks, A., Ciechanover, A. & Masucci, M. G. Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc. Natl Acad. Sci. USA* **94,** 12616–12621 (1997).
39. Flynn, J. L., Goldstein, M. A., Treibold, K. J., Koller, B. & Bloom, B. R. Major histocompatability complex class-I restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection. *Proc. Natl Acad. Sci. USA* **89,** 12013–12017 (1992).
40. Bloom, B. R. & Fine, P. E. M. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 531–557 (Am. Soc. Microbiol., Washington DC, 1994).
41. Collins, D. M. In search of tuberculosis virulence genes. *Trends Microbiol.* **4,** 426–430 (1996).
42. Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T. & Riley, L. W. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261,** 1454–1457 (1993).
43. Baumler, A. J., Kusters, J. G., Stojikovic, I. & Heffron, F. *Salmonella typhimurium* loci involved in survival within macrophages. *Infect. Immun.* **62,** 1623–1630 (1994).
44. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8,** 562–566 (1998).
45. Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **24,** 4992–4999 (1995).
46. Krogh, A., Mian, I. S. & Haussler, D. A hidden Markov model that finds genes in *E. coli* DNA. *Nucleic Acids Res.* **22,** 4768–4778 (1994).
47. Bairoch, A., Bucher, P. & Hofmann, K. The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25,** 217–221 (1997).
48. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. A basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
49. Pearson, W. & Lipman, D. Improved tools for biological sequence comparisons. *Proc. Natl Acad. USA* **85,** 2444–2448 (1988).
50. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic DNA. *Nucleic Acids Res.* **25,** 955–964 (1997).

**Acknowledgements.** We thank Y. Av-Gay, F.-C. Bange, A. Danchin, B. Dujon, W. R. Jacobs Jr, L. Jones, M. McNeil, I. Moszer, P. Rice and J. Stephenson for advice, reagents and support. This work was supported by the Wellcome Trust. Additional funding was provided by the Association Française Raoul Follereau, the World Health Organisation and the Institut Pasteur. S.V.G. received a Wellcome Trust travelling research fellowship.

Correspondence and requests for materials should be addressed to B.G.B. (barrell@sanger.ac.uk) or S.T.C. (stcole@pasteur.fr). The complete sequence has been deposited in EMBL/GenBank/DDBJ as MTBH37RV, accession number AL123456.

## Table 1. Functional classification of *Mycobacterium tuberculosis* protein-coding genes

**I. Small-molecule metabolism**

*A. Degradation*

1. Carbon compounds

| | | |
|---|---|---|
| Rv0186 | *bglS* | β-glucosidase |
| Rv2202c | *cbhK* | carbohydrate kinase |
| Rv0727c | *fucA* | L-fuculose phosphate aldolase |
| Rv1731 | *gabD1* | succinate-semialdehyde dehydrogenase |
| Rv0234c | *gabD2* | succinate-semialdehyde dehydrogenase |
| Rv0501 | *galE1* | UDP-glucose 4-epimerase |
| Rv0536 | *galE2* | UDP-glucose 4-epimerase |
| Rv0620 | *galK* | galactokinase |
| Rv0619 | *galT* | galactose-1-phosphate uridylyltransferase C-term |
| Rv0618 | *galT'* | galactose-1-phosphate uridylyltransferase N-term |
| Rv0993 | *galU* | UTP-glucose-1-phosphate uridylyltransferase |
| Rv3696c | *glpK* | ATP:glycerol 3-phosphotransferase |
| Rv3255c | *manA* | mannose-6-phosphate isomerase |
| Rv3441c | *mrsA* | phosphoglucomutase or phosphomannomutase |
| Rv0118c | *oxcA* | oxalyl-CoA decarboxylase |
| Rv3068c | *pgmA* | phosphoglucomutase |
| Rv3257c | *pmmA* | phosphomannomutase |
| Rv3308 | *pmmB* | phosphomannomutase |
| Rv2702 | *ppgK* | polyphosphate glucokinase |
| Rv0408 | *pta* | phosphate acetyltransferase |
| Rv0729 | *xylB* | xylulose kinase |
| Rv1096 | - | carbohydrate degrading enzyme |

2. Amino acids and amines

| | | |
|---|---|---|
| Rv1905c | *aao* | D-amino acid oxidase |
| Rv2531c | *adi* | ornithine/arginine decarboxylase |
| Rv2780 | *ald* | L-alanine dehydrogenase |
| Rv1538c | *ansA* | L-asparaginase |
| Rv1001 | *arcA* | arginine deiminase |
| Rv0753c | *mmsA* | methylmalmonate semialdehyde dehydrogenase |
| Rv0751c | *mmsB* | methylmalmonate semialdehyde oxidoreductase |
| Rv1187 | *rocA* | pyrroline-5-carboxylate dehydrogenase |
| Rv2322c | *rocD1* | ornithine aminotransferase |
| Rv2321c | *rocD2* | ornithine aminotransferase |
| Rv1848 | *ureA* | urease γ subunit |
| Rv1849 | *ureB* | urease β subunit |
| Rv1850 | *ureC* | urease α subunit |
| Rv1853 | *ureD* | urease accessory protein |
| Rv1851 | *ureF* | urease accessory protein |
| Rv1852 | *ureG* | urease accessory protein |
| Rv2913c | - | probable D-amino acid aminohydrolase |
| Rv3551 | - | possible glutaconate CoA-transferase |

3. Fatty acids

| | | |
|---|---|---|
| Rv2501c | *accA1* | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv0973c | *accA2* | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv2502c | *accD1* | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv0974c | *accD2* | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv3667 | *acs* | acetyl-CoA synthase |
| Rv3409c | *choD* | cholesterol oxidase |
| Rv0222 | *echA1* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0456c | *echA2* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0632c | *echA3* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0673 | *echA4* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0675 | *echA5* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0905 | *echA6* | enoyl-CoA hydratase/isomerase superfamily (aka *eccH*) |
| Rv0971c | *echA7* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1070c | *echA8* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1071c | *echA9* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1142c | *echA10* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1141c | *echA11* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1472 | *echA12* | enoyl-CoA hydratase/isomerase superfamily |
| Rv1935c | *echA13* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2486 | *echA14* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2679 | *echA15* | enoyl-CoA hydratase/isomerase superfamily |
| Rv2831 | *echA16* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3039c | *echA17* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3373 | *echA18* | enoyl-CoA hydratase/isomerase superfamily, N-term |
| Rv3374 | *echA18'* | enoyl-CoA hydratase/isomerase superfamily, C-term |
| Rv3516 | *echA19* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3550 | *echA20* | enoyl-CoA hydratase/isomerase superfamily |
| Rv3774 | *echA21* | enoyl-CoA hydratase/isomerase superfamily |
| Rv0859 | *fadA* | β oxidation complex, β subunit (acetyl-CoA C-acetyltransferase) |
| Rv0243 | *fadA2* | acetyl-CoA C-acetyltransferase |
| Rv1074c | *fadA3* | acetyl-CoA C-acetyltransferase |
| Rv1323 | *fadA4* | acetyl-CoA C-acetyltransferase (aka *thiL*) |
| Rv3546 | *fadA5* | acetyl-CoA C-acetyltransferase |
| Rv3556c | *fadA6* | acetyl-CoA C-acetyltransferase |
| Rv0860 | *fadB* | β oxidation complex, α subunit (multiple activities) |
| Rv0468 | *fadB2* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1715 | *fadB3* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv3141 | *fadB4* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1912c | *fadB5* | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1750c | *fadD1* | acyl-CoA synthase |
| Rv0270 | *fadD2* | acyl-CoA synthase |
| Rv3561 | *fadD3* | acyl-CoA synthase |
| Rv0214 | *fadD4* | acyl-CoA synthase |
| Rv0166 | *fadD5* | acyl-CoA synthase |
| Rv1206 | *fadD6* | acyl-CoA synthase |
| Rv0119 | *fadD7* | acyl-CoA synthase |
| Rv0551c | *fadD8* | acyl-CoA synthase |
| Rv2590 | *fadD9* | acyl-CoA synthase |
| Rv0099 | *fadD10* | acyl-CoA synthase |
| Rv1550 | *fadD11* | acyl-CoA synthase, N-term |
| Rv1549 | *fadD11'* | acyl-CoA synthase, C-term |
| Rv1427c | *fadD12* | acyl-CoA synthase |
| Rv3089 | *fadD13* | acyl-CoA synthase |
| Rv1058 | *fadD14* | acyl-CoA synthase |
| Rv2187 | *fadD15* | acyl-CoA synthase |
| Rv0852 | *fadD16* | acyl-CoA synthase |
| Rv3506 | *fadD17* | acyl-CoA synthase |
| Rv3513c | *fadD18* | acyl-CoA synthase |
| Rv3515c | *fadD19* | acyl-CoA synthase |
| Rv1185c | *fadD21* | acyl-CoA synthase |
| Rv2948c | *fadD22* | acyl-CoA synthase |
| Rv3826 | *fadD23* | acyl-CoA synthase |
| Rv1529 | *fadD24* | acyl-CoA synthase |
| Rv1521 | *fadD25* | acyl-CoA synthase |
| Rv2930 | *fadD26* | acyl-CoA synthase |
| Rv0275c | *fadD27* | acyl-CoA synthase |
| Rv2941 | *fadD28* | acyl-CoA synthase |
| Rv2950c | *fadD29* | acyl-CoA synthase |
| Rv0404 | *fadD30* | acyl-CoA synthase |
| Rv1925 | *fadD31* | acyl-CoA synthase |
| Rv3801c | *fadD32* | acyl-CoA synthase |
| Rv1345 | *fadD33* | acyl-CoA synthase |
| Rv0035 | *fadD34* | acyl-CoA synthase |
| Rv2505c | *fadD35* | acyl-CoA synthase |
| Rv1193 | *fadD36* | acyl-CoA synthase |
| Rv0131c | *fadE1* | acyl-CoA dehydrogenase |
| Rv0154c | *fadE2* | acyl-CoA dehydrogenase |
| Rv0215c | *fadE3* | acyl-CoA dehydrogenase |
| Rv0231 | *fadE4* | acyl-CoA dehydrogenase |
| Rv0244c | *fadE5* | acyl-CoA dehydrogenase |
| Rv0271c | *fadE6* | acyl-CoA dehydrogenase |
| Rv0400c | *fadE7* | acyl-CoA dehydrogenase |
| Rv0672 | *fadE8* | acyl-CoA dehydrogenase (aka *aidB*) |
| Rv0752c | *fadE9* | acyl-CoA dehydrogenase |
| Rv0873 | *fadE10* | acyl-CoA dehydrogenase |
| Rv0972c | *fadE12* | acyl-CoA dehydrogenase |
| Rv0975c | *fadE13* | acyl-CoA dehydrogenase |
| Rv1346 | *fadE14* | acyl-CoA dehydrogenase |
| Rv1467c | *fadE15* | acyl-CoA dehydrogenase |
| Rv1679 | *fadE16* | acyl-CoA dehydrogenase |
| Rv1934c | *fadE17* | acyl-CoA dehydrogenase |
| Rv1933c | *fadE18* | acyl-CoA dehydrogenase |
| Rv2500c | *fadE19* | acyl-CoA dehydrogenase (aka *mmgC*) |
| Rv2724c | *fadE20* | acyl-CoA dehydrogenase |
| Rv2789c | *fadE21* | acyl-CoA dehydrogenase |
| Rv3061c | *fadE22* | acyl-CoA dehydrogenase |
| Rv3140 | *fadE23* | acyl-CoA dehydrogenase |
| Rv3139 | *fadE24* | acyl-CoA dehydrogenase |
| Rv3274c | *fadE25* | acyl-CoA dehydrogenase |
| Rv3504 | *fadE26* | acyl-CoA dehydrogenase |
| Rv3505 | *fadE27* | acyl-CoA dehydrogenase |
| Rv3544c | *fadE28* | acyl-CoA dehydrogenase |
| Rv3543c | *fadE29* | acyl-CoA dehydrogenase |
| Rv3560c | *fadE30* | acyl-CoA dehydrogenase |
| Rv3562 | *fadE31* | acyl-CoA dehydrogenase |
| Rv3563 | *fadE32* | acyl-CoA dehydrogenase |
| Rv3564 | *fadE33* | acyl-CoA dehydrogenase |
| Rv3573c | *fadE34* | acyl-CoA dehydrogenase |
| Rv3797 | *fadE35* | acyl-CoA dehydrogenase |
| Rv3761c | *fadE36* | acyl-CoA dehydrogenase |
| Rv1175c | *fadH* | 2,4-Dienoyl-CoA Reductase |
| Rv0855 | *far* | fatty acyl-CoA racemase |
| Rv1143 | *mcr* | α-methyl acyl-CoA racemase |
| Rv1492 | *mutA* | methylmalonyl-CoA mutase, β subunit |
| Rv1493 | *mutB* | methylmalonyl-CoA mutase, α subunit |
| Rv2504c | *scoA* | 3-oxo acid:CoA transferase, α subunit |
| Rv2503c | *scoB* | 3-oxo acid:CoA transferase, β subunit |
| Rv1136 | - | probable carnitine racemase |
| Rv1683 | - | possible acyl-CoA synthase |

4. Phosphorous compounds

| | | |
|---|---|---|
| Rv2368c | *phoH* | ATP-binding *pho* regulon component |
| Rv1095 | *phoH2* | PhoH-like protein |
| Rv3628 | *ppa* | probable inorganic pyrophosphatase |
| Rv2984 | *ppk* | polyphosphate kinase |

*B. Energy metabolism*

1. Glycolysis

| | | |
|---|---|---|
| Rv1023 | *eno* | enolase |
| Rv0363c | *fba* | fructose bisphosphate aldolase |
| Rv1436 | *gap* | glyceraldehyde 3-phosphate dehydrogenase |
| Rv0489 | *gpm* | phosphoglycerate mutase I |
| Rv3010c | *pfkA* | phosphofructokinase I |
| Rv2029c | *pfkB* | phosphofructokinase II |
| Rv0946c | *pgi* | glucose-6-phosphate isomerase |
| Rv1437 | *pgk* | phosphoglycerate kinase |
| Rv1617 | *pykA* | pyruvate kinase |
| Rv1438 | *tpi* | triosephosphate isomerase |
| Rv2419c | - | putative phosphoglycerate mutase |
| Rv3837c | - | putative phosphoglycerate mutase |

2. Pyruvate dehydrogenase

| | | |
|---|---|---|
| Rv2241 | *aceE* | pyruvate dehydrogenase E1 component |
| Rv3303c | *lpdA* | dihydrolipoamide dehydrogenase |
| Rv2497c | *pdhA* | pyruvate dehydrogenase E1 component α subunit |
| Rv2496c | *pdhB* | pyruvate dehydrogenase E1 component β subunit |
| Rv2495c | *pdhC* | dihydrolipoamide acetyltransferase |
| Rv0462 | - | probable dihydrolipoamide dehydrogenase |

3. TCA cycle

| | | |
|---|---|---|
| Rv1475c | *acn* | aconitate hydratase |
| Rv0889c | *citA* | citrate synthase 2 |
| Rv2498c | *citE* | citrate lyase β chain |
| Rv1098c | *fum* | fumarase |
| Rv1131 | *gltA1* | citrate synthase 3 |
| Rv0896 | *gltA2* | citrate synthase 1 |
| Rv3339c | *icd1* | isocitrate dehydrogenase |
| Rv0066c | *icd2* | isocitrate dehydrogenase |
| Rv0794c | *lpdB* | dihydrolipoamide dehydrogenase |
| Rv1240 | *mdh* | malate dehydrogenase |
| Rv2967c | *pca* | pyruvate carboxylase |
| Rv3318 | *sdhA* | succinate dehydrogenase A |
| Rv3319 | *sdhB* | succinate dehydrogenase B |
| Rv3316 | *sdhC* | succinate dehydrogenase C subunit |
| Rv3317 | *sdhD* | succinate dehydrogenase D subunit |
| Rv1248c | *sucA* | 2-oxoglutarate dehydrogenase |
| Rv2215 | *sucB* | dihydrolipoamide succinyltransferase |
| Rv0951 | *sucC* | succinyl-CoA synthase β chain |
| Rv0952 | *sucD* | succinyl-CoA synthase α chain |

4. Glyoxylate bypass

| | | |
|---|---|---|
| Rv0467 | *aceA* | isocitrate lyase |
| Rv1915 | *aceAa* | isocitrate lyase, α module |
| Rv1916 | *aceAb* | isocitrate lyase, β module |
| Rv1837c | *glcB* | malate synthase |
| Rv3323c | *gphA* | phosphoglycolate phosphatase |

5. Pentose phosphate pathway

| | | |
|---|---|---|
| Rv1445c | *devB* | glucose-6-phosphate 1-dehydrogenase |
| Rv1844c | *gnd* | 6-phosphogluconate dehydrogenase (Gram –) |
| Rv1122 | *gnd2* | 6-phosphogluconate dehydrogenase (Gram +) |
| Rv1446c | *opcA* | unknown function, may aid G6PDH |

| | | |
|---|---|---|
| Rv2436 | *rbsK* | ribokinase |
| Rv1408 | *rpe* | ribulose-phosphate 3-epimerase |
| Rv2465c | *rpi* | phosphopentose isomerase |
| Rv1448c | *tal* | transaldolase |
| Rv1449c | *tkt* | transketolase |
| Rv1121 | *zwf* | glucose-6-phosphate 1-dehydrogenase |
| Rv1447c | *zwf2* | glucose-6-phosphate 1-dehydrogenase |

6. Respiration

*a. aerobic*

| | | |
|---|---|---|
| Rv0527 | *ccsA* | cytochrome *c*-type biogenesis protein |
| Rv0529 | *ccsB* | cytochrome *c*-type biogenesis protein |
| Rv1451 | *ctaB* | cytochrome *c* oxidase assembly factor |
| Rv2200c | *ctaC* | cytochrome *c* oxidase chain II |
| Rv3043c | *ctaD* | cytochrome *c* oxidase polypeptide I |
| Rv2193 | *ctaE* | cytochrome *c* oxidase polypeptide III |
| Rv1542c | *glbN* | hemoglobin-like, oxygen carrier |
| Rv2470 | *glbO* | hemoglobin-like, oxygen carrier |
| Rv2249c | *glpD1* | glycerol-3-phosphate dehydrogenase |
| Rv3302c | *glpD2* | glycerol-3-phosphate dehydrogenase |
| Rv0694 | *lldD1* | L-lactate dehydrogenase (cytochrome) |
| Rv1872c | *lldD2* | L-lactate dehydrogenase |
| Rv1854c | *ndh* | probable NADH dehydrogenase |
| Rv3145 | *nuoA* | NADH dehydrogenase chain A |
| Rv3146 | *nuoB* | NADH dehydrogenase chain B |
| Rv3147 | *nuoC* | NADH dehydrogenase chain C |
| Rv3148 | *nuoD* | NADH dehydrogenase chain D |
| Rv3149 | *nuoE* | NADH dehydrogenase chain E |
| Rv3150 | *nuoF* | NADH dehydrogenase chain F |
| Rv3151 | *nuoG* | NADH dehydrogenase chain G |
| Rv3152 | *nuoH* | NADH dehydrogenase chain H |
| Rv3153 | *nuoI* | NADH dehydrogenase chain I |
| Rv3154 | *nuoJ* | NADH dehydrogenase chain J |
| Rv3155 | *nuoK* | NADH dehydrogenase chain K |
| Rv3156 | *nuoL* | NADH dehydrogenase chain L |
| Rv3157 | *nuoM* | NADH dehydrogenase chain M |
| Rv3158 | *nuoN* | NADH dehydrogenase chain N |
| Rv2195 | *qcrA* | Rieske iron-sulphur component of *ubiQ-cytB* reductase |
| Rv2196 | *qcrB* | cytochrome β component of *ubiQ-cytB* reductase |
| Rv2194 | *qcrC* | cytochrome *b/c* component of *ubiQ-cytB* reductase |

*b. anaerobic*

| | | |
|---|---|---|
| Rv2392 | *cysH* | 3'-phosphoadenylylsulfate (PAPS) reductase |
| Rv2899c | *fdhD* | affects formate dehydrogenase-N |
| Rv2900c | *fdhF* | molybdopterin-containing oxidoreductase |
| Rv1552 | *frdA* | fumarate reductase flavoprotein subunit |
| Rv1553 | *frdB* | fumarate reductase iron sulphur protein |
| Rv1554 | *frdC* | fumarate reductase 15kD anchor protein |
| Rv1555 | *frdD* | fumarate reductase 13kD anchor protein |
| Rv1161 | *narG* | nitrate reductase α subunit |
| Rv1162 | *narH* | nitrate reductase β chain |
| Rv1164 | *narI* | nitrate reductase γ chain |
| Rv1163 | *narJ* | nitrate reductase δ chain |
| Rv1736c | *narX* | fused nitrate reductase |
| Rv2391 | *nirA* | probable nitrite reductase/sulphite reductase |
| Rv0252 | *nirB* | nitrite reductase flavoprotein |
| Rv0253 | *nirD* | probable nitrite reductase small subunit |

*c. Electron transport*

| | | |
|---|---|---|
| Rv0409 | *ackA* | acetate kinase |
| Rv1623c | *appC* | cytochrome *bd-II* oxidase subunit I |
| Rv1622c | *cydB* | cytochrome *d* ubiquinol oxidase subunit II |
| Rv1620c | *cydC* | ABC transporter |
| Rv1621c | *cydD* | ABC transporter |
| Rv2007c | *fdxA* | ferredoxin |
| Rv3554 | *fdxB* | ferredoxin |
| Rv1177 | *fdxC* | ferredoxin 4Fe-4S |
| Rv3503c | *fdxD* | probable ferredoxin |
| Rv3029c | *fixA* | electron transfer flavoprotein β subunit |
| Rv3028c | *fixB* | electron transfer flavoprotein α subunit |
| Rv3106 | *fprA* | adrenodoxin and NADPH ferredoxin reductase |
| Rv0886 | *fprB* | ferredoxin, ferredoxin-NADP reductase |
| Rv3251c | *rubA* | rubredoxin A |
| Rv3250c | *rubB* | rubredoxin B |

7. Miscellaneous oxidoreductases and oxygenases 171

8. ATP-proton motive force

| | | |
|---|---|---|
| Rv1308 | *atpA* | ATP synthase α chain |
| Rv1304 | *atpB* | ATP synthase α chain |
| Rv1311 | *atpC* | ATP synthase ε chain |
| Rv1310 | *atpD* | ATP synthase β chain |
| Rv1305 | *atpE* | ATP synthase c chain |
| Rv1306 | *atpF* | ATP synthase b chain |
| Rv1309 | *atpG* | ATP synthase γ chain |
| Rv1307 | *atpH* | ATP synthase δ chain |

*C. Central intermediary metabolism*

1. General

| | | |
|---|---|---|
| Rv2589 | *gabT* | 4-aminobutyrate aminotransferase |
| Rv3432c | *gadB* | glutamate decarboxylase |
| Rv1832 | *gcvB* | glycine decarboxylase |
| Rv1826 | *gcvH* | glycine cleavage system H protein |
| Rv2211c | *gcvT* | T protein of glycine cleavage system |
| Rv1213 | *glgC* | glucose-1-phosphate adenylyltransferase |
| Rv3842c | *glpQ1* | glycerophosphoryl diester phosphodiesterase |
| Rv0317c | *glpQ2* | glycerophosphoryl diester phosphodiesterase |
| Rv3566c | *nhoA* | N-hydroxyarylamine *o*-acetyltransferase |
| Rv0155 | *pntAA* | pyridine transhydrogenase subunit α1 |
| Rv0156 | pntAB | pyridine transhydrogenase subunit α2 |
| Rv0157 | *pntB* | pyridine transhydrogenase subunit β |
| Rv1127c | *ppdK* | similar to pyruvate, phosphate dikinase |

2. Gluconeogenesis

| | | |
|---|---|---|
| Rv0211 | *pckA* | phosphoenolpyruvate carboxykinase |
| Rv0069c | *sdaA* | L-serine dehydratase 1 |

3. Sugar nucleotides

| | | |
|---|---|---|
| Rv1512 | *epiA* | nucleotide sugar epimerase |
| Rv3784 | *epiB* | probable UDP-galactose 4-epimerase |
| Rv1511 | *gmdA* | GDP-mannose 4,6 dehydratase |
| Rv0334 | *rmlA* | glucose-1-phosphate thymidyltransferase |
| Rv3264c | *rmlA2* | glucose-1-phosphate thymidyltransferase |
| Rv3464 | *rmlB* | dTDP-glucose 4,6-dehydratase |
| Rv3634c | *rmlB2* | dTDP-glucose 4,6-dehydratase |
| Rv3468c | *rmlB3* | dTDP-glucose 4,6-dehydratase |
| Rv3465 | *rmlC* | dTDP-4-dehydrorhamnose 3,5-epimerase |
| Rv3266c | *rmlD* | dTDP-4-dehydrorhamnose reductase |
| Rv0322 | *udgA* | UDP-glucose dehydrogenase/GDP-mannose 6-dehydrogenase |
| Rv3265c | *wbbL* | dTDP-rhamnosyl transferase |
| Rv1525 | *wbbl2* | dTDP-rhamnosyl transferase |
| Rv3400 | - | probable β-phosphoglucomutase |

4. Amino sugars

| | | |
|---|---|---|
| Rv3436c | *glmS* | glucosamine-fructose-6-phosphate aminotransferase |

5. Sulphur metabolism

| | | |
|---|---|---|
| Rv0711 | *atsA* | arylsulfatase |
| Rv3299c | *atsB* | proable arylsulfatase |
| Rv0663 | *atsD* | proable arylsulfatase |
| Rv3077 | *atsF* | proable arylsulfatase |
| Rv0296c | *atsG* | proable arylsulfatase |
| Rv3796 | *atsH* | proable arylsulfatase |
| Rv1285 | *cysD* | ATP:sulphurylase subunit 2 |
| Rv1286 | *cysN* | ATP:sulphurylase subunit 1 |
| Rv2131c | *cysQ* | homologue of *M.leprae cysQ* |
| Rv3248c | *sahH* | adenosylhomocysteinase |
| Rv3283 | *sseA* | thiosulfate sulfurtransferase |
| Rv2291 | *sseB* | thiosulfate sulfurtransferase |
| Rv3118 | *sseC* | thiosulfate sulfurtransferase |
| Rv0814c | *sseC2* | thiosulfate sulfurtransferase |
| Rv3762c | - | probable alkyl sulfatase |

*D. Amino acid biosynthesis*

1. Glutamate family

| | | |
|---|---|---|
| Rv1654 | *argB* | acetylglutamate kinase |
| Rv1652 | *argC* | N-acetyl-γ-glutamyl-phosphate reductase |
| Rv1655 | *argD* | acetylornithine aminotransferase |
| Rv1656 | *argF* | ornithine carbamoyltransferase |
| Rv1658 | *argG* | arginosuccinate synthase |
| Rv1659 | *argH* | arginosuccinate lyase |
| Rv1653 | *argJ* | glutamate N-acetyltransferase |
| Rv2220 | *glnA1* | glutamine synthase class I |
| Rv2222c | *glnA2* | glutamine synthase class II |
| Rv1878 | *glnA3* | probable glutamine synthase |
| Rv2860c | *glnA4* | proable glutamine synthase |
| Rv2918c | *glnD* | uridylyltransferase |
| Rv2221c | *glnE* | glutamate-ammonia-ligase adenyltransferase |
| Rv3859c | *gltB* | ferredoxin-dependent glutamate synthase |
| Rv3858c | *gltD* | small subunit of NADH-dependent glutamate synthase |
| Rv3704c | *gshA* | possible γ-glutamylcysteine synthase |
| Rv2427c | *proA* | γ-glutamyl phosphate reductase |
| Rv2439c | *proB* | glutamate 5-kinase |
| Rv0500 | *proC* | pyrroline-5-carboxylate reductase |

2. Aspartate family

| | | |
|---|---|---|
| Rv3708c | *asd* | aspartate semialdehyde dehydrogenase |
| Rv3709c | *ask* | aspartokinase |
| Rv2201 | *asnB* | asparagine synthase B |
| Rv3565 | *aspB* | aspartate aminotransferase |
| Rv0337c | *aspC* | aspartate aminotransferase |
| Rv2753c | *dapA* | dihydrodipicolinate synthase |
| Rv2773c | *dapB* | dihydrodipicolinate reductase |
| Rv1202 | *dapE* | succinyl-diaminopimelate desuccinylase |
| Rv2141c | *dapE2* | ArgE/DapE/Acy1/Cpg2/yscS family |
| Rv2726c | *dapF* | diaminopimelate epimerase |
| Rv1293 | *lysA* | diaminopimelate decarboxylase |
| Rv3341 | *metA* | homoserine *o*-acetyltransferase |
| Rv1079 | *metB* | cystathionine γ-synthase |
| Rv3340 | *metC* | cystathionine β-lyase |
| Rv1133c | *metE* | 5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase |
| Rv2124c | *metH* | 5-methyltetrahydrofolate-homocysteine methyltransferase |
| Rv1392 | *metK* | *S*-adenosylmethionine synthase |
| Rv0391 | *metZ* | *o*-succinylhomoserine sulfhydrylase |
| Rv1294 | *thrA* | homoserine dehydrogenase |
| Rv1296 | *thrB* | homoserine kinase |
| Rv1295 | *thrC* | homoserine synthase |

3. Serine family

| | | |
|---|---|---|
| Rv0815c | *cysA2* | thiosulfate sulfurtransferase |
| Rv3117 | *cysA3* | thiosulfate sulfurtransferase |
| Rv2335 | *cysE* | serine acetyltransferase |
| Rv0511 | *cysG* | uroporphyrin-III c-methyltransferase |
| Rv2847c | *cysG2* | multifunctional enzyme, siroheme synthase |
| Rv2334 | *cysK* | cysteine synthase A |
| Rv1336 | *cysM* | cysteine synthase B |
| Rv1077 | *cysM2* | cystathionine β-synthase |
| Rv0848 | *cysM3* | putative cysteine synthase |
| Rv1093 | *glyA* | serine hydroxymethyltransferase |
| Rv0070c | glyA2 | serine hydroxymethyltransferase |
| Rv2996c | *serA* | D-3-phosphoglycerate dehydrogenase |
| Rv0505c | *serB* | probable phosphoserine phosphatase |
| Rv3042c | *serB2* | C-term similar to phosphoserine phosphatase |
| Rv0884c | *serC* | phosphoserine aminotransferase |

4. Aromatic amino acid family

| | | |
|---|---|---|
| Rv3227 | *aroA* | 3-phosphoshikimate 1-carboxyvinyl transferase |
| Rv2538c | *aroB* | 3-dehydroquinate synthase |
| Rv2537c | *aroD* | 3-dehydroquinate dehydratase |
| Rv2552c | *aroE* | shikimate 5-dehydrogenase |
| Rv2540c | *aroF* | chorismate synthase |
| Rv2178c | *aroG* | DAHP synthase |
| Rv2539c | *aroK* | shikimate kinase I |
| Rv3838c | *pheA* | prephenate dehydratase |
| Rv1613 | *trpA* | tryptophan synthase α chain |
| Rv1612 | *trpB* | tryptophan synthase β chain |
| Rv1611 | *trpC* | indole-3-glycerol phosphate synthase |
| Rv2192c | *trpD* | anthranilate phosphoribosyltransferase |
| Rv1609 | *trpE* | anthranilate synthase component I |
| Rv2386c | *trpE2* | anthranilate synthase component I |
| Rv3754 | *tyrA* | prephenate dehydrogenase |

5. Histidine

| | | |
|---|---|---|
| Rv1603 | *hisA* | phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide isomerase |
| Rv1601 | *hisB* | imidazole glycerol-phosphate dehydratase |
| Rv1600 | *hisC* | histidinol-phosphate aminotransferase |
| Rv3772 | *hisC2* | histidinol-phosphate aminotransferase |
| Rv1599 | *hisD* | histidinol dehydrogenase |

| | | |
|---|---|---|
| Rv1605 | *hisF* | imidazole glycerol-phosphate synthase |
| Rv2121c | *hisG* | ATP phosphoribosyltransferase |
| Rv1602 | *hisH* | amidotransferase |
| Rv2122c | *hisI* | phosphoribosyl-AMP cyclohydrolase |
| Rv1606 | *hisI2* | probable phosphoribosyl-AMP 1,6 cyclohydrolase |
| Rv0114 | - | similar to HisB |

6. Pyruvate family

| | | |
|---|---|---|
| Rv3423c | *alr* | alanine racemase |

7. Branched amino acid family

| | | |
|---|---|---|
| Rv1559 | *ilvA* | threonine deaminase |
| Rv3003c | *ilvB* | acetolactate synthase I large subunit |
| Rv3470c | *ilvB2* | acetolactate synthase large subunit |
| Rv3001c | *ilvC* | ketol-acid reductoisomerase |
| Rv0189c | *ilvD* | dihydroxy-acid dehydratase |
| Rv2210c | *ilvE* | branched-chain-amino-acid transaminase |
| Rv1820 | *ilvG* | acetolactate synthase II |
| Rv3002c | *ilvN* | acetolactate synthase I small subunit |
| Rv3509c | *ilvX* | probable acetohydroxyacid synthase I large subunit |
| Rv3710 | *leuA* | α-isopropyl malate synthase |
| Rv2995c | *leuB* | 3-isopropylmalate dehydrogenase |
| Rv2988c | *leuC* | 3-isopropylmalate dehydratase large subunit |
| Rv2987c | *leuD* | 3-isopropylmalate dehydratase small subunit |

*E. Polyamine synthesis*

| | | |
|---|---|---|
| Rv2601 | *speE* | spermidine synthase |

*F. Purines, pyrimidines, nucleosides and nucleotides*

1. Purine ribonucleotide biosynthesis

| | | |
|---|---|---|
| Rv1389 | *gmk* | putative guanylate kinase |
| Rv3396c | *guaA* | GMP synthase |
| Rv1843c | *guaB1* | inosine-5'-monophosphate dehydrogenase |
| Rv3411c | *guaB2* | inosine-5'-monophosphate dehydrogenase |
| Rv3410c | *guaB3* | inosine-5'-monophosphate dehydrogenase |
| Rv1017c | *prsA* | ribose-phosphate pyrophosphokinase |
| Rv0357c | *purA* | adenylosuccinate synthase |
| Rv0777 | *purB* | adenylosuccinate lyase |
| Rv0780 | *purC* | phosphoribosylaminoimidazole-succinocarboxamide synthase |
| Rv0772 | *purD* | phosphoribosylamine-glycine ligase |
| Rv3275c | *purE* | phosphoribosylaminoimidazole carboxylase |
| Rv0808 | *purF* | amidophosphoribosyltransferase- |
| Rv0957 | *purH* | phosphoribosylaminoimidazole-carboxamide formyltransferase |
| Rv3276c | *purK* | phosphoribosylaminoimidazole carboxylase ATPase subunit |
| Rv0803 | *purL* | phosphoribosylformylglycinamidine synthase II |
| Rv0809 | *purM* | 5'-phosphoribosyl-5-aminoimidazole synthase |
| Rv0956 | *purN* | phosphoribosylglycinamide formyltransferase I |
| Rv0788 | *purQ* | phosphoribosylformylglycinamidine synthase I |
| Rv0389 | *purT* | phosphoribosylglycinamide formyltransferase II |
| Rv2964 | *purU* | formyltetrahydrofolate deformylase |

2. Pyrimidine ribonucleotide biosynthesis

| | | |
|---|---|---|
| Rv1383 | *carA* | carbamoyl-phosphate synthase subunit |
| Rv1384 | *carB* | carbamoyl-phosphate synthase subunit |
| Rv1380 | *pyrB* | aspartate carbamoyltransferase |
| Rv1381 | *pyrC* | dihydroorotase |
| Rv2139 | pyrD | dihydroorotate dehydrogenase |
| Rv1385 | pyrF | orotidine 5'-phosphate decarboxylase |
| Rv1699 | *pyrG* | CTP synthase |
| Rv2883c | *pyrH* | uridylate kinase |
| Rv0382c | *umpA* | probable uridine 5'-monophosphate synthase |

3. 2'-deoxyribonucleotide metabolism

| | | |
|---|---|---|
| Rv0321 | *dcd* | deoxycytidine triphosphate deaminase |
| Rv2697c | *dut* | deoxyuridine triphosphatase |
| Rv0233 | *nrdB* | ribonucleoside-diphosphate reductase B2 (eukaryotic-like) |
| Rv3051c | *nrdE* | ribonucleoside diphosphate reductase α chain |
| Rv1981c | *nrdF* | ribonucleotide reductase small subunit |
| Rv3048c | *nrdG* | ribonucleoside-diphosphate small subunit |
| Rv3053c | *nrdH* | glutaredoxin electron transport component of NrdEF system |
| Rv3052c | *nrdI* | NrdI/YgaO/YmaA family |
| Rv3247c | *tmk* | thymidylate kinase |
| Rv2764c | *thyA* | thymidylate synthase |
| Rv0570 | *nrdZ* | ribonucleotide reductase, class II |
| Rv3752c | - | probable cytidine/deoxycytidylate deaminase |

4. Salvage of nucleosides and nucleotides

| | | |
|---|---|---|
| Rv3313c | *add* | probable adenosine deaminase |
| Rv2584c | *apt* | adenine phosphoribosyltransferases |
| Rv3315c | *cdd* | probable cytidine deaminase |
| Rv3314c | *deoA* | thymidine phosphorylase |
| Rv0478 | *deoC* | deoxyribose-phosphate aldolase |
| Rv3307 | *deoD* | probable purine nucleoside phosphorylase |
| Rv3624c | *hpt* | probable hypoxanthine-guanine phosphoribosyltransferase |
| Rv3393 | *iunH* | probable inosine-uridine preferring nucleoside hydrolase |
| Rv0535 | *pnp* | phosphorylase from Pnp/MtaP family 2 |
| Rv3309c | *upp* | uracil phophoribosyltransferase |

5. Miscellaneous nucleoside/nucleotide reactions

| | | |
|---|---|---|
| Rv0733 | *adk* | probable adenylate kinase |
| Rv2364c | *bex* | GTP-binding protein of Era/ThdF family |
| Rv1712 | *cmk* | cytidylate kinase |
| Rv2344c | *dgt* | probable deoxyguanosine triphosphate hydrolase |
| Rv2404c | *lepA* | GTP-binding protein LepA |
| Rv2727c | *miaA* | tRNA δ(2)-isopentenylpyrophosphate transferase |
| Rv2445c | *ndkA* | nucleoside diphosphate kinase |
| Rv2440c | *obg* | Obg GTP-binding protein |
| Rv2583c | *relA* | (p)ppGpp synthase I |

*G. Biosynthesis of cofactors, prosthetic groups and carriers*

1. Biotin

| | | |
|---|---|---|
| Rv1568 | *bioA* | adenosylmethionine-8-amino-7-oxononanoate aminotransferase |
| Rv1589 | *bioB* | biotin synthase |
| Rv1570 | *bioD* | dethiobiotin synthase |
| Rv1569 | *bioF* | 8-amino-7-oxononanoate synthase |
| Rv0032 | *bioF2* | C-terminal similar to *B. subtilis* BioF |
| Rv3279c | *birA* | biotin apo-protein ligase |
| Rv1442 | *bisC* | biotin sulfoxide reductase |
| Rv0089 | - | possible *bioC* biotin synthesis gene |

2. Folic acid

| | | |
|---|---|---|
| Rv2763c | *dfrA* | dihydrofolate reductase |
| Rv2447c | *folC* | folylpolyglutamate synthase |
| Rv3356c | *folD* | methylenetetrahydrofolate dehydrogenase |
| Rv3609c | *folE* | GTP cyclohydrolase I |
| Rv3606c | *folK* | 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase |
| Rv3608c | *folP* | dihydropteroate synthase |
| Rv1207 | *folP2* | dihydropteroate synthase |
| Rv3607c | *folX* | may be involved in folate biosynthesis |
| Rv0013 | *pabA* | *p*-aminobenzoate synthase glutamine amidotransferase |
| Rv1005c | *pabB* | *p*-aminobenzoate synthase |
| Rv0812 | *pabC* | aminodeoxychorismate lyase |

3. Lipoate

| | | |
|---|---|---|
| Rv2218 | *lipA* | lipoate biosynthesis protein A |
| Rv2217 | *lipB* | lipoate biosynthesis protein B |

4. Molybdopterin

| | | |
|---|---|---|
| Rv3109 | *moaA* | molybdenum cofactor biosynthesis, protein A |
| Rv0869c | *moaA2* | molybdenum cofactor biosynthesis, protein A |
| Rv0438c | *moaA3* | molybdenum cofactor biosynthesis, protein A |
| Rv3110 | *moaB* | molybdenum cofactor biosynthesis, protein B |
| Rv0984 | *moaB2* | molybdenum cofactor biosynthesis, protein B |
| Rv3111 | *moaC* | molybdenum cofactor biosynthesis, protein C |
| Rv0864 | *moaC2* | molybdenum cofactor biosynthesis, protein C |
| Rv3324c | *moaC3* | molybdenum cofactor biosynthesis, protein C |
| Rv3112 | *moaD* | molybdopterin converting factor subunit 1 |
| Rv0868c | *moaD2* | molybdopterin converting factor subunit 1 |
| Rv3119 | *moaE* | molybdopterin-converting factor subunit 2 |
| Rv0866 | *moaE2* | molybdopterin-converting factor subunit 2 |
| Rv3322c | *moaE3* | molybdopterin-converting factor subunit 2 |
| Rv0994 | *moeA* | molybdopterin biosynthesis |
| Rv3116 | *moeB* | molybdopterin biosynthesis |
| Rv2338c | *moeW* | molybdopterin biosynthesis |
| Rv1681 | *moeX* | weak similarity to *E. coli* MoaA |
| Rv1355c | *moeY* | weak similarity to *E. coli* MoeB |
| Rv3206c | *moeZ* | probably involved in molybdopterin biosynthesis |
| Rv0865 | *mog* | molybdopterin biosynthesis |

5. Pantothenate

| | | |
|---|---|---|
| Rv1092c | *coaA* | pantothenate kinase |
| Rv2225 | *panB* | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| Rv3602c | *panC* | pantoate-β-alanine ligase |
| Rv3601c | *panD* | aspartate 1-decarboxylase |

6. Pyridoxine

| | | |
|---|---|---|
| Rv2607 | *pdxH* | pyridoxamine 5'-phosphate oxidase |

7. Pyridine nucleotide

| | | |
|---|---|---|
| Rv1594 | *nadA* | quinolinate synthase |
| Rv1595 | *nadB* | L-aspartate oxidase |
| Rv1596 | *nadC* | nicotinate-nucleotide pyrophosphatase |
| Rv0423c | *thiC* | thiamine synthesis, pyrimidine moiety |

8. Thiamine

| | | |
|---|---|---|
| Rv0422c | *thiD* | phosphomethylpyrimidine kinase |
| Rv0414c | *thiE* | thiamine synthesis, thiazole moiety |
| Rv0417 | *thiG* | thiamine synthesis, thiazole moiety |
| Rv2977c | *thiL* | probable thiamine-monophosphate kinase |

9. Riboflavin

| | | |
|---|---|---|
| Rv1940 | *ribA* | GTP cyclohydrolase II |
| Rv1415 | *ribA2* | probable GTP cyclohydrolase II |
| Rv1412 | *ribC* | riboflavin synthase α chain |
| Rv2671 | *ribD* | probable riboflavin deaminase |
| Rv2786c | *ribF* | riboflavin kinase |
| Rv1409 | *ribG* | riboflavin biosynthesis |
| Rv1416 | *ribH* | riboflavin synthase β chain |
| Rv3300c | - | probable deaminase, riboflavin synthesis |

10. Thioredoxin, glutaredoxin and mycothiol

| | | |
|---|---|---|
| Rv0773c | *ggtA* | putative γ-glutamyl transpeptidase |
| Rv2394 | *ggtB* | γ -glutamyltranspeptidase precursor |
| Rv2855 | *gorA* | glutathione reductase homologue |
| Rv0816c | *thiX* | equivalent to *M. leprae* ThiX |
| Rv1470 | *trxA* | thioredoxin |
| Rv1471 | *trxB* | thioredoxin reductase |
| Rv3913 | *trxB2* | thioredoxin reductase |
| Rv3914 | *trxC* | thioredoxin |

11. Menaquinone, PQQ, ubiquinone and other terpenoids

| | | |
|---|---|---|
| Rv2682c | *dxs* | 1-deoxy-D-xylulose 5-phosphate synthase |
| Rv0562 | *grcC1* | heptaprenyl diphosphate synthase II |
| Rv0989c | *grcC2* | heptaprenyl diphosphate synthase II |
| Rv3398c | *idsA* | geranylgeranyl pyrophosphate synthase |
| Rv2173 | *idsA2* | geranylgeranyl pyrophosphate synthase |
| Rv3383c | *idsB* | transfergeranyl, similar geranyl pyrophosphate synthase |
| Rv0534c | *menA* | 4-dihydroxy-2-naphthoate octaprenyltransferase |
| Rv0548c | *menB* | naphthoate synthase |
| Rv0553 | *menC* | *o*-succinylbenzoate-CoA synthase |
| Rv0555 | *menD* | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| Rv0542c | *menE* | *o*-succinylbenzoic acid-CoA ligase |
| Rv3853 | *menG* | *S*-adenosylmethionine: 2-demethylmenaquinone |
| Rv3397c | *phyA* | phytoene synthase |
| Rv0693 | *pqqE* | coenzyme PQQ synthesis protein E |
| Rv0558 | *ubiE* | ubiquinone/menaquinone biosynthesis methyltransferase |

12. Heme and porphyrin

| | | |
|---|---|---|
| Rv0509 | *hemA* | glutamyl-tRNA reductase |
| Rv0512 | *hemB* | δ-aminolevulinic acid dehydratase |
| Rv0510 | *hemC* | porphobilinogen deaminase |
| Rv2678c | *hemE* | uroporphyrinogen decarboxylase |

| | | |
|---|---|---|
| Rv1300 | *hemK* | protoporphyrinogen oxidase |
| Rv0524 | *hemL* | glutamate-1-semialdehyde amino-transferase |
| Rv2388c | *hemN* | oxygen-independent coproporphyrinogen III oxidase |
| Rv2677c | *hemY'* | protoporphyrinogen oxidase |
| Rv1485 | *hemZ* | ferrochelatase |

13. Cobalamin

| | | |
|---|---|---|
| Rv2849c | *cobA* | cob(I)alamin adenosyltransferase |
| Rv2848c | *cobB* | cobyrinic acid a,c-diamide synthase |
| Rv2231c | *cobC* | aminotransferase |
| Rv2236c | *cobD* | cobinamide synthase |
| Rv2064 | *cobG* | percorrin reductase |
| Rv2065 | *cobH* | precorrin isomerase |
| Rv2066 | *cobI* | CobI-CobJ fusion protein |
| Rv2070c | *cobK* | precorrin reductase |
| Rv2072c | *cobL* | probable methyltransferase |
| Rv2071c | *cobM* | precorrin-3 methylase |
| Rv2062c | *cobN* | cobalt insertion |
| Rv2208 | *cobS* | cobalamin (5'-phosphate) synthase |
| Rv2207 | *cobT* | nicotinate-nucleotide-dimethyl-benzimidazole transferase |
| Rv0254c | *cobU* | cobinamide kinase |
| Rv0255c | *cobQ* | cobyric acid synthase |
| Rv3713 | *cobQ2* | possible cobyric acid synthase |
| Rv0306 | - | similar to BluB cobalamin synthesis protein *R. capsulatus* |

14. Iron utilization

| | | |
|---|---|---|
| Rv1876 | *bfrA* | bacterioferritin |
| Rv3841 | *bfrB* | bacterioferritin |
| Rv3215 | *entC* | probable isochorismate synthase |
| Rv3214 | *entD* | weak similarity to many phosphoglycerate mutases |
| Rv2895c | viuB | similar to proteins involved in vibriobactin uptake |
| Rv3525c | - | similar to ferripyochelin binding protein |

*H. Lipid biosynthesis*

1. Synthesis of fatty and mycolic acids

| | | |
|---|---|---|
| Rv3285 | *accA3* | acetyl/propionyl CoA carboxylase α subunit |
| Rv0904c | *accD3* | acetyl/propionyl CoA carboxylase β subunit |
| Rv3799c | *accD4* | acetyl/propionyl CoA carboxylase β subunit |
| Rv3280 | *accD5* | acetyl/propionyl CoA carboxylase β subunit |
| Rv2247 | *accD6* | acetyl/propionyl CoA carboxylase β subunit |
| Rv2244 | *acpM* | acyl carrier protein (meromycolate extension) |
| Rv2523c | *acpS* | CoA:apo-[ACP] pantethienephosphotransferase |
| Rv2243 | *fabD* | malonyl CoA-[ACP] transacylase |
| Rv0649 | *fabD2* | malonyl CoA-[ACP] transacylase |
| Rv1483 | *fabG1* | 3-oxoacyl-[ACP] reductase (aka MabA) |
| Rv1350 | *fabG2* | 3-oxoacyl-[ACP] Reductase |
| Rv2002 | *fabG3* | 3-oxoacyl-[ACP] reductase |
| Rv0242c | *fabG4* | 3-oxoacyl-[ACP] reductase |
| Rv2766c | *fabG5* | 3-oxoacyl-[ACP] reductase |
| Rv0533c | *fabH* | β-ketoacyl-ACP synthase III |
| Rv2524c | *fas* | fatty acid synthase |
| Rv1484 | *inhA* | enoyl-[ACP] reductase |
| Rv2245 | *kasA* | β-ketoacyl-ACP synthase (meromycolate extension) |
| Rv2246 | *kasB* | β-ketoacyl-ACP synthase (meromycolate extension) |
| Rv1618 | *tesB1* | thioesterase II |
| Rv2605c | *tesB2* | thioesterase II |
| Rv0033 | - | possible acyl carrier protein |
| Rv1344 | - | possible acyl carrier protein |
| Rv1722 | - | possible biotin carboxylase |
| Rv3221c | - | resembles biotin carboxyl carrier |
| Rv3472 | - | possible acyl carrier protein |

2. Modification of fatty and mycolic acids

| | | |
|---|---|---|
| Rv3391 | *acrA1* | fatty acyl-CoA reductase |
| Rv3392c | *cmaA1* | cyclopropane mycolic acid synthase 1 |
| Rv0503c | *cmaA2* | cyclopropane mycolic acid synthase 2 |
| Rv0824c | *desA1* | acyl-[ACP] desaturase |
| Rv1094 | *desA2* | acyl-[ACP] desaturase |
| Rv3229c | *desA3* | acyl-[ACP] desaturase |
| Rv0645c | *mmaA1* | methoxymycolic acid synthase 1 |
| Rv0644c | *mmaA2* | methoxymycolic acid synthase 2 |
| Rv0643c | *mmaA3* | methoxymycolic acid synthase 3 |
| Rv0642c | *mmaA4* | methoxymycolic acid synthase 4 |
| Rv0447c | *ufaA1* | unknown fatty acid methyltransferase |
| Rv3538 | *ufaA2* | unknown fatty acid methyltransferase |
| Rv0469 | *umaA1* | unknown mycolic acid methyltransferase |
| Rv0470c | *umaA2* | unknown mycolic acid methyltransferase |

3. Acyltransferases, mycoloyltransferases and phospholipid synthesis

| | | |
|---|---|---|
| Rv2289 | *cdh* | CDP-diacylglycerol phosphatidylhydrolase |
| Rv2881c | *cdsA* | phosphatidate cytidylyltransferase |
| Rv3804c | *fbpA* | antigen 85A, mycolyltransferase |
| Rv1886c | *fbpB* | antigen 85B, mycolyltransferase |
| Rv3803c | *fbpC1* | antigen 85C, mycolyltransferase |
| Rv0129c | *fbpC2* | antigen 85C', mycolyltransferase |
| Rv0564c | *gpdA1* | glycerol-3-phosphate dehydrogenase |
| Rv2982c | *gpdA2* | glycerol-3-phosphate dehydrogenase |
| Rv2612c | *pgsA* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv1822 | *pgsA2* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv2746c | *pgsA3* | CDP-diacylglycerol-glycerol-3-phosphate phosphatidyltransferase |
| Rv1551 | *plsB1* | glycerol-3-phosphate acyltransferase |
| Rv2482c | *plsB2* | glycerol-3-phosphate acyltransferase |
| Rv0437c | *psd* | putative phosphatidylserine decarboxylase |
| Rv0436c | *pssA* | CDP-diacylglycerol-serine *o*-phosphatidyltransferase |
| Rv0045c | - | possible dihydrolipoamide acetyltransferase |
| Rv0914c | - | lipid transfer protein |
| Rv1543 | - | probable fatty-acyl CoA reductase |
| Rv1627c | - | lipid carrier protein |
| Rv1814 | - | possible C-5 sterol desaturase |
| Rv1867 | - | similar to acetyl CoA synthase/lipid carriers |
| Rv2261c | - | apolipoprotein N-acyltransferase-a |
| Rv2262c | - | apolipoprotein N-acyltransferase-b |
| Rv3523 | - | lipid carrier protein |
| Rv3720 | - | C-term similar to cyclopropane fatty acid synthases |

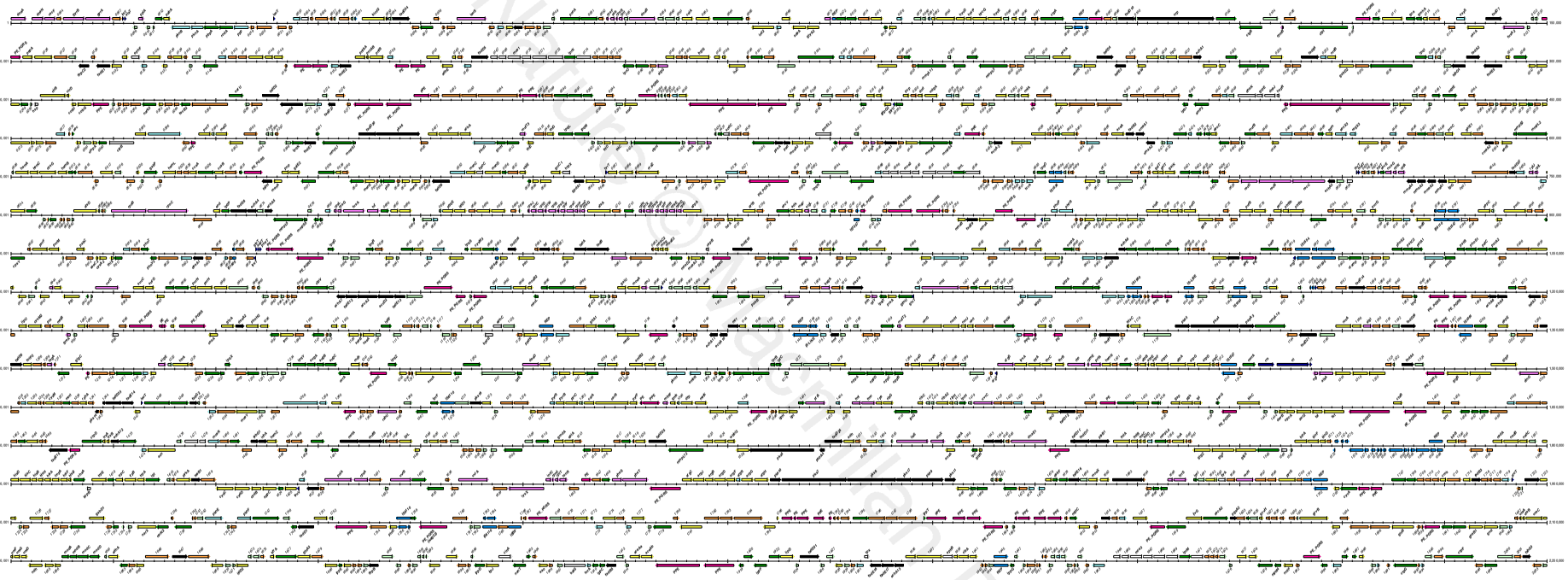
*I. Polyketide and non-ribosomal peptide synthesis*

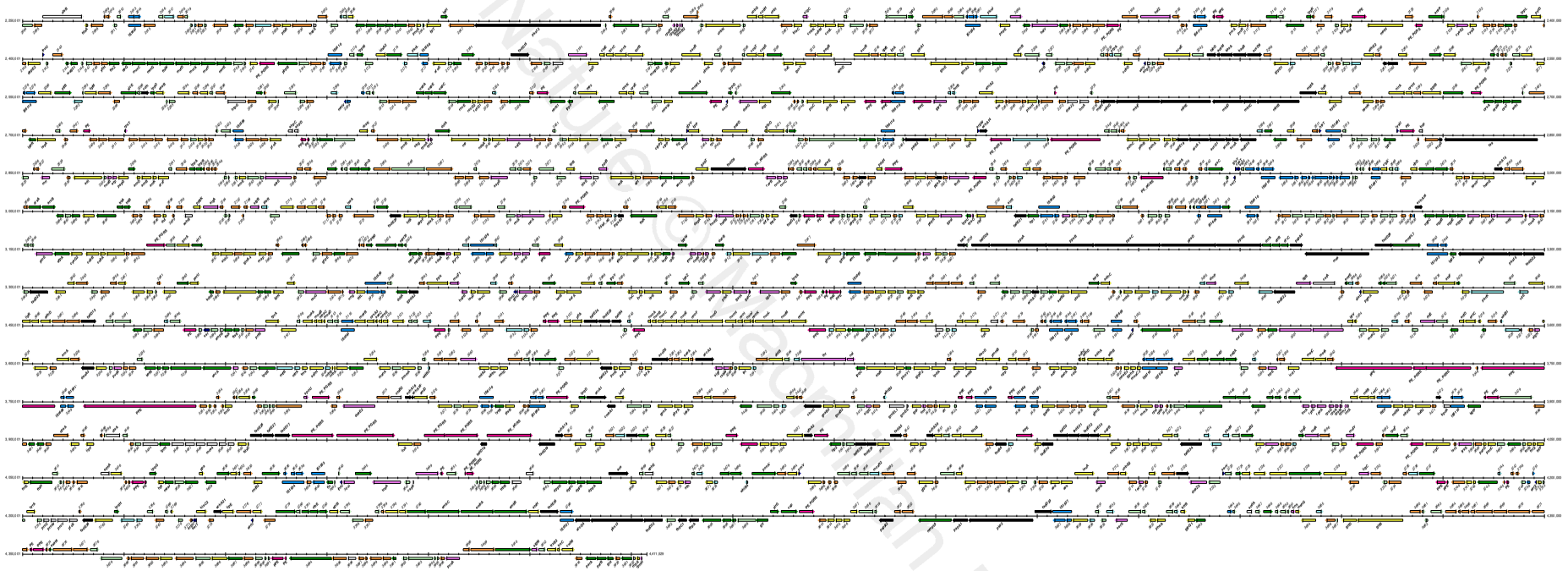
| | | |
|---|---|---|
| Rv2940c | *mas* | mycocerosic acid synthase |
| Rv2384 | *mbtA* | mycobactin/exochelin synthesis (salicylate-AMP ligase) |
| Rv2383c | *mbtB* | mycobactin/exochelin synthesis (serine/threonine ligation) |
| Rv2382c | *mbtC* | mycobactin/exochelin synthesis |
| Rv2381c | *mbtD* | mycobactin/exochelin synthesis (polyketide synthase) |
| Rv2380c | *mbtE* | mycobactin/exochelin synthesis (lysine ligation) |
| Rv2379c | *mbtF* | mycobactin/exochelin synthesis (lysine ligation) |
| Rv2378c | *mbtG* | mycobactin/exochelin synthesis (lysine hydroxylase) |
| Rv2377c | *mbtH* | mycobactin/exochelin synthesis |
| Rv0101 | *nrp* | unknown non-ribosomal peptide synthase |
| Rv1153c | *omt* | PKS *o*-methyltransferase |
| Rv3824c | *papA1* | PKS-associated protein, unknown function |
| Rv3820c | *papA2* | PKS-associated protein, unknown function |
| Rv1182 | *papA3* | PKS-associated protein, unknown function |
| Rv1528c | *papA4* | PKS-associated protein, unknown function |
| Rv2939 | *papA5* | PKS-associated protein, unknown function |
| Rv2946c | *pks1* | polyketide synthase |
| Rv1660 | *pks10* | polyketide synthase (chalcone synthase-like) |
| Rv1665 | *pks11* | polyketide synthase (chalcone synthase-like) |
| Rv2048c | *pks12* | polyketide synthase (erythronolide synthase-like) |
| Rv3800c | *pks13* | polyketide synthase |
| Rv1342c | *pks14* | polyketide synthase (chalcone synthase-like) |
| Rv2947c | *pks15* | polyketide synthase |
| Rv1013 | *pks16* | polyketide synthase |
| Rv1663 | *pks17* | polyketide synthase |
| Rv1372 | *pks18* | polyketide synthase |
| Rv3825c | *pks2* | polyketide synthase |
| Rv1180 | *pks3* | polyketide synthase |
| Rv1181 | *pks4* | polyketide synthase |
| Rv1527c | *pks5* | polyketide synthase |
| Rv0405 | *pks6* | polyketide synthase |
| Rv1661 | *pks7* | polyketide synthase |
| Rv1662 | *pks8* | polyketide synthase |
| Rv1664 | *pks9* | polyketide synthase |
| Rv2931 | *ppsA* | phenolpthiocerol synthesis (*pksB*) |
| Rv2932 | *ppsB* | phenolpthiocerol synthesis (*pksC*) |
| Rv2933 | *ppsC* | phenolpthiocerol synthesis (*pksD*) |
| Rv2934 | *ppsD* | phenolpthiocerol synthesis (*pksE*) |
| Rv2935 | *ppsE* | phenolpthiocerol synthesis (*pksF*) |
| Rv2928 | *tesA* | thioesterase |
| Rv1544 | - | probable ketoacyl reductase |

*J. Broad regulatory functions*

1. Repressors/activators

| | | |
|---|---|---|
| Rv1657 | *argR* | arginine repressor |
| Rv1267c | *embR* | regulator of *embAB* genes (AfsR/DndI/RedD family) |
| Rv1909c | *furA* | ferric uptake regulatory protein |
| Rv2359 | *furB* | ferric uptake regulatory protein |
| Rv2919c | *glnB* | nitrogen regulatory protein |
| Rv2711 | *ideR* | iron dependent repressor, IdeR |
| Rv2720 | *lexA* | LexA, SOS repressor protein |
| Rv1479 | *moxR* | transcriptional regulator, MoxR homologue |
| Rv3692 | *moxR2* | transcriptional regulator, MoxR homologue |
| Rv3164c | *moxR3* | transcriptional regulator, MoxR homologue |
| Rv0212c | *nadR* | similar to *E.coli* NadR |
| Rv0117 | *oxyS* | transcriptional regulator (LysR family) |
| Rv1379 | *pyrR* | regulatory protein pyrimidine biosynthesis |
| Rv2788 | *sirR* | iron-dependent transcriptional repressor |
| Rv3082c | *virS* | putative virulence regulating protein (AraC/XylS family) |
| Rv3219 | *whiB1* | WhiB transcriptional activator homologue |
| Rv3260c | *whiB2* | WhiB transcriptional activator homologue |
| Rv3416 | *whiB3* | WhiB transcriptional activator homologue |
| Rv3681c | *whiB4* | WhiB transcriptional activator homologue |
| Rv0023 | - | putative transcriptional regulator |
| Rv0043c | - | transcriptional regulator (GntR family) |
| Rv0067c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0078 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0081 | - | transcriptional regulator (ArsR family) |
| Rv0135c | - | putative transcriptional regulator |
| Rv0144 | - | putative transcriptional regulator |
| Rv0158 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0165c | - | transcriptional regulator (GntR family) |
| Rv0195 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0196 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0232 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0238 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0273c | - | putative transcriptional regulator |
| Rv0302 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0324 | - | putative transcriptional regulator |
| Rv0328 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0348 | - | putative transcriptional regulator |
| Rv0377 | - | transcriptional regulator (LysR family) |
| Rv0386 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0452 | - | putative transcriptional regulator |
| Rv0465c | - | transcriptional regulator (PbsX/Xre family) |
| Rv0472c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0474 | - | transcriptional regulator (PbsX/Xre family) |
| Rv0485 | - | transcriptional regulator (ROK family) |
| Rv0494 | - | transcriptional regulator (GntR family) |
| Rv0552 | - | putative transcriptional regulator |
| Rv0576 | - | putative transcriptional regulator |
| Rv0586 | - | transcriptional regulator (GntR family) |
| Rv0650 | - | transcriptional regulator (ROK family) |
| Rv0653c | - | putative transcriptional regulator |
| Rv0681 | - | transcriptional regulator (TetR/AcrR family) |
| Rv0691c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0737 | - | putative transcriptional regulator |
| Rv0744c | - | putative transcriptional regulator |
| Rv0792c | - | transcriptional regulator (GntR |





| | | |
|---|---|---|
| | | family) |
| Rv0823c | - | transcriptional regulator (NifR3/Smm1 family) |
| Rv0827c | - | transcriptional regulator (ArsR family) |
| Rv0890c | - | transcriptional regulator (LuxR/UhpA family) |
| Rv0891c | - | putative transcriptional regulator |
| Rv0894 | - | putative transcriptional regulator |
| Rv1019 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1049 | - | transcriptional regulator (MarR family) |
| Rv1129c | - | transcriptional regulator (PbsX/Xre family) |
| Rv1151c | - | putative transcriptional regulator |
| Rv1152 | - | transcriptional regulator (GntR family) |
| Rv1167c | - | putative transcriptional regulator |
| Rv1219c | - | putative transcriptional regulator |
| Rv1255c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1332 | - | putative transcriptional regulator |
| Rv1353c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1358 | - | transcriptional regulator (LuxR/UhpA family) |
| Rv1359 | - | putative transcriptional regulator |
| Rv1395 | - | transcriptional regulator (AraC/XylS family) |
| Rv1404 | - | transcriptional regulator (MarR family) |
| Rv1423 | - | putative transcriptional regulator |
| Rv1460 | - | putative transcriptional regulator |
| Rv1474c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1534 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1556 | - | putative transcriptional regulator |
| Rv1674c | - | putative transcriptional regulator |
| Rv1675c | - | putative transcriptional regulator |
| Rv1719 | - | transcriptional regulator (IclR family) |
| Rv1773c | - | transcriptional regulator (IclR family) |
| Rv1776c | - | putative transcriptional regulator |
| Rv1816 | - | putative transcriptional regulator |
| Rv1846c | - | putative transcriptional regulator |
| Rv1931c | - | transcriptional regulator (AraC/XylS family) |
| Rv1956 | - | putative transcriptional regulator |
| Rv1963c | - | putative transcriptional regulator |
| Rv1985c | - | transcriptional regulator (LysR family) |
| Rv1990c | - | putative transcriptional regulator |
| Rv1994c | - | transcriptional regulator (MerR family) |
| Rv2017 | - | putative transcriptional regulator (PbsX/Xre family) |
| Rv2021c | - | putative transcriptional regulator |
| Rv2034 | - | transcriptional regulator (ArsR family) |
| Rv2175c | - | putative transcriptional regulator |
| Rv2250c | - | putative transcriptional regulator |
| Rv2258c | - | putative transcriptional regulator |
| Rv2282c | - | transcriptional regulator (LysR family) |
| Rv2308 | - | putative transcriptional regulator |
| Rv2324 | - | transcriptional regulator (Lrp/AsnC family) |
| Rv2358 | - | transcriptional regulator (ArsR family) |
| Rv2488c | - | transcriptional regulator (LuxR/UhpA family) |
| Rv2506 | - | transcriptional regulator (TetR/AcrR family) |
| Rv2621c | - | putative transcriptional regulator |
| Rv2640c | - | transcriptional regulator (ArsR family) |
| Rv2642 | - | transcriptional regulator (ArsR family) |
| Rv2669 | - | putative transcriptional regulator |
| Rv2745c | - | putative transcriptional regulator |
| Rv2779c | - | transcriptional regulator (Lrp/AsnC family) |
| Rv2887 | - | transcriptional regulator (MarR family) |
| Rv2912c | - | transcriptional regulator (TetR/AcrR family) |
| Rv2989 | - | transcriptional regulator (IclR family) |
| Rv3050c | - | putative transcriptional regulator |
| Rv3055 | - | putative transcriptional regulator |
| Rv3058c | - | putative transcriptional regulator |
| Rv3060c | - | transcriptional regulator (GntR family) |
| Rv3066 | - | putative transcriptional regulator |
| Rv3095 | - | putative transcriptional regulator |
| Rv3124 | - | transcriptional regulator (AfsR/DndI/RedD family) |
| Rv3160c | - | putative transcriptional regulator |
| Rv3167c | - | putative transcriptional regulator |
| Rv3173c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3183 | - | putative transcriptional regulator |
| Rv3208 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3249c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3291c | - | transcriptional regulator (Lrp/AsnC family) |
| Rv3295 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3334 | - | transcriptional regulator (MerR family) |
| Rv3405c | - | putative transcriptional regulator |
| Rv3522 | - | putative transcriptional regulator |
| Rv3557c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3574 | - | transcriptional regulator (TetR/AcrR family) |
| Rv3575c | - | transcriptional regulator (LacI family) |
| Rv3583c | - | putative transcriptional regulator |
| Rv3676 | - | transcriptional regulator (Crp/Fnr family) |
| Rv3678c | - | transcriptional regulator (LysR family) |
| Rv3736 | - | transcriptional regulator (AraC/XylS family) |
| Rv3744 | - | transcriptional regulator (ArsR family) |
| Rv3830c | - | transcriptional regulator (TetR/AcrR family) |
| Rv3833 | - | transcriptional regulator (AraC/XylS family) |
| Rv3840 | - | putative transcriptional regulator |
| Rv3855 | - | putative transcriptional regulator |

2. Two component systems

| | | |
|---|---|---|
| Rv1028c | *kdpD* | sensor histidine kinase |
| Rv1027c | *kdpE* | two-component response regulator |
| Rv3246c | *mtrA* | two-component response regulator |
| Rv3245c | *mtrB* | sensor histidine kinase |
| Rv0844c | *narL* | two-component response regulator |
| Rv0757 | *phoP* | two-component response regulator |
| Rv0758 | *phoR* | sensor histidine kinase |
| Rv0491 | *regX3* | two-component response regulator |
| Rv0490 | *senX3* | sensor histidine kinase |
| Rv0602c | *tcrA* | two-component response regulator |
| Rv0260c | - | two-component response regulator |
| Rv0600c | - | sensor histidine kinase |
| Rv0601c | - | sensor histidine kinase |
| Rv0818 | - | two-component response regulator |
| Rv0845 | - | sensor histidine kinase |
| Rv0902c | - | sensor histidine kinase |
| Rv0903c | - | two-component response regulator |
| Rv0981 | - | two-component response regulator |
| Rv0982 | - | sensor histidine kinase |
| Rv1032c | - | sensor histidine kinase |
| Rv1033c | - | two-component response regulator |
| Rv1626 | - | two-component response regulator |
| Rv2027c | - | sensor histidine kinase |
| Rv2884 | - | two-component response regulator |
| Rv3132c | - | sensor histidine kinase |
| Rv3133c | - | two-component response regulator |
| Rv3143 | - | putative sensory transduction protein |
| Rv3220c | - | sensor histidine kinase |
| Rv3764c | - | sensor histidine kinase |
| Rv3765c | - | two-component response regulator |

3. Serine-threonine protein kinases and phosphoprotein phosphatases

| | | |
|---|---|---|
| Rv0015c | *pknA* | serine-threonine protein kinase |
| Rv0014c | *pknB* | serine-threonine protein kinase |
| Rv0931c | *pknD* | serine-threonine protein kinase |
| Rv1743 | *pknE* | serine-threonine protein kinase |
| Rv1746 | *pknF* | serine-threonine protein kinase |
| Rv0410c | *pknG* | serine-threonine protein kinase |
| Rv1266c | *pknH* | serine-threonine protein kinase |
| Rv2914c | *pknI* | serine-threonine protein kinase |
| Rv2088 | *pknJ* | serine-threonine protein kinase |
| Rv3080c | *pknK* | serine-threonine protein kinase |
| Rv2176 | *pknL* | serine-threonine protein kinase, truncated |
| Rv0018c | *ppp* | putative phosphoprotein phosphatase |
| Rv2234 | *ptpA* | low molecular weight protein-tyrosine-phosphatase |
| Rv0153c | - | putative protein-tyrosine-phosphatase |

**II. Macromolecule metabolism**

*A. Synthesis and modification of macromolecules*

1. Ribosomal protein synthesis and modification

| | | |
|---|---|---|
| Rv3420c | *rimI* | ribosomal protein S18 acetyl transferase |
| Rv0995 | *rimJ* | acetylation of 30S S5 subunit |
| Rv0641 | *rplA* | 50S ribosomal protein L1 |
| Rv0704 | *rplB* | 50S ribosomal protein L2 |
| Rv0701 | *rplC* | 50S ribosomal protein L3 |
| Rv0702 | *rplD* | 50S ribosomal protein L4 |
| Rv0716 | *rplE* | 50S ribosomal protein L5 |
| Rv0719 | *rplF* | 50S ribosomal protein L6 |
| Rv0056 | *rplI* | 50S ribosomal protein L9 |
| Rv0651 | *rplJ* | 50S ribosomal protein L10 |
| Rv0640 | *rplK* | 50S ribosomal protein L11 |
| Rv0652 | *rplL* | 50S ribosomal protein L7/L12 |
| Rv3443c | *rplM* | 50S ribosomal protein L13 |
| Rv0714 | *rplN* | 50S ribosomal protein L14 |
| Rv0723 | *rplO* | 50S ribosomal protein L15 |
| Rv0708 | *rplP* | 50S ribosomal protein L16 |
| Rv3456c | *rplQ* | 50S ribosomal protein L17 |
| Rv0720 | *rplR* | 50S ribosomal protein L18 |
| Rv2904c | *rplS* | 50S ribosomal protein L19 |
| Rv1643 | *rplT* | 50S ribosomal protein L20 |
| Rv2442c | *rplU* | 50S ribosomal protein L21 |
| Rv0706 | *rplV* | 50S ribosomal protein L22 |
| Rv0703 | *rplW* | 50S ribosomal protein L23 |
| Rv0715 | *rplX* | 50S ribosomal protein L24 |
| Rv1015c | *rplY* | 50S ribosomal protein L25 |
| Rv2441c | *rpmA* | 50S ribosomal protein L27 |
| Rv0105c | *rpmB* | 50S ribosomal protein L28 |
| Rv2058c | *rpmB2* | 50S ribosomal protein L28 |
| Rv0709 | *rpmC* | 50S ribosomal protein L29 |
| Rv0722 | *rpmD* | 50S ribosomal protein L30 |
| Rv1298 | *rpmE* | 50S ribosomal protein L31 |
| Rv2057c | *rpmG* | 50S ribosomal protein L33 |
| Rv3924c | *rpmH* | 50S ribosomal protein L34 |
| Rv1642 | *rpmI* | 50S ribosomal protein L35 |
| Rv3461c | *rpmJ* | 50S ribosomal protein L36 |
| Rv1630 | *rpsA* | 30S ribosomal protein S1 |
| Rv2890c | *rpsB* | 30S ribosomal protein S2 |
| Rv0707 | *rpsC* | 30S ribosomal protein S3 |
| Rv3458c | *rpsD* | 30S ribosomal protein S4 |
| Rv0721 | *rpsE* | 30S ribosomal protein S5 |
| Rv0053 | *rpsF* | 30S ribosomal protein S6 |
| Rv0683 | *rpsG* | 30S ribosomal protein S7 |
| Rv0718 | *rpsH* | 30S ribosomal protein S8 |
| Rv3442c | *rpsI* | 30S ribosomal protein S9 |
| Rv0700 | *rpsJ* | 30S ribosomal protein S10 |
| Rv3459c | *rpsK* | 30S ribosomal protein S11 |
| Rv0682 | *rpsL* | 30S ribosomal protein S12 |
| Rv3460c | *rpsM* | 30S ribosomal protein S13 |
| Rv0717 | *rpsN* | 30S ribosomal protein S14 |
| Rv2056c | *rpsN2* | 30S ribosomal protein S14 |
| Rv2785c | *rpsO* | 30S ribosomal protein S15 |
| Rv2909c | *rpsP* | 30S ribosomal protein S16 |
| Rv0710 | *rpsQ* | 30S ribosomal protein S17 |
| Rv0055 | *rpsR* | 30S ribosomal protein S18 |
| Rv2055c | *rpsR2* | 30S ribosomal protein S18 |
| Rv0705 | *rpsS* | 30S ribosomal protein S19 |
| Rv2412 | *rpsT* | 30S ribosomal protein S20 |
| Rv3241c | - | member of S30AE ribosomal protein family |

2. Ribosome modification and maturation

| | | |
|---|---|---|
| Rv1010 | *ksgA* | 16S rRNA dimethyltransferase |
| Rv2838c | *rbfA* | ribosome-binding factor A |
| Rv2907c | *rimM* | 16S rRNA processing protein |

3. Aminoacyl tRNA synthases and their modification

| | | |
|---|---|---|
| Rv2555c | *alaS* | alanyl-tRNA synthase |
| Rv1292 | argS | arginyl-tRNA synthase |
| Rv2572c | *aspS* | aspartyl-tRNA synthase |
| Rv3580c | *cysS* | cysteinyl-tRNA synthase |
| Rv2130c | *cysS2* | cysteinyl-tRNA synthase |
| Rv1406 | *fmt* | methionyl-tRNA formyltransferase |
| Rv3011c | *gatA* | glu-tRNA-gln amidotransferase, subunit B |
| Rv3009c | *gatB* | glu-tRNA-gln amidotransferase, subunit A |
| Rv3012c | *gatC* | glu-tRNA-gln amidotransferase, subunit C |
| Rv2992c | *gltS* | glutamyl-tRNA synthase |
| Rv2357c | *glyS* | glycyl-tRNA synthase |
| Rv2580c | *hisS* | histidyl-tRNA synthase |
| Rv1536 | *ileS* | isoleucyl-tRNA synthase |
| Rv0041 | *leuS* | leucyl-tRNA synthase |
| Rv3598c | *lysS* | lysyl-tRNA synthase |
| Rv1640c | *lysX* | C-term lysyl-tRNA synthase |
| Rv1007c | *metS* | methionyl-tRNA synthase |
| Rv1649 | *pheS* | phenylalanyl-tRNA synthase α subunit |

| | | |
|---|---|---|
| Rv1650 | *pheT* | phenylalanyl-tRNA synthase β subunit |
| Rv2845c | *proS* | prolyl-tRNA synthase |
| Rv3834c | *serS* | seryl-tRNA synthase |
| Rv2614c | *thrS* | threonyl-tRNA synthase |
| Rv2906c | *trmD* | tRNA (guanine-N1)-methyltransferase |
| Rv3336c | *trpS* | tryptophanyl tRNA synthase |
| Rv1689 | *tyrS* | tyrosyl-tRNA synthase |
| Rv2448c | *valS* | valyl-tRNA synthase |

4. Nucleoproteins

| | | |
|---|---|---|
| Rv1407 | *fmu* | similar to Fmu protein |
| Rv3852 | *hns* | HU-histone protein |
| Rv2986c | *hupB* | DNA-binding protein II |
| Rv1388 | *mIHF* | integration host factor |

5. DNA replication, repair, recombination and restriction/modification

| | | |
|---|---|---|
| Rv1317c | *alkA* | DNA-3-methyladenine glycosidase II |
| Rv2836c | *dinF* | DNA-damage-inducible protein F |
| Rv1329c | *dinG* | probable ATP-dependent helicase |
| Rv3056 | *dinP* | DNA-damage-inducible protein |
| Rv1537 | *dinX* | probable DNA-damage-inducible protein |
| Rv0001 | *dnaA* | chromosomal replication initiator protein |
| Rv0058 | *dnaB* | DNA helicase (contains intein) |
| Rv1547 | *dnaE1* | DNA polymerase III, α subunit |
| Rv3370c | *dnaE2* | DNA polymerase III α chain |
| Rv2343c | *dnaG* | DNA primase |
| Rv0002 | *dnaN* | DNA polymerase III, β subunit |
| Rv3711c | *dnaQ* | DNA polymerase III ε chain |
| Rv3721c | *dnaZX* | DNA polymerase III, γ (dnaZ) and τ (dnaX) |
| Rv2924c | *fpg* | formamidopyrimidine-DNA glycosylase |
| Rv0006 | *gyrA* | DNA gyrase subunit A |
| Rv0005 | *gyrB* | DNA gyrase subunit B |
| Rv2092c | *helY* | probable helicase, Ski2 subfamily |
| Rv2101 | *helZ* | probable helicase, Snf2/Rad54 family |
| Rv2756c | *hsdM* | type I restriction/modification system DNA methylase |
| Rv2755c | *hsdS'* | type I restriction/modification system specificity determinant |
| Rv3296 | *lhr* | ATP-dependent helicase |
| Rv3014c | *ligA* | DNA ligase |
| Rv3062 | *ligB* | DNA ligase |
| Rv3731 | *ligC* | probable DNA ligase |
| Rv1020 | *mfd* | transcription-repair coupling factor |
| Rv2528c | *mrr* | restriction system protein |
| Rv2985 | *mutT1* | MutT homologue |
| Rv1160 | *mutT2* | MutT homologue |
| Rv0413 | *mutT3* | MutT homologue |
| Rv3589 | *mutY* | probable DNA glycosylase |
| Rv3297 | *nei* | probable endonuclease VIII |
| Rv3674c | *nth* | probable endonuclease III |
| Rv1316c | *ogt* | methylated-DNA-protein-cysteine methyltransferase |
| Rv1629 | *polA* | DNA polymerase I |
| Rv1402 | *priA* | putative primosomal protein n' (replication factor Y) |
| Rv3585 | *radA* | probable DNA repair RadA homologue |
| Rv2737c | *recA* | recombinase (contains intein) |
| Rv0630c | *recB* | exodeoxyribonuclease V |
| Rv0631c | *recC* | exodeoxyribonuclease V |
| Rv0629c | *recD* | exodeoxyribonuclease V |
| Rv0003 | *recF* | DNA replication and SOS induction |
| Rv2973c | *recG* | ATP-dependent DNA helicase |
| Rv1696 | *recN* | recombination and DNA repair |
| Rv3715c | *recR* | RecBC-Independent process of DNA repair |
| Rv2736c | *recX* | regulatory protein for RecA |
| Rv2593c | *ruvA* | Holliday junction binding protein, DNA helicase |
| Rv2592c | *ruvB* | Holliday junction binding protein |
| Rv2594c | *ruvC* | Holliday junction resolvase, endodeoxyribonuclease |
| Rv0054 | *ssb* | single strand binding protein |
| Rv1210 | *tagA* | DNA-3-methyladenine glycosidase I |
| Rv3646c | *topA* | DNA topoisomerase |
| Rv2976c | *ung* | uracil-DNA glycosylase |
| Rv1638 | *uvrA* | excinuclease ABC subunit A |
| Rv1633 | uvrB | excinuclease ABC subunit B |
| Rv1420 | uvrC | excinuclease ABC subunit C |
| Rv0949 | uvrD | DNA-dependent ATPase I and helicase II |
| Rv3198c | uvrD2 | putative UvrD |
| Rv0427c | xthA | exodeoxyribonuclease III |
| Rv0071 | - | group II intron maturase |
| Rv0861c | - | probable DNA helicase |
| Rv0944 | - | possible formamidopyrimidine-DNA glycosylase |
| Rv1688 | - | probable 3-methylpurine DNA glycosylase |
| Rv2090 | - | partially similar to DNA polymerase I |
| Rv2191 | - | similar to both PolC and UvrC proteins |
| Rv2464c | - | probable DNA glycosylase, endonuclease VIII |
| Rv3201c | - | probable ATP-dependent DNA helicase |
| Rv3202c | - | similar to UvrD proteins |
| Rv3263 | - | probable DNA methylase |
| Rv3644c | - | similar in N-term to DNA polymerase III |

6. Protein translation and modification

| | | |
|---|---|---|
| Rv0429c | *def* | polypeptide deformylase |
| Rv2534c | *efp* | elongation factor P |
| Rv2882c | *frr* | ribosome recycling factor |
| Rv0684 | *fusA* | elongation factor G |
| Rv0120c | *fusA2* | elongation factor G |
| Rv1080c | *greA* | transcription elongation factor G |
| Rv3462c | *infA* | initiation factor IF-1 |
| Rv2839c | *infB* | initiation factor IF-2 |
| Rv1641 | *infC* | initiation factor IF-3 |
| Rv0009 | *ppiA* | peptidyl-prolyl *cis-trans* isomerase |
| Rv2582 | *ppiB* | peptidyl-prolyl *cis-trans* isomerase |
| Rv1299 | *prfA* | peptide chain release factor 1 |
| Rv3105c | *prfB* | peptide chain release factor 2 |
| Rv2889c | *tsf* | elongation factor EF-Ts |
| Rv0685 | tuf | elongation factor EF-Tu |

7. RNA synthesis, RNA modification and DNA transcription

| | | |
|---|---|---|
| Rv1253 | *deaD* | ATP-dependent DNA/RNA helicase |
| Rv2783c | *gpsI* | pppGpp synthase and polyribonucleotide phosphorylase |
| Rv2841c | *nusA* | transcription termination factor |
| Rv2533c | *nusB* | N-utilization substance protein B |
| Rv0639 | *nusG* | transcription antitermination protein |
| Rv3907c | *pcnA* | polynucleotide polymerase |
| Rv3232c | *pvdS* | alternative sigma factor for siderophore production |
| Rv3211 | *rhlE* | probable ATP-dependent RNA helicase |
| Rv1297 | *rho* | transcription termination factor rho |
| Rv3457c | *rpoA* | α subunit of RNA polymerase |
| Rv0667 | *rpoB* | β subunit of RNA polymerase |
| Rv0668 | *rpoC* | β' subunit of RNA polymerase |
| Rv1364c | *rsbU* | SigB regulation protein |
| Rv3287c | *rsbW* | anti-sigma B factor |
| Rv2703 | *sigA* | RNA polymerase sigma factor (aka MysA, RpoV) |
| Rv2710 | *sigB* | RNA polymerase sigma factor (aka MysB) |
| Rv2069 | *sigC* | ECF subfamily sigma subunit |
| Rv3414c | *sigD* | ECF subfamily sigma subunit |
| Rv1221 | *sigE* | ECF subfamily sigma subunit |
| Rv3286c | *sigF* | ECF subfamily sigma subunit |
| Rv0182c | *sigG* | sigma-70 factors ECF subfamily |
| Rv3223c | *sigH* | ECF subfamily sigma subunit |
| Rv1189 | *sigI* | ECF family sigma factor |
| Rv3328c | *sigJ* | similar to SigI, ECF family |
| Rv0445c | *sigK* | ECF-type sigma factor |
| Rv0735 | *sigL* | sigma-70 factors ECF subfamily |
| Rv3911 | *sigM* | probable sigma factor, similar to SigE |
| Rv3366 | *spoU* | probable rRNA methylase |
| Rv3455c | *truA* | probable pseudouridylate synthase |
| Rv2793c | *truB* | tRNA pseudouridine 55 synthase |
| Rv1644 | *tsnR* | putative 23S rRNA methyltransferase |
| Rv3649 | - | ATP-dependent DNA/RNA helicase |

8. Polysaccharides (cytoplasmic)

| | | |
|---|---|---|
| Rv1326c | *glgB* | 1,4-α-glucan branching enzyme |
| Rv1328 | *glgP* | probable glycogen phosphorylase |
| Rv1564c | *glgX* | probable glycogen debranching enzyme |
| Rv1563c | *glgY* | putative α-amylase |
| Rv1562c | *glgZ* | maltooligosyltrehalose trehalohydrolase |
| Rv0126 | - | probable glycosyl hydrolase |
| Rv1781c | - | probable 4-α-glucanotransferase |
| Rv2471 | - | probable maltase α-glucosidase |

*B. Degradation of macromolecules*

1. RNA

| | | |
|---|---|---|
| Rv1014c | *pth* | peptidyl-tRNA hydrolase |
| Rv2925c | *rnc* | RNAse III |
| Rv2444c | *rne* | similar at C-term to ribonuclease E |
| Rv2902c | *rnhB* | ribonuclease HII |
| Rv3923c | *rnpA* | ribonuclease P protein component |
| Rv1340 | *rphA* | ribonuclease PH |

2. DNA

| | | |
|---|---|---|
| Rv0670 | *end* | endonuclease IV (apurinase) |
| Rv1108c | *xseA* | exonuclease VII large subunit |
| Rv1107c | *xseB* | exonuclease VII small subunit |

3. Proteins, peptides and glycopeptides

| | | |
|---|---|---|
| Rv3305c | *amiA* | probable aminohydrolase |
| Rv3306c | *amiB* | probable aminohydrolase |
| Rv3596c | *clpC* | ATP-dependent Clp protease |
| Rv2461c | *clpP* | ATP-dependent Clp protease proteolytic subunit |
| Rv2460c | *clpP2* | ATP-dependent Clp protease proteolytic subunit |
| Rv2457c | *clpX* | ATP-dependent Clp protease ATP-binding subunit ClpX |
| Rv2667 | *clpX'* | similar to ClpC from *M. leprae* but shorter |
| Rv3419c | *gcp* | glycoprotease |
| Rv2725c | *hflX* | GTP-binding protein |
| Rv1223 | *htrA* | serine protease |
| Rv2861c | *map* | methionine aminopeptidase |
| Rv0734 | *map'* | probable methionine aminopeptidase |
| Rv0319 | *pcp* | pyrrolidone-carboxylate peptidase |
| Rv0125 | *pepA* | probable serine protease |
| Rv2213 | *pepB* | aminopeptidase A/I |
| Rv0800 | *pepC* | aminopeptidase I |
| Rv2467 | *pepD* | probable aminopeptidase |
| Rv2089c | *pepE* | cytoplasmic peptidase |
| Rv2535c | *pepQ* | cytoplasmic peptidase |
| Rv2782c | *pepR* | protease/peptidase, M16 family (insulinase) |
| Rv2109c | *prcA* | proteasome α-type subunit 1 |
| Rv2110c | *prcB* | proteasome β-type subunit 2 |
| Rv0782 | *ptrBa* | protease II, α subunit |
| Rv0781 | *ptrBb* | protease II, β subunit |
| Rv0724 | *sppA* | protease IV, signal peptide peptidase |
| Rv0198c | - | probable zinc metalloprotease |
| Rv0457c | - | probable peptidase |
| Rv0840c | - | probable proline iminopeptidase |
| Rv0983 | - | probable serine protease |
| Rv1977 | - | probable zinc metallopeptidase |
| Rv3668c | - | probable alkaline serine protease |
| Rv3671c | - | probable serine protease |
| Rv3883c | - | probable secreted protease |
| Rv3886c | - | protease |

4. Polysaccharides, lipopolysaccharides and phospholipids

| | | |
|---|---|---|
| Rv0062 | *celA* | cellulase/endoglucanase |
| Rv3915 | *cwlM* | hydrolase |
| Rv0315 | - | probable β-1,3-glucanase |
| Rv1090 | - | probable inactivated cellulase/endoglucanase |
| Rv1327c | - | probable glycosyl hydrolase, α-amylase family |
| Rv1333 | - | probable hydrolase |
| Rv3463 | - | probable neuraminidase |
| Rv3717 | - | possible N-acetylmuramoyl-L-alanine amidase |

5. Esterases and lipases

| | | |
|---|---|---|
| Rv0220 | *lipC* | probable esterase |
| Rv1923 | *lipD* | probable esterase |
| Rv3775 | *lipE* | probable hydrolase |
| Rv3487c | *lipF* | probable esterase |
| Rv0646c | *lipG* | probable hydrolase |
| Rv1399c | *lipH* | probable lipase |
| Rv1400c | *lipI* | probable lipase |
| Rv1900c | *lipJ* | probable esterase |
| Rv2385 | *lipK* | probable acetyl-hydrolase |
| Rv1497 | *lipL* | esterase |
| Rv2284 | *lipM* | probable esterase |
| Rv2970c | *lipN* | probable lipase/esterase |
| Rv1426c | *lipO* | probable esterase |
| Rv2463 | *lipP* | probable esterase |
| Rv2485c | *lipQ* | probable carboxylesterase |
| Rv3084 | *lipR* | probable acetyl-hydrolase |
| Rv3176c | *lipS* | probable esterase/lipase |
| Rv2045c | *lipT* | probable carboxylesterase |
| Rv1076 | *lipU* | probable esterase |
| Rv3203 | *lipV* | probable lipase |
| Rv0217c | *lipW* | probable esterase |
| Rv2351c | *plcA* | phospholipase C precursor |
| Rv2350c | *plcB* | phospholipase C precursor |
| Rv2349c | *plcC* | phospholipase C precursor |
| Rv1755c | *plcD* | partial CDS for phospholipase C |
| Rv1104 | - | probable esterase pseudogene |
| Rv1105 | - | probable esterase pseudogene |

6. Aromatic hydrocarbons

| | | |
|---|---|---|
| Rv3469c | *mhpE* | probable 4-hydroxy-2-oxovalerate aldolase |
| Rv0316 | - | probable muconolactone isomerase |
| Rv0771 | - | probable 4-carboxymuconolactone decarboxylase |
| Rv0939 | - | probable dehydrase |
| Rv1723 | - | 6-aminohexanoate-dimer hydro- |

| | | |
|---|---|---|
| | | lase |
| Rv2715 | - | 2-hydroxymuconic semialdehyde hydrolase |
| Rv3530c | - | probable *cis*-diol dehydrogenase |
| Rv3534c | - | 4-hydroxy-2-oxovalerate aldolase |
| Rv3536c | - | aromatic hydrocarbon degradation |

*C. Cell envelope*

1. Lipoproteins (*lppA-lpr0*) 65

2. Surface polysaccharides, lipopolysaccharides, proteins and antigens

| | | |
|---|---|---|
| Rv0806c | *cpsY* | probable UDP-glucose-4-epimerase |
| Rv3811 | *csp* | secreted protein |
| Rv1677 | *dsbF* | highly similar to C-term Mpt53 |
| Rv3794 | *embA* | involved in arabinogalactan synthesis |
| Rv3795 | *embB* | involved in arabinogalactan synthesis |
| Rv3793 | *embC* | involved in arabinogalactan synthesis |
| Rv3875 | *esat6* | early secretory antigen target |
| Rv0112 | *gca* | probable GDP-mannose dehydratase |
| Rv0113 | *gmhA* | phosphoheptose isomerase |
| Rv2965c | *kdtB* | lipopolysaccharide core biosynthesis protein |
| Rv2878c | *mpt53* | secreted protein Mpt53 |
| Rv1980c | *mpt64* | secreted immunogenic protein Mpb64/Mpt64 |
| Rv2875 | *mpt70* | major secreted immunogenic protein Mpt70 precursor |
| Rv2873 | *mpt83* | surface lipoprotein Mpt83 |
| Rv0899 | *ompA* | member of OmpA family |
| Rv3810 | *pirG* | cell surface protein precursor (Erp protein) |
| Rv3782 | *rfbE* | similar to rhamnosyl transferase |
| Rv1302 | *rfe* | undecaprenyl-phosphate α-N-acetylglucosaminyltransferase |
| Rv2145c | *wag31* | antigen 84 (aka wag31) |
| Rv0431 | - | tuberculin related peptide (AT103) |
| Rv0954 | - | cell envelope antigen |
| Rv1514c | - | involved in polysaccharide synthesis |
| Rv1518 | - | involved in exopolysaccharide synthesis |
| Rv1758 | - | partial cutinase |
| Rv1910c | - | probable secreted protein |
| Rv1919c | - | weak similarity to pollen antigens |
| Rv1984c | - | probable secreted protein |
| Rv1987 | - | probable secreted protein |
| Rv2223c | - | probable exported protease |
| Rv2224c | - | probable exported protease |
| Rv2301 | - | probable cutinase |
| Rv2345 | - | precursor of probable membrane protein |
| Rv2672 | - | putative exported protease |
| Rv3019c | - | similar to Esat6 |
| Rv3036c | - | probable secreted protein |
| Rv3449 | - | probable precursor of serine protease |
| Rv3451 | - | probable cutinase |
| Rv3452 | - | probable cutinase precursor |
| Rv3724 | - | probable cutinase precursor |

3. Murein sacculus and peptidoglycan

| | | |
|---|---|---|
| Rv2911 | *dacB* | penicillin binding protein |
| Rv2981c | *ddlA* | D-alanine-D-alanine ligase A |
| Rv3809c | *glf* | UDP-galactopyranose mutase |
| Rv1018c | *glmU* | UDP-N-acetylglucosamine pyrophosphorylase |
| Rv3382c | *lytB* | LytB protein homologue |
| Rv1110 | *lytB'* | very similar to LytB |
| Rv1315 | *murA* | UDP-N-acetylglucosamine-1-carboxyvinyltransferase |
| Rv0482 | *murB* | UDP-N-acetylenolpyruvoylglucosamine reductase |
| Rv2152c | *murC* | UDP-N-acetyl-muramate-alanine ligase |
| Rv2155c | *murD* | UDP-N-acetylmuramoylalanine-D-glutamate ligase |
| Rv2158c | *murE* | *meso*-diaminopimelate-adding enzyme |
| Rv2157c | *murF* | D-alanine:D-alanine-adding enzyme |
| Rv2153c | *murG* | transferase in peptidoglycan synthesis |
| Rv1338 | *murI* | glutamate racemase |
| Rv2156c | *murX* | phospho-N-acetylmuramoyl-petapeptide transferase |
| Rv3332 | *nagA* | N-acetylglucosamine-6-P-deacetylase |
| Rv0016c | *pbpA* | penicillin-binding protein |
| Rv2163c | *pbpB* | penicillin-binding protein 2 |
| Rv0050 | *ponA* | penicillin-bonding protein |
| Rv3682 | *ponA'* | class A penicillin binding protein |
| Rv0017c | *rodA* | FtsW/RodA/SpovE family |
| Rv0907 | - | probable penicillin binding protein |
| Rv1367c | - | probable penicillin binding protein |
| Rv1730c | - | probable penicillin binding protein |
| Rv1922 | - | probable penicillin binding protein |
| Rv2864c | - | probable penicillin binding protein |
| Rv3330 | - | probable penicillin binding protein |
| Rv3627c | - | probable penicillin binding protein |

4. Conserved membrane proteins

| | | |
|---|---|---|
| Rv0402c | *mmpL1* | conserved large membrane protein |
| Rv0507 | *mmpL2* | conserved large membrane protein |
| Rv0206c | *mmpL3* | conserved large membrane protein |
| Rv0450c | *mmpL4* | conserved large membrane protein |
| Rv0676c | *mmpL5* | conserved large membrane protein |
| Rv1557 | *mmpL6* | conserved large membrane protein |
| Rv2942 | *mmpL7* | conserved large membrane protein |
| Rv3823c | *mmpL8* | conserved large membrane protein |
| Rv2339 | *mmpL9* | conserved large membrane protein |
| Rv1183 | *mmpL10* | conserved large membrane protein |
| Rv0202c | *mmpL11* | conserved large membrane protein |
| Rv1522c | *mmpL12* | conserved large membrane protein |
| Rv0403c | *mmpS1* | conserved small membrane protein |
| Rv0506 | *mmpS2* | conserved small membrane protein |
| Rv2198c | *mmpS3* | conserved small membrane protein |
| Rv0451c | *mmpS4* | conserved small membrane protein |
| Rv0677c | *mmpS5* | conserved small membrane protein |

5. Other membrane proteins 211

**III. Cell processes**

*A. Transport/binding proteins*

1. Amino acids

| | | |
|---|---|---|
| Rv2127 | *ansP* | L-asparagine permease |
| Rv0346c | *aroP2* | probable aromatic amino acid permease |
| Rv0917 | *betP* | glycine betaine transport |
| Rv1704c | *cycA* | transport of D-alanine, D-serine and glycine |
| Rv3666c | *dppA* | probable peptide transport system permease |
| Rv3665c | *dppB* | probable peptide transport system permease |
| Rv3664c | *dppC* | probable peptide transport system permease |
| Rv3663c | *dppD* | probable ABC-transporter |
| Rv0522 | *gabP* | probable 4-amino butyrate transporter |
| Rv0411c | *glnH* | putative glutamine binding protein |
| Rv2564 | *glnQ* | probable ATP-binding transport protein |
| Rv1280c | *oppA* | probable oligopeptide transport protein |
| Rv1283c | *oppB* | oligopeptide transport protein |
| Rv1282c | *oppC* | oligopeptide transport system permease |
| Rv1281c | *oppD* | probable peptide transport protein |
| Rv2320c | *rocE* | arginine/ornithine transporter |
| Rv3253c | - | probable cationic amino acid transport |
| Rv3454 | - | possible proline permease |

2. Cations

| | | |
|---|---|---|
| Rv2920c | *amt* | putative ammonium transporter |
| Rv1607 | *chaA* | putative calcium/proton antiporter |
| Rv1239c | *corA* | probable magnesium and cobalt transport protein |
| Rv0092 | *ctpA* | cation-transporting ATPase |
| Rv0103c | *ctpB* | cation transport ATPase |
| Rv3270 | *ctpC* | cation transport ATPase |
| Rv1469 | *ctpD* | probable cadmium-transporting ATPase |
| Rv0908 | *ctpE* | probable cation transport ATPase |
| Rv1997 | *ctpF* | probable cation transport ATPase |
| Rv1992c | *ctpG* | probable cation transport ATPase |
| Rv0425c | *ctpH* | C-terminal region putative cation-transporting ATPase |
| Rv0107c | *ctpI* | probable magnesium transport ATPase |
| Rv0969 | *ctpV* | cation transport ATPase |
| Rv3044 | *fecB* | putative FeIII-dicitrate transporter |
| Rv0265c | *fecB2* | iron transport protein FeIII dicitrate transporter |
| Rv1029 | *kdpA* | potassium-transporting ATPase A chain |
| Rv1030 | *kdpB* | potassium-transporting ATPase B chain |
| Rv1031 | *kdpC* | potassium-transporting ATPase C chain |
| Rv3236c | *kefB* | probable glutathione-regulated potassium-efflux protein |
| Rv2877c | *merT* | possible mercury resistance transport system |
| Rv1811 | *mgtC* | probable magnesium transport ATPase protein C |
| Rv0362 | *mgtE* | putative magnesium ion transporter |
| Rv2856 | *nicT* | probable nickel transport protein |
| Rv0924c | *nramp* | transmembrane protein belonging to Nramp family |
| Rv2691 | *trkA* | probable potassium uptake protein |
| Rv2692 | *trkB* | probable potassium uptake protein |
| Rv2287 | *yjcE* | probable Na+/H+ exchanger |
| Rv2723 | - | probable membrane protein, tellurium resistance |
| Rv3162c | - | probable membrane protein |
| Rv3237c | - | possible potassium channel protein |
| Rv3743c | - | probable cation-transporting ATPase |

3. Carbohydrates, organic acids and alcohols

| | | |
|---|---|---|
| Rv2443 | *dctA* | C4-dicarboxylate transport protein |
| Rv3476c | *kgtP* | sugar transport protein |
| Rv1902c | *nanT* | probable sialic acid transporter |
| Rv1236 | *sugA* | membrane protein probably involved in sugar transport |
| Rv1237 | *sugB* | sugar transport protein |
| Rv1238 | *sugC* | ABC transporter component of sugar uptake system |
| Rv3331 | *sugI* | probable sugar transport protein |
| Rv2835c | *ugpA* | *sn*-glycerol-3-phosphate permease |
| Rv2833c | *ugpB* | *sn*-glycerol-3-phosphate-binding periplasmic lipoprotein |
| Rv2832c | *ugpC* | *sn*-glycerol-3-phosphate transport ATP-binding protein |
| Rv2834c | *ugpE* | *sn*-glycerol-3-phosphate transport system protein |
| Rv2316 | *uspA* | sugar transport protein |
| Rv2318 | *uspC* | sugar transport protein |
| Rv2317 | *uspE* | sugar transport protein |
| Rv1200 | - | probable sugar transporter |
| Rv2038c | - | probable ABC sugar transporter |
| Rv2039c | - | probable sugar transporter |
| Rv2040c | - | probable sugar transporter |
| Rv2041c | - | probable sugar transporter |

4. Anions

| | | |
|---|---|---|
| Rv2684 | *arsA* | probable arsenical pump |
| Rv2685 | *arsB* | probable arsenical pump |
| Rv3578 | *arsB2* | probable arsenical pump |
| Rv2643 | *arsC* | probable arsenical pump |
| Rv2397c | *cysA* | sulphate transport ATP-binding protein |
| Rv2399c | *cysT* | sulphate transport system permease protein |
| Rv2398c | *cysW* | sulphate transport system permease protein |
| Rv1857 | *modA* | molybdate binding protein |
| Rv1858 | *modB* | transport system permease, molybdate uptake |
| Rv1859 | *modC* | molybdate uptake ABC-transporter |
| Rv1860 | *modD* | precursor of Apa (45/47 kD secreted protein) |
| Rv2329c | *narK1* | probable nitrite extrusion protein |
| Rv1737c | *narK2* | nitrite extrusion protein |
| Rv0261c | *narK3* | nitrite extrusion protein1 |
| Rv0267 | *narU* | similar to nitrite extrusion protein 2 |
| Rv0934 | *phoS1* | PstS component of phosphate uptake |
| Rv0928 | *phoS2* | PstS component of phosphate uptake |
| Rv0820 | *phoT* | phosphate transport system ABC transporter |
| Rv3301c | *phoY1* | phosphate transport system regulator |
| Rv0821c | *phoY2* | phosphate transport system regulator |
| Rv0545c | *pitA* | low-affinity inorganic phosphate transporter |
| Rv2281 | *pitB* | phosphate permease |
| Rv0930 | *pstA1* | PstA component of phosphate uptake |
| Rv0936 | *pstA2* | PstA component of phosphate uptake |
| Rv0933 | *pstB* | ABC transport component of phosphate uptake |
| Rv0935 | *pstC* | PstC component of phosphate uptake |
| Rv0929 | *pstC2* | membrane-bound component of |

| | | |
|---|---|---|
| | | phosphate transport system |
| Rv0932c | *pstS* | PstS component of phosphate uptake |
| Rv2400c | *subI* | sulphate binding precursor |
| Rv0143c | - | probable chloride channel |
| Rv1707 | - | probable sulphate permease |
| Rv1739c | - | possible sulphate transporter |
| Rv3679 | - | possible anion transporter |
| Rv3680 | - | probable anion transporter |

5. Fatty acid transport

| | | |
|---|---|---|
| Rv2790c | *ltp1* | non-specific lipid transport protein |
| Rv3540c | *ltp2* | non-specific lipid transport protein |

6. Efflux proteins

| | | |
|---|---|---|
| Rv2936 | *drrA* | similar daunorubicin resistance ABC-transporter |
| Rv2937 | *drrB* | similar daunorubicin resistance transmembrane protein |
| Rv2938 | *drrC* | similar daunorubicin resistance transmembrane protein |
| Rv2846c | *efpA* | putative efflux protein |
| Rv3065 | *emrE* | resistance to ethidium bromide |
| Rv0783c | - | multidrug resistance protein |
| Rv0849 | - | possible quinolone efflux pump |
| Rv1145 | - | probable drug transporter |
| Rv1146 | - | probable drug transporter |
| Rv1250 | - | probable drug efflux protein |
| Rv1258c | - | probable multidrug resistance pump |
| Rv1410c | - | probable drug efflux protein |
| Rv1634 | - | probable drug efflux protein |
| Rv1819c | - | probable multidrug resistance pump |
| Rv2136c | - | putative bacitracin resistance protein |
| Rv2209 | - | probable drug efflux protein |
| Rv2333c | - | probable tetracenomycin C resistance protein |
| Rv2994 | - | probable fluoroquinolone efflux protein |
| Rv1877 | - | probable drug efflux protein |
| Rv2459 | - | probable drug efflux protein |

*B. Chaperones/Heat shock*

| | | |
|---|---|---|
| Rv0384c | *clpB* | heat shock protein |
| Rv0352 | *dnaJ* | acts with GrpE to stimulate DnaK ATPase |
| Rv2373c | *dnaJ2* | DnaJ homologue |
| Rv0350 | *dnaK* | 70 kD heat shock protein, chromosome replication |
| Rv3417c | *groEL1* | 60 kD chaperonin 1 |
| Rv0440 | *groEL2* | 60 kD chaperonin 2 |
| Rv3418c | *groES* | 10 kD chaperone |
| Rv0351 | *grpE* | stimulates DnaK ATPase activity |
| Rv2374c | *hrcA* | heat-inducible transcription repressor |
| Rv0251c | *hsp* | possible heat shock protein |
| Rv0353 | *hspR* | heat shock regulator |
| Rv2031c | *hspX* | 14kD antigen, heat shock protein Hsp20 family |
| Rv2299c | *htpG* | heat shock protein Hsp90 family |
| Rv0563 | *htpX* | probable (transmembrane) heat shock protein |
| Rv2701c | *suhB* | putative extragenic suppressor protein |
| Rv3269 | - | probable heat shock protein |

*C. Cell division*

| | | |
|---|---|---|
| Rv3641c | *fic* | possible cell division protein |
| Rv3102c | *ftsE* | membrane protein |
| Rv3610c | *ftsH* | inner membrane protein, chaperone |
| Rv2748c | *ftsK* | chromosome partitioning |
| Rv2151c | *ftsQ* | ingrowth of wall at septum |
| Rv2154c | *ftsW* | membrane protein (shape determination) |
| Rv3101c | *ftsX* | membrane protein |
| Rv2921c | *ftsY* | cell division protein FtsY |
| Rv2150c | *ftsZ* | circumferential ring, GTPase |
| Rv3919c | *gid* | glucose inhibited division protein B |
| Rv3625c | *mesJ* | probable cell cycle protein |
| Rv3917c | *parA* | chromosome partitioning; DNA - binding |
| Rv3918c | *parB* | possibly involved in chromosome partitioning |
| Rv2922c | *smc* | member of Smc1/Cut3/Cut14 family |
| Rv0012 | - | possible cell division protein |
| Rv0435c | - | ATPase of AAA-family |
| Rv2115c | - | ATPase of AAA-family |
| Rv3213c | - | possible role in chromosome segregation |
| Rv1708 | - | possible role in chromosome partitioning |

*D. Protein and peptide secretion*

| | | |
|---|---|---|
| Rv2916c | *ffh* | signal recognition particle protein |
| Rv2903c | *lepB* | signal peptidase I |
| Rv1614 | *lgt* | prolipoprotein diacylglyceryl transferase |
| Rv1539 | *lspA* | lipoprotein signal peptidase |
| Rv0379 | *sec* | probable transport protein SecE/Sec61- $\gamma$ family |
| Rv3240c | *secA* | SecA, preprotein translocase subunit |
| Rv1821 | *secA2* | SecA, preprotein translocase subunit |
| Rv2587c | *secD* | protein-export membrane protein |
| Rv0638 | *secE* | SecE preprotein translocase |
| Rv2586c | *secF* | protein-export membrane protein |
| Rv1440 | *secG* | protein-export membrane protein SecG |
| Rv0732 | *secY* | SecY subunit of preprotein translocase |
| Rv2462c | *tig* | chaperone protein, similar to trigger factor |
| Rv2813 | - | probable general secretion pathway protein |

*E. Adaptations and atypical conditions*

| | | |
|---|---|---|
| Rv1901 | *cinA* | competence damage protein |
| Rv3648c | *cspA* | cold shock protein, transcriptional regulator |
| Rv0871 | *cspB* | probable cold shock protein |
| Rv3063 | *cstA* | starvation-induced stress response protein |
| Rv3490 | *otsA* | probable $\alpha,\alpha$-trehalose-phosphate synthase |
| Rv2006 | *otsB* | trehalose-6-phosphate phosphatase |
| Rv3372 | *otsB2* | trehalose-6-phosphate phosphatase |
| Rv3758c | *proV* | osmoprotection ABC transporter |
| Rv3757c | *proW* | transport system permease |
| Rv3759c | *proX* | similar to osmoprotection proteins |
| Rv3756c | *proZ* | transport system permease |
| Rv1026 | - | probable pppGpp-5'phosphohydrolase |

*F. Detoxification*

| | | |
|---|---|---|
| Rv2428 | *ahpC* | alkyl hydroperoxide reductase |
| Rv2429 | *ahpD* | member of AhpC/TSA family |
| Rv2238c | *ahpE* | member of AhpC/TSA family |
| Rv2521 | *bcp* | bacterioferritin comigratory protein |
| Rv1608c | *bcpB* | probable bacterioferritin comigratory protein |
| Rv3473c | *bpoA* | probable non-heme bromoperoxidase |
| Rv1123c | *bpoB* | probable non-heme bromoperoxidase |
| Rv0554 | *bpoC* | probable non-heme bromoperoxidase |
| Rv3617 | *ephA* | probable epoxide hydrolase |
| Rv1938 | *ephB* | probable epoxide hydrolase |
| Rv1124 | *ephC* | probable epoxide hydrolase |
| Rv2214c | *ephD* | probable epoxide hydrolase |
| Rv3670 | *ephE* | probable epoxide hydrolase |
| Rv0134 | *ephF* | probable epoxide hydrolase |
| Rv3171c | *hpx* | probable non-heme haloperoxidase |
| Rv1908c | *katG* | catalase-peroxidase |
| Rv3846 | *sodA* | superoxide dismutase |
| Rv0432 | *sodC* | superoxide dismutase precursor - (Cu-Zn) |
| Rv1932 | *tpx* | thiol peroxidase |
| Rv0634c | - | putative glyoxylase II |
| Rv2581c | - | putative glyoxylase II |
| Rv3177 | - | probable non-heme haloperoxidase |

IV. Other

*A. Virulence*

| | | |
|---|---|---|
| Rv0169 | *mce1* | cell invasion protein |
| Rv0589 | *mce2* | cell invasion protein |
| Rv1966 | *mce3* | cell invasion protein |
| Rv3499c | *mce4* | cell invasion protein |
| Rv3100c | *smpB* | probable small protein b |
| Rv1694 | *tlyA* | cytotoxin/hemolysin homologue |
| Rv0024 | - | putative p60 homologue |
| Rv0167 | - | part of *mce1* operon |
| Rv0168 | - | part of *mce1* operon |
| Rv0170 | - | part of *mce1* operon |
| Rv0171 | - | part of *mce1* operon |
| Rv0172 | - | part of *mce1* operon |
| Rv0174 | - | part of *mce1* operon |
| Rv0587 | - | part of *mce2* operon |
| Rv0588 | - | part of *mce2* operon |
| Rv0590 | - | part of *mce2* operon |
| Rv0591 | - | part of *mce2* operon |
| Rv0592 | - | part of *mce2* operon |
| Rv0594 | - | part of *mce2* operon |
| Rv1085c | - | possible hemolysin |
| Rv1477 | - | putative exported p60 protein homologue |
| Rv1478 | - | putative exported p60 protein homologue |
| Rv1566c | - | putative exported p60 protein homologue |
| Rv1964 | - | part of *mce3* operon |
| Rv1965 | - | part of *mce3* operon |
| Rv1967 | - | part of *mce3* operon |
| Rv1968 | - | part of *mce3* operon |
| Rv1969 | - | part of *mce3* operon |
| Rv1971 | - | part of *mce3* operon |
| Rv2190c | - | putative p60 homologue |
| Rv3494c | - | part of *mce4* operon |
| Rv3496c | - | part of *mce4* operon |
| Rv3497c | - | part of *mce4* operon |
| Rv3498c | - | part of *mce4* operon |
| Rv3500c | - | part of *mce4* operon |
| Rv3501c | - | part of *mce4* operon |
| Rv3896c | - | putative p60 homologue |
| Rv3922c | - | possible hemolysin |

*B. IS elements, Repeated sequences, and Phage*

1. IS elements

| | |
|---|---|
| IS*6110* | 16 copies |
| IS*1081* | 6 copies |
| Others | 37 copies |

2. REP*13E12* family 7 copies

3. Phage-related functions

| | | |
|---|---|---|
| Rv2894c | *xerC* | integrase/recombinase |
| Rv1701 | *xerD* | integrase/recombinase |
| Rv1054 | - | integrase-a |
| Rv1055 | - | integrase-b |
| Rv1573 | - | phiRV1 phage related protein |
| Rv1574 | - | phiRV1 phage related protein |
| Rv1575 | - | phiRV1 phage related protein |
| Rv1576c | - | phiRV1 phage related protein |
| Rv1577c | - | phiRV1 possible prohead protease |
| Rv1578c | - | phiRV1 phage related protein |
| Rv1579c | - | phiRV1 phage related protein |
| Rv1580c | - | phiRV1 phage related protein |
| Rv1581c | - | phiRV1 phage related protein |
| Rv1582c | - | phiRV1 phage related protein |
| Rv1583c | - | phiRV1 phage related protein |
| Rv1584c | - | phiRV1 phage related protein |
| Rv1585c | - | phiRV1 phage related protein |
| Rv1586c | - | phiRV1 integrase |
| Rv2309c | - | integrase |
| Rv2310 | - | excisionase |
| Rv2646 | - | phiRV2 integrase |
| Rv2647 | - | phiRV2 phage related protein |
| Rv2650c | - | phiRV2 phage related protein |
| Rv2651c | - | phiRV2 prohead protease |
| Rv2652c | - | phiRV2 phage related protein |
| Rv2653c | - | phiRV2 phage related protein |
| Rv2654c | - | phiRV2 phage related protein |
| Rv2655c | - | phiRV2 phage related protein |
| Rv2656c | - | phiRV2 phage related protein |
| Rv2657c | - | similar to gp36 of mycobacteriophage L5 |
| Rv2658c | - | phiRV2 phage related protein |
| Rv2659c | - | phiRV2 integrase |
| Rv2830c | - | similar to phage P1 *phd* gene |
| Rv3750c | - | excisionase |
| Rv3751 | - | putative integrase |

*C. PE and PPE families*

1. PE family

| | |
|---|---|
| PE subfamily | 38 members |
| PE_PGRS subfamily | 61 members |

2. PPE family 68 members

*D. Antibiotic production and resistance*

| | | |
|---|---|---|
| Rv2068c | *blaC* | class A $\beta$-lactamase |
| Rv3290c | *lat* | lysine-$\epsilon$ aminotransferase |
| Rv2043c | *pncA* | pyrazinamide resistance/sensitivity |
| Rv0133 | - | possible puromycin N-acetyltransferase |
| Rv0262c | - | aminoglycoside 2'-N-acetyltransferase |
| Rv0802c | - | acetyltransferase |
| Rv1082 | - | similar to *S. lincolnensis lmbE* |
| Rv1170 | - | similar to *S. lincolnensis lmbE* |
| Rv1347c | - | possible aminoglycoside 6'-N-acetyltransferase |
| Rv2036 | - | similar to lincomycin production genes |
| Rv2303c | - | similar to *S. griseus* macrotetrolide resistance protein |
| Rv3225c | - | probable aminoglycoside 3'-phosphotransferases |
| Rv3700c | - | probable acetyltransferase |
| Rv3817 | - | probable aminoglycoside 3'-phosphotransferase |

| | |
|---|---|
| *E. Bacteriocin-like proteins* | 3 |
| *F. Cytochrome P450 enzymes* | 22 |
| *G. Coenzyme F420-dependent enzymes* | 3 |
| *H. Miscellaneous transferases* | 61 |
| *I. Miscellaneous phosphatases, lyases, and hydrolases* | 18 |
| *J. Cyclases* | 6 |
| *K. Chelatases* | 2 |
| *V. Conserved hypotheticals* | 912 |
| *VI. Unknowns* | 606 |
| TOTAL | 3924 |