

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 . . . . . GSGAPGGAGGAAGLWGTGGAGGAGSSAGGGGAGGAGGAGGWLGDGGAGGIGGAST . . .
. . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
BCG . . . . . GSGAPGGAGGAAGLWTGGA-----GGAGGWLGDGGAGGIGGAST . . .
```



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-fucose 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* phosphoglucosyltransferase or phosphomannosyltransferase
 Rv0118c *oxcA* oxalyl-CoA decarboxylase
 Rv3068c *pgmA* phosphoglucosyltransferase
 Rv3257c *pmmA* phosphomannosyltransferase
 Rv3308 *pmmB* phosphomannosyltransferase
 Rv2702 *ppgK* polyphosphate glucokinase
 Rv0408 *pta* phosphate acetyltransferase
 Rv0729 *xyiB* 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* methylmalonate semialdehyde dehydrogenase
 Rv0751c *mmsB* methylmalonate 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 glutamate 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 *echH*)
 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

Rv2831 *echA16* superfamily enoyl-CoA hydratase/isomerase
 Rv3039c *echA17* superfamily enoyl-CoA hydratase/isomerase
 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* dihydroliipoamide dehydrogenase
 Rv2497c *pdhA* pyruvate dehydrogenase E1 component α subunit
 Rv2496c *pdhB* pyruvate dehydrogenase E1 component β subunit
 Rv2495c *pdhC* dihydroliipoamide acetyltransferase
 Rv0462 - probable dihydroliipoamide dehydrogenase

3. TCA cycle

Rv1475c *acon* aconitate hydratase
 Rv0889c *citA* citrate synthase 2
 Rv2498c *citE* citrate lyase β chain
 Rv1098c *fum* fumarase
 Rv1131 *glfA1* citrate synthase 3
 Rv0896 *glfA2* citrate synthase 1
 Rv3339c *icd1* isocitrate dehydrogenase
 Rv0066c *icd2* isocitrate dehydrogenase
 Rv0794c *lpdB* dihydroliipoamide 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* dihydroliipoamide 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 *rhsK* ribokinase
 Rv1408 *rpe* ribulose-phosphate 3-epimerase
 Rv2465c *rpi* phosphopentose isomerase
 Rv1448c *tal* transaldolase
 Rv1449c *tkl* 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-hydroxyarylamino *o*-acetyltransferase
 Rv0155 *pntAA* pyridine transhydrogenase subunit $\alpha 1$
 Rv0156 *pntAB* pyridine transhydrogenase subunit $\alpha 2$
 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-dehydrothiamine 3,5-epimerase
 Rv3266c *rmlD* dTDP-4-dehydrothiamine 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* probable arylsulfatase
 Rv0663 *atsD* probable arylsulfatase
 Rv3077 *atsF* probable arylsulfatase
 Rv0296c *atsG* probable arylsulfatase
 Rv3796 *atsH* probable 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* probable glutamine synthase
 Rv2918c *glnD* uridylyltransferase
 Rv2221c *glnE* glutamate-ammonia-ligase
 Rv3859c *gltB* adenylyltransferase
 Rv3858c *gltD* ferredoxin-dependent glutamate synthase
 Rv3704c *gshA* small subunit of NADH-dependent glutamate synthase
 Rv2427c *proA* possible γ -glutamylcysteine synthase
 Rv2439c *proB* γ -glutamyl phosphate reductase
 Rv0500 *proC* glutamate 5-kinase
 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-methyltetrahydropteroylglutamate-homocysteine methyltransferase
 Rv2124c *meth* 5-methyltetrahydrofolate-homocysteine methyltransferase
 Rv1392 *metK* S-adenosylmethionine synthase
 Rv0391 *metZ* *o*-succinylhomoserine sulfurydrylase
 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 *o*-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 | <i>hisF</i> | imidazole glycerol-phosphate synthase | Rv3048c | <i>nrdG</i> | subunit ribonucleoside-diphosphate small subunit | Rv3119 | <i>moaE</i> | subunit 1 molybdopterin-converting factor subunit 2 |
| Rv2121c | <i>hisG</i> | ATP phosphoribosyltransferase | Rv3053c | <i>nrdH</i> | glutaredoxin electron transport component of NrdEF system | Rv0866 | <i>moaE2</i> | molybdopterin-converting factor subunit 2 |
| Rv1602 | <i>hisH</i> | amidotransferase | Rv3052c | <i>nrdI</i> | NrdI/YgaO/YmaA family | Rv3322c | <i>moaE3</i> | molybdopterin-converting factor subunit 2 |
| Rv2122c | <i>hisI</i> | phosphoribosyl-AMP cyclohydrolase | Rv3247c | <i>tmk</i> | thymidylate kinase | Rv0994 | <i>moeA</i> | molybdopterin biosynthesis |
| Rv1606 | <i>hisI2</i> | probable phosphoribosyl-AMP 1,6 cyclohydrolase | Rv2764c | <i>thyA</i> | thymidylate synthase | Rv3116 | <i>moeB</i> | molybdopterin biosynthesis |
| Rv0114 | - | similar to HisB | Rv0570 | <i>nrdZ</i> | ribonucleotide reductase, class II | Rv2338c | <i>moeW</i> | molybdopterin biosynthesis |
| 6. Pyruvate family | | | Rv3752c | - | probable cytidine/deoxycytidylate deaminase | Rv1681 | <i>moeX</i> | weak similarity to <i>E. coli</i> MoeA |
| Rv3423c | <i>alr</i> | alanine racemase | 4. Salvage of nucleosides and nucleotides | | | Rv1355c | <i>moeY</i> | weak similarity to <i>E. coli</i> MoeB |
| 7. Branched amino acid family | | | Rv3313c | <i>add</i> | probable adenosine deaminase | Rv3206c | <i>moeZ</i> | probably involved in molybdopterin biosynthesis |
| Rv1559 | <i>ilvA</i> | threonine deaminase | Rv2584c | <i>apt</i> | adenine phosphoribosyltransferases | Rv0865 | <i>mog</i> | molybdopterin biosynthesis |
| Rv3003c | <i>ilvB</i> | acetolactate synthase I large subunit | Rv3315c | <i>cdd</i> | probable cytidine deaminase | 5. Pantothenate | | |
| Rv3470c | <i>ilvB2</i> | acetolactate synthase large subunit | Rv3314c | <i>deoA</i> | thymidine phosphorylase | Rv1092c | <i>coaA</i> | pantothenate kinase |
| Rv3001c | <i>ilvC</i> | ketol-acid reductoisomerase | Rv0478 | <i>deoC</i> | deoxyribose-phosphate aldolase | Rv2225 | <i>panB</i> | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| Rv0189c | <i>ilvD</i> | dihydroxy-acid dehydratase | Rv3307 | <i>deoD</i> | probable purine nucleoside phosphorylase | Rv3602c | <i>panC</i> | pantoate- β -alanine ligase |
| Rv2210c | <i>ilvE</i> | branched-chain-amino-acid transaminase | Rv3624c | <i>hpt</i> | probable hypoxanthine-guanine phosphoribosyltransferase | Rv3601c | <i>panD</i> | aspartate 1-decarboxylase |
| Rv1820 | <i>ilvG</i> | acetolactate synthase II | Rv3393 | <i>iunH</i> | probable inosine-uridine preferring nucleoside hydrolase | 6. Pyridoxine | | |
| Rv3002c | <i>ilvN</i> | acetolactate synthase I small subunit | Rv0535 | <i>pnp</i> | phosphorylase from Pnp/MtaP family 2 | Rv2607 | <i>pxdH</i> | pyridoxamine 5'-phosphate oxidase |
| Rv3509c | <i>ilvX</i> | probable acetohydroxyacid synthase I large subunit | Rv3309c | <i>upp</i> | uracil phosphoribosyltransferase | 7. Pyridine nucleotide | | |
| Rv3710 | <i>leuA</i> | α -isopropyl malate synthase | 5. Miscellaneous nucleoside/nucleotide reactions | | | Rv1594 | <i>nadA</i> | quinolinate synthase |
| Rv2995c | <i>leuB</i> | 3-isopropylmalate dehydrogenase | Rv0733 | <i>adk</i> | probable adenylate kinase | Rv1595 | <i>nadB</i> | L-aspartate oxidase |
| Rv2988c | <i>leuC</i> | 3-isopropylmalate dehydratase large subunit | Rv2364c | <i>bex</i> | GTP-binding protein of Era/ThdF family | Rv1596 | <i>nadC</i> | nicotinate-nucleotide pyrophosphatase |
| Rv2987c | <i>leuD</i> | 3-isopropylmalate dehydratase small subunit | Rv1712 | <i>cmk</i> | cytidylate kinase | Rv0423c | <i>thiC</i> | thiamine synthesis, pyrimidine moiety |
| E. Polyamine synthesis | | | Rv2344c | <i>dgt</i> | probable deoxyguanosine triphosphate hydrolase | 8. Thiamine | | |
| Rv2601 | <i>speE</i> | spermidine synthase | Rv2404c | <i>lepA</i> | GTP-binding protein LepA | Rv0422c | <i>thiD</i> | phosphomethylpyrimidine kinase |
| F. Purines, pyrimidines, nucleosides and nucleotides | | | Rv2727c | <i>miaA</i> | tRNA δ (2)-isopentenylpyrophosphate transferase | Rv0414c | <i>thiE</i> | thiamine synthesis, thiazole moiety |
| 1. Purine ribonucleotide biosynthesis | | | Rv2445c | <i>ndkA</i> | nucleoside diphosphate kinase | Rv0417 | <i>thiG</i> | thiamine synthesis, thiazole moiety |
| Rv1389 | <i>gmk</i> | putative guanylate kinase | Rv2440c | <i>obg</i> | Obg GTP-binding protein | Rv2977c | <i>thiL</i> | probable thiamine-monophosphate kinase |
| Rv3396c | <i>guaA</i> | GMP synthase | Rv2583c | <i>relA</i> | (pppGpp synthase I) | 9. Riboflavin | | |
| Rv1843c | <i>guaB1</i> | inosine-5'-monophosphate dehydrogenase | G. Biosynthesis of cofactors, prosthetic groups and carriers | | | Rv1940 | <i>ribA</i> | GTP cyclohydrolase II |
| Rv3411c | <i>guaB2</i> | inosine-5'-monophosphate dehydrogenase | 1. Biotin | | | Rv1415 | <i>ribA2</i> | probable GTP cyclohydrolase II |
| Rv3410c | <i>guaB3</i> | inosine-5'-monophosphate dehydrogenase | Rv1568 | <i>bioA</i> | adenosylmethionine-8-amino-7-oxononanoate aminotransferase | Rv1412 | <i>ribC</i> | riboflavin synthase α chain |
| Rv1017c | <i>prsA</i> | ribose-phosphate pyrophosphokinase | Rv1589 | <i>bioB</i> | biotin synthase | Rv2671 | <i>ribD</i> | probable riboflavin deaminase |
| Rv0357c | <i>purA</i> | adenylosuccinate synthase | Rv1570 | <i>bioD</i> | dethiobiotin synthase | Rv2786c | <i>ribF</i> | riboflavin kinase |
| Rv0777 | <i>purB</i> | adenylosuccinate lyase | Rv1569 | <i>bioF</i> | 8-amino-7-oxononanoate synthase | Rv1409 | <i>ribG</i> | riboflavin biosynthesis |
| Rv0780 | <i>purC</i> | phosphoribosylaminoimidazole-succinocarboxamide synthase | Rv0032 | <i>bioF2</i> | C-terminal similar to <i>B. subtilis</i> BioF | Rv1416 | <i>ribH</i> | riboflavin synthase β chain |
| Rv0772 | <i>purD</i> | phosphoribosylamine-glycine ligase | Rv3279c | <i>birA</i> | biotin apo-protein ligase | Rv3300c | - | probable deaminase, riboflavin synthesis |
| Rv3275c | <i>purE</i> | phosphoribosylaminoimidazole carboxylase | Rv1442 | <i>bisC</i> | biotin sulfoxide reductase | 10. Thioredoxin, glutaredoxin and mycothiol | | |
| Rv0808 | <i>purF</i> | amidophosphoribosyltransferase | Rv0089 | - | possible <i>bioC</i> biotin synthesis gene | Rv0773c | <i>ggtA</i> | putative γ -glutamyl transpeptidase |
| Rv0957 | <i>purH</i> | phosphoribosylaminoimidazole-carboxamide formyltransferase | 2. Folic acid | | | Rv2394 | <i>ggtB</i> | γ -glutamyltranspeptidase precursor |
| Rv3276c | <i>purK</i> | phosphoribosylaminoimidazole carboxylase ATPase subunit | Rv2763c | <i>dfra</i> | dihydrofolate reductase | Rv2855 | <i>gorA</i> | glutathione reductase homologue |
| Rv0803 | <i>purL</i> | phosphoribosylformylglycinamide synthase II | Rv2447c | <i>foIC</i> | folylpolyglutamate synthase | Rv0816c | <i>thiX</i> | equivalent to <i>M. leprae</i> ThiX |
| Rv0809 | <i>purM</i> | 5'-phosphoribosyl-5-aminoimidazole synthase | Rv3356c | <i>foID</i> | methylene-tetrahydrofolate dehydrogenase | Rv1470 | <i>trxA</i> | thioredoxin |
| Rv0956 | <i>purN</i> | phosphoribosylglycinamide formyltransferase I | Rv3609c | <i>folE</i> | GTP cyclohydrolase I | Rv1471 | <i>trxB</i> | thioredoxin reductase |
| Rv0788 | <i>purQ</i> | phosphoribosylformylglycinamide synthase I | Rv3606c | <i>folK</i> | 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase | Rv3913 | <i>trxC</i> | thioredoxin reductase |
| Rv0389 | <i>purT</i> | phosphoribosylglycinamide formyltransferase II | Rv3608c | <i>folP</i> | dihydropterolate synthase | Rv3914 | <i>trx2</i> | thioredoxin |
| Rv2964 | <i>purU</i> | formyltetrahydrofolate deformylase | Rv1207 | <i>folP2</i> | dihydropterolate synthase | 11. Menaquinone, PQQ, ubiquinone and other terpenoids | | |
| 2. Pyrimidine ribonucleotide biosynthesis | | | Rv3607c | <i>folX</i> | may be involved in folate biosynthesis | Rv2682c | <i>dxs</i> | 1-deoxy-D-xylulose 5-phosphate synthase |
| Rv1383 | <i>carA</i> | carbamoyl-phosphate synthase subunit | Rv0013 | <i>pabA</i> | <i>p</i> -aminobenzoate synthase | Rv0562 | <i>grcC1</i> | heptaprenyl diphosphate synthase I |
| Rv1384 | <i>carB</i> | carbamoyl-phosphate synthase subunit | Rv1005c | <i>pabB</i> | <i>p</i> -aminobenzoate synthase | Rv0989c | <i>grcC2</i> | heptaprenyl diphosphate synthase II |
| Rv1380 | <i>pyrB</i> | aspartate carbamoyltransferase | Rv0812 | <i>pabC</i> | aminodeoxychorismate lyase | Rv3398c | <i>idsA</i> | geranylgeranyl pyrophosphate synthase |
| Rv1381 | <i>pyrC</i> | dihydroorotate | 3. Lipoate | | | Rv2173 | <i>idsA2</i> | geranylgeranyl pyrophosphate synthase |
| Rv2139 | <i>pyrD</i> | dihydroorotate dehydrogenase | Rv2218 | <i>lipA</i> | lipoate biosynthesis protein A | Rv3383c | <i>idsB</i> | transfergeranyl, similar geranyl pyrophosphate synthase |
| Rv1385 | <i>pyrF</i> | orotidine 5'-phosphate decarboxylase | Rv2217 | <i>lipB</i> | lipoate biosynthesis protein B | Rv0534c | <i>menA</i> | 4-dihydroxy-2-naphthoate octaprenyltransferase |
| Rv1699 | <i>pyrG</i> | CTP synthase | 4. Molybdopterin | | | Rv0548c | <i>menB</i> | naphthoate synthase |
| Rv2883c | <i>pyrH</i> | uridylylate kinase | Rv3109 | <i>moaA</i> | molybdenum cofactor biosynthesis, protein A | Rv0553 | <i>menC</i> | <i>o</i> -succinylbenzoate-CoA synthase |
| Rv0382c | <i>umpA</i> | probable uridine 5'-monophosphate synthase | Rv0869c | <i>moaA2</i> | molybdenum cofactor biosynthesis, protein A | Rv0555 | <i>menD</i> | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| 3. 2'-deoxyribonucleotide metabolism | | | Rv0438c | <i>moaA3</i> | molybdenum cofactor biosynthesis, protein A | Rv0542c | <i>menE</i> | <i>o</i> -succinylbenzoic acid-CoA ligase |
| Rv0321 | <i>dcd</i> | deoxycytidine triphosphate deaminase | Rv3110 | <i>moaB</i> | molybdenum cofactor biosynthesis, protein B | Rv3853 | <i>menG</i> | <i>S</i> -adenosylmethionine: 2-demethylmenaquinone phytoene synthase |
| Rv2697c | <i>dut</i> | deoxyuridine triphosphatase | Rv0984 | <i>moaB2</i> | molybdenum cofactor biosynthesis, protein B | Rv3397c | <i>phyA</i> | coenzyme PQQ synthesis protein E |
| Rv0233 | <i>nrdB</i> | ribonucleoside-diphosphate reductase B2 (eukaryotic-like) | Rv3111 | <i>moaC</i> | molybdenum cofactor biosynthesis, protein C | Rv0693 | <i>pqqE</i> | coenzyme PQQ synthesis protein E |
| Rv3051c | <i>nrdE</i> | ribonucleoside diphosphate reductase α chain | Rv0864 | <i>moaC2</i> | molybdenum cofactor biosynthesis, protein C | Rv0558 | <i>ubiE</i> | ubiquinone/menaquinone biosynthesis methyltransferase |
| Rv1981c | <i>nrdF</i> | ribonucleotide reductase small subunit | Rv3324c | <i>moaC3</i> | molybdenum cofactor biosynthesis, protein C | 12. Heme and porphyrin | | |
| | | | Rv3112 | <i>moaD</i> | molybdopterin converting factor subunit 1 | Rv0509 | <i>hemA</i> | glutaryl-tRNA reductase |
| | | | Rv0868c | <i>moaD2</i> | molybdopterin converting factor subunit 2 | Rv0512 | <i>hemB</i> | δ -aminolevulinic acid dehydratase |
| | | | | | | Rv0510 | <i>hemC</i> | porphobilinogen deaminase |
| | | | | | | Rv2678c | <i>hemE</i> | uroporphyrinogen decarboxylase |

Rv1300 *hemK* protoporphyrinogen oxidase
 Rv0524 *hemL* glutamate-1-semialdehyde amino-
 transferase
 Rv2388c *hemN* oxygen-independent copropor-
 phyrinogen 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* precorrin reductase
 Rv2065 *cobH* precorrin isomerase
 Rv2066 *cobI* Cobl-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* cobyrinic acid synthase
 Rv3713 *cobQ2* possible cobyrinic acid synthase
 Rv0306 - similar to BluB cobalamin synthe-
 sis protein *R. capsulatus*

14. Iron utilization

Rv1876 *bfrA* bacterioferritin
 Rv3841 *bfrB* bacterioferritin
 Rv3215 *entC* probable isochorismate synthase
 Rv3214 *entD* weak similarity to many phospho-
 glycerate mutases
 Rv2895c *viuB* similar to proteins involved in
 vibriobactin uptake
 Rv3525c - similar to ferrityochelin 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] pantethienephos-
 photransferase
 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 syn-
 thase 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 methyltrans-
 ferase
 Rv3538 *ufaA2* unknown fatty acid methyltrans-
 ferase
 Rv0469 *umaA1* unknown mycolic acid methyl-
 transferase
 Rv0470c *umaA2* unknown mycolic acid methyl-

transferase
 3. Acyltransferases, mycolyltransferases and
 phospholipid synthesis
 Rv2289 *cdh* CDP-diacylglycerol phosphatidyl-
 hydrolase
 Rv2881c *cdsA* phosphatidate cytidyltransferase
 Rv3804c *fbpA* antigen 85A, mycolyltransferase
 Rv1886c *fbpB* antigen 85B, mycolyltransferase
 Rv0129c *fbpC* antigen 85C, mycolyltransferase
 Rv3803c *fbpD* antigen MPT51, mycolyltrans-
 ferase
 Rv0564c *gpdA1* glycerol-3-phosphate dehydroge-
 nase
 Rv2982c *gpdA2* glycerol-3-phosphate dehydroge-
 nase
 Rv2612c *pgsA* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv1822 *pgsA2* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv2746c *pgsA3* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv1551 *plsB1* glycerol-3-phosphate acyltrans-
 ferase
 Rv2482c *plsB2* glycerol-3-phosphate acyltrans-
 ferase
 Rv0437c *psd* putative phosphatidylserine
 decarboxylase
 Rv0436c *psaA* CDP-diacylglycerol-serine
 α -phosphatidyltransferase
 Rv0045c - possible dihydroipoamide acetyl-
 transferase
 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-acyltrans-
 ferase-a
 Rv2262c - apolipoprotein N-acyltrans-
 ferase-b
 Rv3523 - lipid carrier protein
 Rv3720 - C-term similar to cyclopropane
 fatty acid synthases

Rv0564c *gpdA1* glycerol-3-phosphate dehydroge-
 nase
 Rv2982c *gpdA2* glycerol-3-phosphate dehydroge-
 nase
 Rv2612c *pgsA* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv1822 *pgsA2* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv2746c *pgsA3* CDP-diacylglycerol-glycerol-3-
 phosphate phosphatidyltrans-
 ferase
 Rv1551 *plsB1* glycerol-3-phosphate acyltrans-
 ferase
 Rv2482c *plsB2* glycerol-3-phosphate acyltrans-
 ferase
 Rv0437c *psd* putative phosphatidylserine
 decarboxylase
 Rv0436c *psaA* CDP-diacylglycerol-serine
 α -phosphatidyltransferase
 Rv0045c - possible dihydroipoamide acetyl-
 transferase
 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-acyltrans-
 ferase-a
 Rv2262c - apolipoprotein N-acyltrans-
 ferase-b
 Rv3523 - lipid carrier protein
 Rv3720 - C-term similar to cyclopropane
 fatty acid synthases

I. Polyketide and non-ribosomal peptide synthesis

Rv2940c *mas* mycoerolic 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 α -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* phenolphthiocerol synthesis (*pkS8*)
 Rv2932 *ppsB* phenolphthiocerol synthesis (*pkS9*)
 Rv2933 *ppsC* phenolphthiocerol synthesis (*pkS10*)
 Rv2934 *ppsD* phenolphthiocerol synthesis (*pkS11*)
 Rv2935 *ppsE* phenolphthiocerol synthesis (*pkS12*)
 Rv2928 *tesA* thioesterase
 Rv1544 - probable ketoacyl reductase

J. Broad regulatory functions

1. Repressors/activators

Rv1657 *argR* arginine repressor
 Rv1267c *embR* regulator of *embAB* genes
 (*AisR/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 *maxR* transcriptional regulator, MoxR
 homologue
 Rv3692 *maxR2* transcriptional regulator, MoxR
 homologue
 Rv3164c *maxR3* 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)

| | | | | | |
|---------|---|---|--|---|---|
| Rv0823c | - | family) transcriptional regulator (NifR3/Smm1 family) | Rv3160c | - | putative transcriptional regulator |
| Rv0827c | - | transcriptional regulator (ArsR family) | Rv3167c | - | putative transcriptional regulator |
| Rv0890c | - | transcriptional regulator (LuxR/UhpA family) | Rv3173c | - | transcriptional regulator (TetR/AcrR family) |
| Rv0891c | - | putative transcriptional regulator | Rv3183 | - | putative transcriptional regulator |
| Rv0894 | - | putative transcriptional regulator | Rv3208 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1019 | - | transcriptional regulator (TetR/AcrR family) | Rv3249c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1049 | - | transcriptional regulator (MarR family) | Rv3291c | - | transcriptional regulator (Lrp/AsnC family) |
| Rv1129c | - | transcriptional regulator (PbsX/Xre family) | Rv3295 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1151c | - | putative transcriptional regulator | Rv3334 | - | transcriptional regulator (MerR family) |
| Rv1152 | - | transcriptional regulator (GntR family) | Rv3405c | - | putative transcriptional regulator |
| Rv1167c | - | putative transcriptional regulator | Rv3522 | - | putative transcriptional regulator |
| Rv1219c | - | putative transcriptional regulator | Rv3557c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1255c | - | transcriptional regulator (TetR/AcrR family) | Rv3574 | - | transcriptional regulator (TetR/AcrR family) |
| Rv1332 | - | putative transcriptional regulator | Rv3575c | - | transcriptional regulator (LacI family) |
| Rv1353c | - | transcriptional regulator (TetR/AcrR family) | Rv3583c | - | putative transcriptional regulator |
| Rv1358 | - | transcriptional regulator (LuxR/UhpA family) | Rv3676 | - | transcriptional regulator (Crp/Fnr family) |
| Rv1359 | - | putative transcriptional regulator | Rv3678c | - | transcriptional regulator (LysR family) |
| Rv1395 | - | transcriptional regulator (AraC/XylS family) | Rv3736 | - | transcriptional regulator (AraC/XylS family) |
| Rv1404 | - | transcriptional regulator (MarR family) | Rv3744 | - | transcriptional regulator (ArsR family) |
| Rv1423 | - | putative transcriptional regulator | Rv3830c | - | transcriptional regulator (TetR/AcrR family) |
| Rv1460 | - | putative transcriptional regulator | Rv3833 | - | transcriptional regulator (AraC/XylS family) |
| Rv1474c | - | transcriptional regulator (TetR/AcrR family) | Rv3840 | - | putative transcriptional regulator |
| Rv1534 | - | transcriptional regulator (TetR/AcrR family) | Rv3855 | - | putative transcriptional regulator |
| Rv1556 | - | putative transcriptional regulator | 2. Two component systems | | |
| Rv1674c | - | putative transcriptional regulator | Rv1028c | - | <i>kdpD</i> sensor histidine kinase |
| Rv1675c | - | putative transcriptional regulator | Rv1027c | - | <i>kdpE</i> two-component response regulator |
| Rv1719 | - | transcriptional regulator (lclR family) | Rv3246c | - | <i>ntrA</i> two-component response regulator |
| Rv1773c | - | transcriptional regulator (lclR family) | Rv3245c | - | <i>ntrB</i> sensor histidine kinase |
| Rv1776c | - | putative transcriptional regulator | Rv0844c | - | <i>narL</i> two-component response regulator |
| Rv1816 | - | putative transcriptional regulator | Rv0757 | - | <i>phoP</i> two-component response regulator |
| Rv1846c | - | putative transcriptional regulator | Rv0758 | - | <i>phoR</i> sensor histidine kinase |
| Rv1931c | - | transcriptional regulator (AraC/XylS family) | Rv0491 | - | <i>regX3</i> two-component response regulator |
| Rv1956 | - | putative transcriptional regulator | Rv0490 | - | <i>senX3</i> sensor histidine kinase |
| Rv1963c | - | putative transcriptional regulator | Rv0602c | - | <i>tcra</i> two-component response regulator |
| Rv1985c | - | transcriptional regulator (LysR family) | Rv0260c | - | two-component response regulator |
| Rv1990c | - | putative transcriptional regulator | Rv0600c | - | sensor histidine kinase |
| Rv1994c | - | transcriptional regulator (MerR family) | Rv0601c | - | sensor histidine kinase |
| Rv2017 | - | putative transcriptional regulator (PbsX/Xre family) | Rv0818 | - | two-component response regulator |
| Rv2021c | - | putative transcriptional regulator | Rv0845 | - | sensor histidine kinase |
| Rv2034 | - | transcriptional regulator (ArsR family) | Rv0902c | - | sensor histidine kinase |
| Rv2175c | - | putative transcriptional regulator | Rv0903c | - | two-component response regulator |
| Rv2250c | - | putative transcriptional regulator | Rv0981 | - | two-component response regulator |
| Rv2258c | - | putative transcriptional regulator | Rv0982 | - | sensor histidine kinase |
| Rv2282c | - | transcriptional regulator (LysR family) | Rv1032c | - | sensor histidine kinase |
| Rv2308 | - | putative transcriptional regulator | Rv1033c | - | two-component response regulator |
| Rv2324 | - | transcriptional regulator (Lrp/AsnC family) | Rv1626 | - | two-component response regulator |
| Rv2358 | - | transcriptional regulator (ArsR family) | Rv2027c | - | sensor histidine kinase |
| Rv2488c | - | transcriptional regulator (LuxR/UhpA family) | Rv2884 | - | two-component response regulator |
| Rv2506 | - | transcriptional regulator (TetR/AcrR family) | Rv3132c | - | sensor histidine kinase |
| Rv2621c | - | putative transcriptional regulator | Rv3133c | - | two-component response regulator |
| Rv2640c | - | transcriptional regulator (ArsR family) | Rv3143 | - | putative sensory transduction protein |
| Rv2642 | - | transcriptional regulator (ArsR family) | Rv3220c | - | sensor histidine kinase |
| Rv2669 | - | putative transcriptional regulator | Rv3764c | - | sensor histidine kinase |
| Rv2745c | - | putative transcriptional regulator | Rv3765c | - | two-component response regulator |
| Rv2779c | - | transcriptional regulator (Lrp/AsnC family) | 3. Serine-threonine protein kinases and phosphoprotein phosphatases | | |
| Rv2887 | - | transcriptional regulator (MarR family) | Rv0015c | - | <i>pknA</i> serine-threonine protein kinase |
| Rv2912c | - | transcriptional regulator (TetR/AcrR family) | Rv0014c | - | <i>pknB</i> serine-threonine protein kinase |
| Rv2989 | - | transcriptional regulator (lclR family) | Rv0931c | - | <i>pknD</i> serine-threonine protein kinase |
| Rv3050c | - | putative transcriptional regulator | Rv1743 | - | <i>pknE</i> serine-threonine protein kinase |
| Rv3055 | - | putative transcriptional regulator | Rv1746 | - | <i>pknF</i> serine-threonine protein kinase |
| Rv3058c | - | putative transcriptional regulator | Rv0410c | - | <i>pknG</i> serine-threonine protein kinase |
| Rv3060c | - | transcriptional regulator (GntR family) | Rv1266c | - | <i>pknH</i> serine-threonine protein kinase |
| Rv3066 | - | putative transcriptional regulator | Rv2914c | - | <i>pknI</i> serine-threonine protein kinase |
| Rv3095 | - | putative transcriptional regulator | Rv2088 | - | <i>pknJ</i> serine-threonine protein kinase |
| Rv3124 | - | transcriptional regulator (AfsR/DndI/RedD family) | Rv3080c | - | <i>pknK</i> serine-threonine protein kinase |
| | | | Rv2176 | - | <i>pknL</i> serine-threonine protein kinase, |

| | | |
|---------|---|---|
| Rv0018c | - | <i>ppp</i> truncated |
| Rv2234 | - | <i>ptpA</i> 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 | - | <i>rimI</i> ribosomal protein S18 acetyl transferase |
| Rv0995 | - | <i>rimJ</i> acetylation of 30S S5 subunit |
| Rv0641 | - | <i>rplA</i> 50S ribosomal protein L1 |
| Rv0704 | - | <i>rplB</i> 50S ribosomal protein L2 |
| Rv0701 | - | <i>rplC</i> 50S ribosomal protein L3 |
| Rv0702 | - | <i>rplD</i> 50S ribosomal protein L4 |
| Rv0716 | - | <i>rplE</i> 50S ribosomal protein L5 |
| Rv0719 | - | <i>rplF</i> 50S ribosomal protein L6 |
| Rv0056 | - | <i>rplI</i> 50S ribosomal protein L9 |
| Rv0651 | - | <i>rplJ</i> 50S ribosomal protein L10 |
| Rv0640 | - | <i>rplK</i> 50S ribosomal protein L11 |
| Rv0652 | - | <i>rplL</i> 50S ribosomal protein L7/L12 |
| Rv3443c | - | <i>rplM</i> 50S ribosomal protein L13 |
| Rv0714 | - | <i>rplN</i> 50S ribosomal protein L14 |
| Rv0723 | - | <i>rplO</i> 50S ribosomal protein L15 |
| Rv0708 | - | <i>rplP</i> 50S ribosomal protein L16 |
| Rv3456c | - | <i>rplQ</i> 50S ribosomal protein L17 |
| Rv0720 | - | <i>rplR</i> 50S ribosomal protein L18 |
| Rv2904c | - | <i>rplS</i> 50S ribosomal protein L19 |
| Rv1643 | - | <i>rplT</i> 50S ribosomal protein L20 |
| Rv2442c | - | <i>rplU</i> 50S ribosomal protein L21 |
| Rv0706 | - | <i>rplV</i> 50S ribosomal protein L22 |
| Rv0703 | - | <i>rplW</i> 50S ribosomal protein L23 |
| Rv0715 | - | <i>rplX</i> 50S ribosomal protein L24 |
| Rv1015c | - | <i>rplY</i> 50S ribosomal protein L25 |
| Rv2441c | - | <i>rpmA</i> 50S ribosomal protein L27 |
| Rv1015c | - | <i>rpmB</i> 50S ribosomal protein L28 |
| Rv2058c | - | <i>rpmB2</i> 50S ribosomal protein L28 |
| Rv0709 | - | <i>rpmC</i> 50S ribosomal protein L29 |
| Rv0722 | - | <i>rpmD</i> 50S ribosomal protein L30 |
| Rv1298 | - | <i>rpmE</i> 50S ribosomal protein L31 |
| Rv2057c | - | <i>rpmG</i> 50S ribosomal protein L33 |
| Rv3924c | - | <i>rpmH</i> 50S ribosomal protein L34 |
| Rv1642 | - | <i>rpmI</i> 50S ribosomal protein L35 |
| Rv3461c | - | <i>rpmJ</i> 50S ribosomal protein L36 |
| Rv1630 | - | <i>rpsA</i> 30S ribosomal protein S1 |
| Rv2890c | - | <i>rpsB</i> 30S ribosomal protein S2 |
| Rv0707 | - | <i>rpsC</i> 30S ribosomal protein S3 |
| Rv3458c | - | <i>rpsD</i> 30S ribosomal protein S4 |
| Rv0721 | - | <i>rpsE</i> 30S ribosomal protein S5 |
| Rv0053 | - | <i>rpsF</i> 30S ribosomal protein S6 |
| Rv0683 | - | <i>rpsG</i> 30S ribosomal protein S7 |
| Rv0718 | - | <i>rpsH</i> 30S ribosomal protein S8 |
| Rv3442c | - | <i>rpsI</i> 30S ribosomal protein S9 |
| Rv0700 | - | <i>rpsJ</i> 30S ribosomal protein S10 |
| Rv3459c | - | <i>rpsK</i> 30S ribosomal protein S11 |
| Rv0682 | - | <i>rpsL</i> 30S ribosomal protein S12 |
| Rv3460c | - | <i>rpsM</i> 30S ribosomal protein S13 |
| Rv0717 | - | <i>rpsN</i> 30S ribosomal protein S14 |
| Rv2056c | - | <i>rpsN2</i> 30S ribosomal protein S14 |
| Rv2785c | - | <i>rpsO</i> 30S ribosomal protein S15 |
| Rv2909c | - | <i>rpsP</i> 30S ribosomal protein S16 |
| Rv0710 | - | <i>rpsQ</i> 30S ribosomal protein S17 |
| Rv0055 | - | <i>rpsR</i> 30S ribosomal protein S18 |
| Rv2055c | - | <i>rpsR2</i> 30S ribosomal protein S18 |
| Rv0705 | - | <i>rpsS</i> 30S ribosomal protein S19 |
| Rv2412 | - | <i>rpsT</i> 30S ribosomal protein S20 |
| Rv3241c | - | member of S30AE ribosomal protein family |

2. Ribosome modification and maturation

| | | |
|---------|---|--|
| Rv1010 | - | <i>ksgA</i> 16S rRNA dimethyltransferase |
| Rv2838c | - | <i>rbfA</i> ribosome-binding factor A |
| Rv2907c | - | <i>rimM</i> 16S rRNA processing protein |

3. Aminoacyl tRNA synthases and their modification

| | | |
|---------|---|---|
| Rv2555c | - | <i>alaS</i> alanyl-tRNA synthase |
| Rv1292 | - | <i>argS</i> arginyl-tRNA synthase |
| Rv2572c | - | <i>aspS</i> aspartyl-tRNA synthase |
| Rv3580c | - | <i>cysS</i> cysteinyl-tRNA synthase |
| Rv2130c | - | <i>cysS2</i> cysteinyl-tRNA synthase |
| Rv1406 | - | <i>fmt</i> methionyl-tRNA formyltransferase |
| Rv3011c | - | <i>gata</i> glu-tRNA-gln amidotransferase, subunit B |
| Rv3009c | - | <i>gatB</i> glu-tRNA-gln amidotransferase, subunit A |
| Rv3012c | - | <i>gatC</i> glu-tRNA-gln amidotransferase, subunit C |
| Rv2992c | - | <i>gltS</i> glutamyl-tRNA synthase |
| Rv2357c | - | <i>glyS</i> glycylyl-tRNA synthase |
| Rv2580c | - | <i>hisS</i> histidyl-tRNA synthase |
| Rv1536 | - | <i>ileS</i> isoleucyl-tRNA synthase |
| Rv0041 | - | <i>leuS</i> leucyl-tRNA synthase |
| Rv3598c | - | <i>lysS</i> lysyl-tRNA synthase |
| Rv1640c | - | <i>lysX</i> C-term lysyl-tRNA synthase |
| Rv1007c | - | <i>metS</i> methionyl-tRNA synthase |
| Rv1649 | - | <i>pheS</i> 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 *fmv* similar to Fmv protein
- Rv3852 *hns* HU-histone protein
- Rv2986c *hupB* DNA-binding protein II
- Rv1388 *mlhF* 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 *heiY* probable helicase, Ski2 subfamily
- Rv2101 *heiZ* 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 *traA* probable pseudouridylylase 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 *mc* RNase III
- Rv2444c *me* similar to C-term to ribonuclease E
- Rv2902c *mhb* ribonuclease HII
- Rv3923c *mpA* 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
- Rv2667 *clpX'* ATP-binding subunit ClpX similar to ClpC from *M. leprae* but shorter
- Rv3419c *gcp* glycoprotease
- Rv2725c *htrX* GTP-binding protein
- Rv1223 *htrA* serine protease
- Rv2861c *mapA1* methionine aminopeptidase
- Rv0734 *mapA2* probable methionine aminopeptidase
- Rv0319 *ppc* 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 *cwIM* 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-

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

C. Cell envelope

1. Lipoproteins (*ippA-ippO*) 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 (Eip 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 *dala* 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-pentapeptide transferase
Rv3332 *nagA* N-acetylglucosamine-6-P-deacetylase
Rv0016c *pbpA* penicillin-binding protein
Rv2163c *pbpB* penicillin-binding protein 2
Rv0050 *ponA1* penicillin-binding protein class A penicillin binding protein
Rv3682 *ponA2* FtsW/RodA/SpovE family
Rv0017c *rodA* probable penicillin binding protein
Rv0907 -

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 *ctpl* probable magnesium transport ATPase
Rv0969 *ctpV* cation transport ATPase
Rv3044 *fecB* putative Fe(III)-dicitrate transporter
Rv0265c *fecB2* iron transport protein Fe(III) dicitrate transporter
Rv1029 *kdpA* potassium-transporting ATPase A chain

Rv1030 *kdpB* potassium-transporting ATPase B chain
Rv1031 *kdpC* potassium-transporting ATPase C chain
Rv3236c *ketB* 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

| | | |
|----------------------------------|---------------|---|
| Rv0932c | <i>pstS</i> | phosphate transport system PstS component of phosphate uptake |
| Rv2400c | <i>subI</i> | 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 | <i>ltp1</i> | non-specific lipid transport protein |
| Rv3540c | <i>ltp2</i> | non-specific lipid transport protein |
| 6. Efflux proteins | | |
| Rv2936 | <i>drvA</i> | similar daunorubicin resistance ABC-transporter |
| Rv2937 | <i>drvB</i> | similar daunorubicin resistance transmembrane protein |
| Rv2938 | <i>drvC</i> | similar daunorubicin resistance transmembrane protein |
| Rv2846c | <i>efpA</i> | putative efflux protein |
| Rv3065 | <i>emrE</i> | 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 | <i>clpB</i> | heat shock protein |
| Rv0352 | <i>dnaJ</i> | acts with GrpE to stimulate DnaK ATPase |
| Rv2373c | <i>dnaJ2</i> | DnaJ homologue |
| Rv0350 | <i>dnaK</i> | 70 kD heat shock protein, chromosome replication |
| Rv3417c | <i>groEL1</i> | 60 kD chaperonin 1 |
| Rv0440 | <i>groEL2</i> | 60 kD chaperonin 2 |
| Rv3418c | <i>groES</i> | 10 kD chaperone |
| Rv0351 | <i>grpE</i> | stimulates DnaK ATPase activity |
| Rv2374c | <i>hrcA</i> | heat-inducible transcription repressor |
| Rv0251c | <i>hsp</i> | possible heat shock protein |
| Rv0353 | <i>hspR</i> | heat shock regulator |
| Rv2031c | <i>hspX</i> | 14kD antigen, heat shock protein Hsp20 family |
| Rv2299c | <i>htpG</i> | heat shock protein Hsp90 family |
| Rv0563 | <i>htpX</i> | probable (transmembrane) heat shock protein |
| Rv2701c | <i>suhB</i> | putative extragenic suppressor protein |
| Rv3269 | - | probable heat shock protein |
| C. Cell division | | |
| Rv3641c | <i>fic</i> | possible cell division protein |
| Rv3102c | <i>ftsE</i> | membrane protein |
| Rv3610c | <i>ftsH</i> | inner membrane protein, chaperone |
| Rv2748c | <i>ftsK</i> | chromosome partitioning |
| Rv2151c | <i>ftsQ</i> | ingrowth of wall at septum |
| Rv2154c | <i>ftsW</i> | membrane protein (shape determination) |
| Rv3101c | <i>ftsX</i> | membrane protein |
| Rv2921c | <i>ftsY</i> | cell division protein FtsY |
| Rv2150c | <i>ftsZ</i> | circumferential ring, GTPase |
| Rv3919c | <i>gid</i> | glucose inhibited division protein B |
| Rv3625c | <i>mesJ</i> | probable cell cycle protein |
| Rv3917c | <i>parA</i> | chromosome partitioning; DNA-binding |
| Rv3918c | <i>parB</i> | possibly involved in chromosome partitioning |
| Rv2922c | <i>smc</i> | 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 | <i>ffh</i> | signal recognition particle protein |
| Rv2903c | <i>lepB</i> | signal peptidase I |
| Rv1614 | <i>lgt</i> | prolipoprotein diacylglycerol transferase |
| Rv1539 | <i>lspA</i> | lipoprotein signal peptidase |
| Rv0379 | <i>sec</i> | probable transport protein SecE/Sec61- γ family |
| Rv3240c | <i>secA</i> | SecA, preprotein translocase sub- |

| | | |
|--|--------------|--|
| Rv1821 | <i>secA2</i> | unit SecA, preprotein translocase sub-unit |
| Rv2587c | <i>secD</i> | protein-export membrane protein |
| Rv0638 | <i>secE</i> | SecE preprotein translocase |
| Rv2586c | <i>secF</i> | protein-export membrane protein |
| Rv1440 | <i>secG</i> | protein-export membrane protein |
| Rv0732 | <i>secY</i> | SecY subunit of preprotein translocase |
| Rv2462c | <i>tig</i> | chaperone protein, similar to trigger factor |
| Rv2813 | - | probable general secretion pathway protein |
| E. Adaptations and atypical conditions | | |
| Rv1901 | <i>cinA</i> | competence damage protein |
| Rv3648c | <i>cspA</i> | cold shock protein, transcriptional regulator |
| Rv0871 | <i>cspB</i> | probable cold shock protein |
| Rv3063 | <i>cstA</i> | starvation-induced stress response protein |
| Rv3490 | <i>otsA</i> | probable α , α -trehalose-phosphate synthase |
| Rv2006 | <i>otsB</i> | trehalose-6-phosphate phosphatase |
| Rv3372 | <i>otsB2</i> | trehalose-6-phosphate phosphatase |
| Rv3758c | <i>proV</i> | osmoprotection ABC transporter |
| Rv3757c | <i>proW</i> | transport system permease |
| Rv3759c | <i>proX</i> | similar to osmoprotection proteins |
| Rv3756c | <i>proZ</i> | transport system permease |
| Rv1026 | - | probable pppGpp-5'phosphohydrolyase |
| F. Detoxification | | |
| Rv2428 | <i>ahpC</i> | alkyl hydroperoxide reductase |
| Rv2429 | <i>ahpD</i> | member of AhpC/TSA family |
| Rv2238c | <i>ahpE</i> | member of AhpC/TSA family |
| Rv2521 | <i>bcp</i> | bacterioferritin comigratory protein |
| Rv1608c | <i>bcpB</i> | probable bacterioferritin comigratory protein |
| Rv3473c | <i>bpoA</i> | probable non-heme bromoperoxidase |
| Rv1123c | <i>bpoB</i> | probable non-heme bromoperoxidase |
| Rv0554 | <i>bpoC</i> | probable non-heme bromoperoxidase |
| Rv3617 | <i>ephA</i> | probable epoxide hydrolase |
| Rv1938 | <i>ephB</i> | probable epoxide hydrolase |
| Rv1124 | <i>ephC</i> | probable epoxide hydrolase |
| Rv2214c | <i>ephD</i> | probable epoxide hydrolase |
| Rv3670 | <i>ephE</i> | probable epoxide hydrolase |
| Rv0134 | <i>ephF</i> | probable epoxide hydrolase |
| Rv3171c | <i>hpx</i> | probable non-heme haloperoxidase |
| Rv1908c | <i>katG</i> | catalase-peroxidase |
| Rv3846 | <i>sodA</i> | superoxide dismutase |
| Rv0432 | <i>sodC</i> | superoxide dismutase precursor - (Cu-Zn) |
| Rv1932 | <i>tpx</i> | thiol peroxidase |
| Rv0634c | - | putative glyoxylase II |
| Rv2581c | - | putative glyoxylase II |
| Rv3177 | - | probable non-heme haloperoxidase |
| IV. Other | | |
| A. Virulence | | |
| Rv0169 | <i>mce1</i> | cell invasion protein |
| Rv0589 | <i>mce2</i> | cell invasion protein |
| Rv1966 | <i>mce3</i> | cell invasion protein |
| Rv3499c | <i>mce4</i> | cell invasion protein |
| Rv3100c | <i>smgB</i> | probable small protein b |
| Rv1694 | <i>tyaA</i> | cytotoxin/hemolysin homologue |
| Rv0024 | - | putative p60 homologue |
| Rv0167 | - | part of <i>mce1</i> operon |
| Rv0168 | - | part of <i>mce1</i> operon |
| Rv0170 | - | part of <i>mce1</i> operon |
| Rv0171 | - | part of <i>mce1</i> operon |
| Rv0172 | - | part of <i>mce1</i> operon |
| Rv0174 | - | part of <i>mce1</i> operon |
| Rv0587 | - | part of <i>mce2</i> operon |
| Rv0588 | - | part of <i>mce2</i> operon |
| Rv0590 | - | part of <i>mce2</i> operon |
| Rv0591 | - | part of <i>mce2</i> operon |
| Rv0592 | - | part of <i>mce2</i> operon |
| Rv0594 | - | part of <i>mce2</i> 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 <i>mce3</i> operon |
| Rv1965 | - | part of <i>mce3</i> operon |
| Rv1967 | - | part of <i>mce3</i> operon |
| Rv1968 | - | part of <i>mce3</i> operon |
| Rv1969 | - | part of <i>mce3</i> operon |
| Rv1971 | - | part of <i>mce3</i> operon |
| Rv2190c | - | putative p60 homologue |
| Rv3494c | - | part of <i>mce4</i> operon |
| Rv3496c | - | part of <i>mce4</i> operon |
| Rv3497c | - | part of <i>mce4</i> operon |
| Rv3498c | - | part of <i>mce4</i> operon |

| | | |
|--|-------------|--|
| Rv3500c | - | part of <i>mce4</i> operon |
| Rv3501c | - | part of <i>mce4</i> operon |
| Rv3896c | - | putative p60 homologue |
| Rv3922c | - | possible hemolysin |
| B. IS elements, Repeated sequences, and Phage | | |
| 1. IS elements | | |
| IS6110 | - | 16 copies |
| IS1081 | - | 6 copies |
| Others | - | 34 copies |
| 2. REP13E12 family 7 copies | | |
| 3. Phage-related functions | | |
| Rv2894c | <i>xerC</i> | integrase/recombinase |
| Rv1701 | <i>xerD</i> | 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 <i>phd</i> 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 | <i>blaC</i> | class A β -lactamase |
| Rv3290c | <i>lat</i> | lysine-c aminotransferase |
| Rv2043c | <i>pncA</i> | pyrazinamide resistance/sensitivity |
| Rv0133 | - | possible puromycin N-acetyltransferase |
| Rv0262c | - | aminoglycoside 2'-N-acetyltransferase |
| Rv0802c | - | acetyltransferase |
| Rv1082 | - | similar to <i>S. lincolnensis</i> <i>lmbE</i> |
| Rv1170 | - | similar to <i>S. lincolnensis</i> <i>lmbE</i> |
| Rv1347c | - | possible aminoglycoside 6'-N-acetyltransferase |
| Rv2036 | - | similar to lincomycin production genes |
| Rv2303c | - | similar to <i>S. griseus</i> 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. Chelataes 2 | | |
| V. Conserved hypotheticals 912 | | |
| VI. Unknowns 606 | | |
| TOTAL | - | 3924 |

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 agent¹. 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 homogeneity². 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 infection³. 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 rare⁶. 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 isolates⁶, 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 replication¹⁰. 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.

Insertion sequences and prophages. Sixteen copies of the promiscuous insertion sequence IS6110 and six copies of the more stable element IS1081 reside within the genome of H37Rv⁸. One copy of IS1081 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 IS3 and IS256 families, although six of them define a new group. There is extensive similarity between IS1561 and IS1552 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 IS6110. This bias was also observed experimentally in a transposon mutagenesis study¹².

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 lost¹³. In a systematic study of the genomic diversity of prophages and insertion sequences (S.V.G. *et al.*, manuscript in preparation), only IS1532 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.

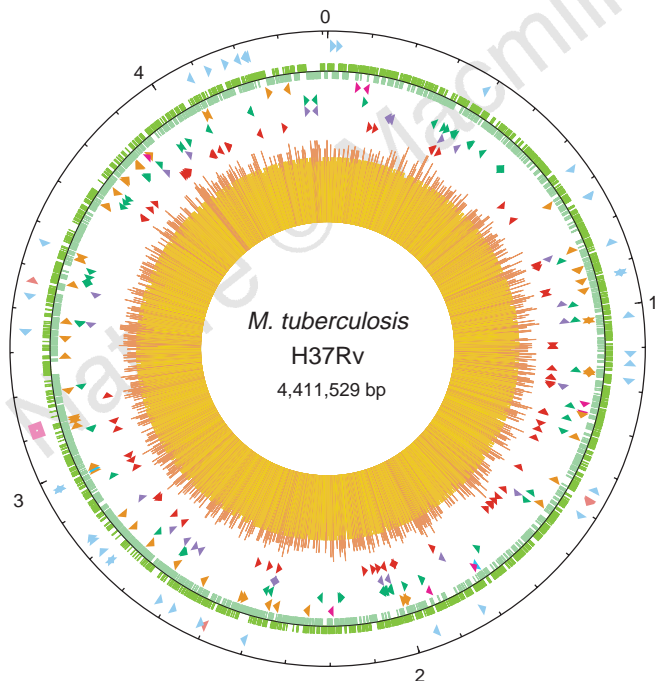


Figure 1 Circular map of the chromosome of *M. tuberculosis* H37Rv. The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes (tRNAs are blue, others are pink) and the direct repeat region (pink cube); the second ring inward 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.

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 sites¹⁵ (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 analysis¹⁶, 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 descent⁶.

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 systems¹⁴, *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 system¹⁸. 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 difficult¹⁹. This resistance is due mainly to the highly hydrophobic cell envelope acting as a permeability barrier⁴, 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.

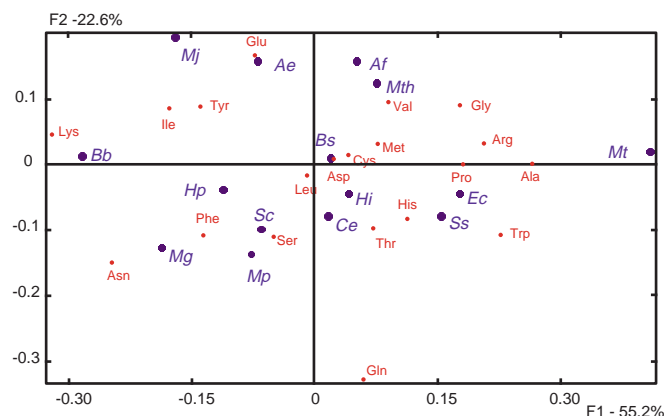


Figure 3 Correspondence analysis of the proteomes from extensively sequenced organisms as a function of amino-acid composition. Note the extreme position of *M. tuberculosis* and the shift in amino-acid preference reflecting increasing G + C content from left to right. Abbreviations used: Ae, *Aquifex aeolicus*; Af, *Archaeoglobus fulgidis*; Bb, *Borrelia burgdorferi*; 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 axes¹⁶.

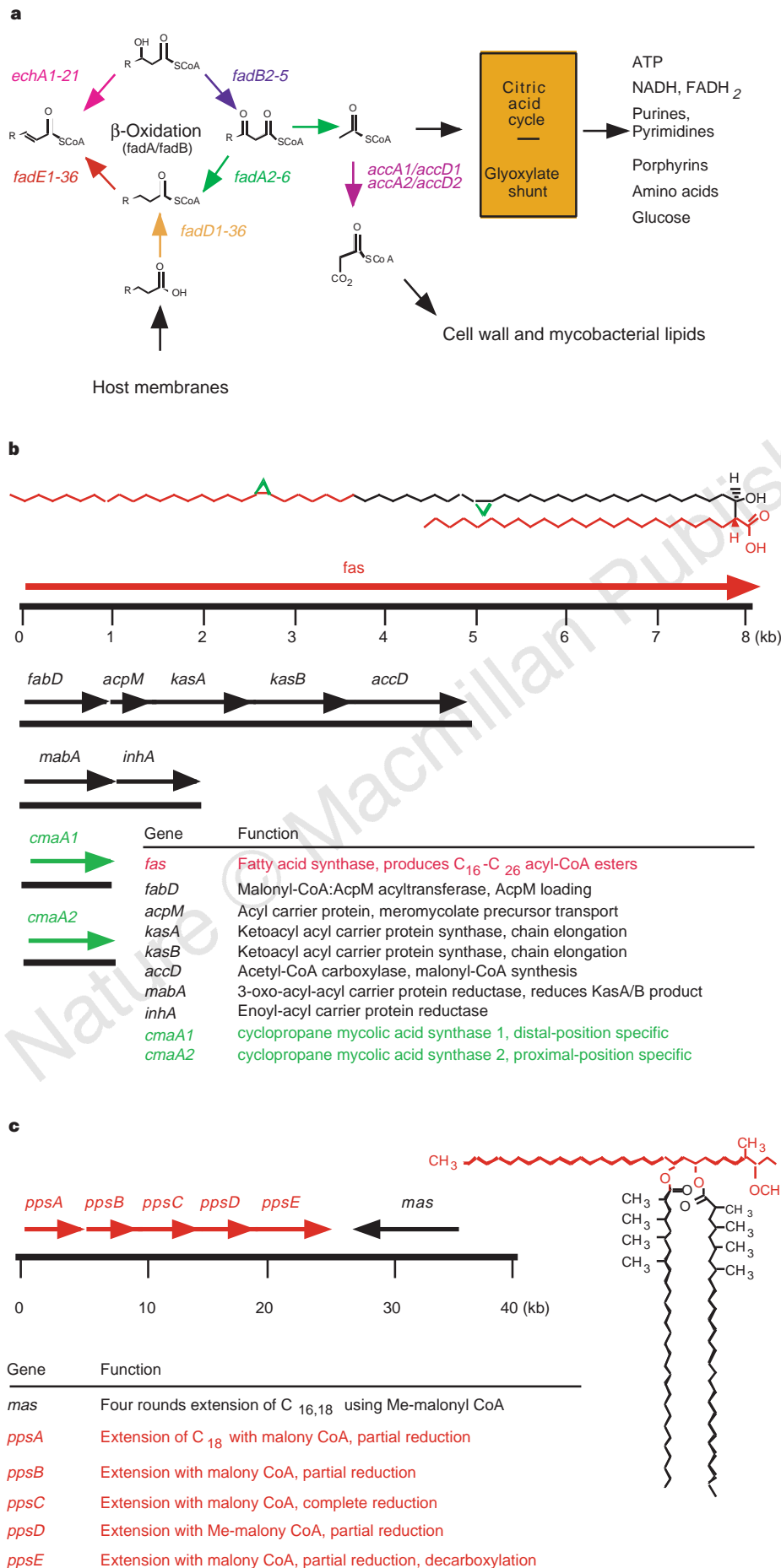


Figure 4 Lipid metabolism. **a**, Degradation of host-cell lipids is vital in the intracellular life of *M. tuberculosis*. Host-cell membranes provide precursors for many metabolic processes, as well as potential precursors of mycobacterial cell-wall constituents, through the actions of a broad family 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 tubercle² (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 primers⁵ 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* species²³. *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

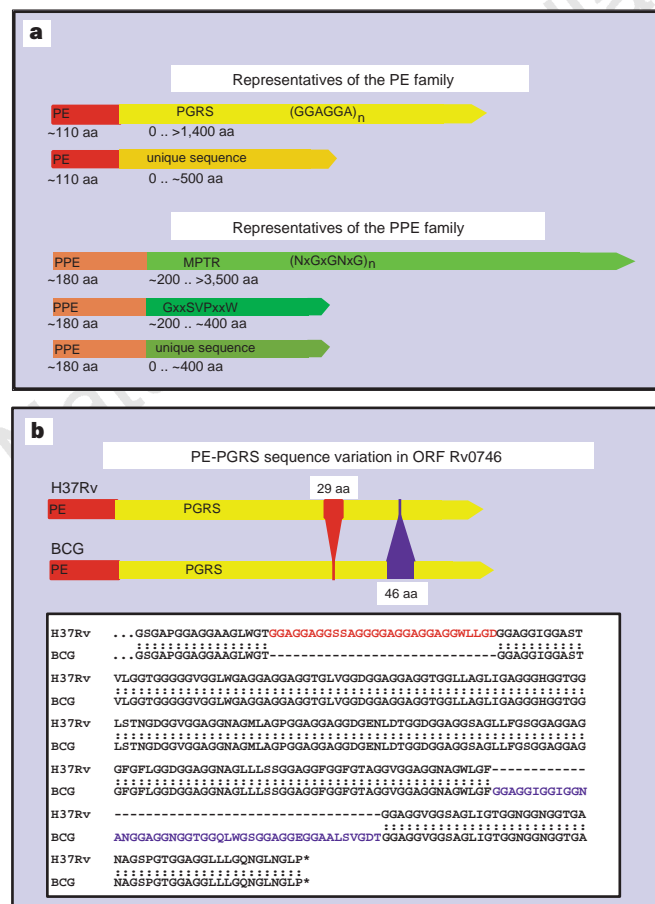


Figure 5 The PE and PPE protein families. **a**, Classification of the PE and PPE protein families. **b**, Sequence variation between *M. tuberculosis* H37Rv and *M. bovis* BCG-Pasteur in the PE-PGRS encoded by open reading frame (ORF) Rv0746.

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 characterized²⁵ 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 complex²⁶. 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 mycolyltransferase 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 biosynthesis²³, is encoded by a very large operon, *ppsABCDE*, and functions in the production of phenolphthiocerol⁵. 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 molecules⁵. 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 mycolipanic acids or the phthioceranic and hydroxyphthioceranic acids, or may even show functional overlap⁵, 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 synthase²³. 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 *pk10*, *pk7*, *pk8* and *pk9*, 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 siderophores²⁸.

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 disease²⁹. 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 cases³³. 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 antigen³⁴. 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 isolates³³. 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 replication³². These mechanisms are known to generate antigenic variability in other bacterial pathogens³⁶.

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 programmes⁴⁰.

Pathogenicity. Despite intensive research efforts, there is little information about the molecular basis of mycobacterial virulence⁴¹. 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 described⁴¹: catalase-peroxidase, which protects against reactive oxygen species produced by the phagocyte; *mce*, which encodes macrophage-colonizing factor⁴²; and a sigma factor gene, *sigA* (aka *rpoV*), mutations in which can lead to attenuation⁴¹. In addition to these single-gene virulence factors, the mycobacterial cell wall⁴ 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 cells⁴². Furthermore, a homologue of *smpB*, which has been implicated in intracellular survival of *Salmonella typhimurium*, has also been identified⁴³. 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 cosmids⁸ and the remainder was obtained from selected BAC clones⁷ 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 libraries⁴⁴ to sequence regions prone to compression or deletion and, in some cases, obtained sequences from products of the polymerase chain reaction or directly from BACs⁷. 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-parse⁴⁶ 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, PROSITE⁴⁷ and in-house databases involved BLASTN, BLASTX⁴⁸, DOTTER (<http://www.sanger.ac.uk>) and FASTA⁴⁹. tRNA genes were located and identified using tRNAscan and tRNAscan-SE⁵⁰. 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, Ravighione, 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. & Labeledan, 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 physiological 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 goodii*. *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 histocompatibility 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., Stojkovic, 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. Sci. 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/DDJB as MTBH37RV, accession number AL123456.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

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

I. Small-molecule metabolism

A. Degradation

1. Carbon compounds

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

2. Amino acids and amines

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

3. Fatty acids

| | | |
|---------|---------------|---|
| Rv2501c | <i>accA1</i> | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv0973c | <i>accA2</i> | acetyl/propionyl-CoA carboxylase, α subunit |
| Rv2502c | <i>accD1</i> | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv0974c | <i>accD2</i> | acetyl/propionyl-CoA carboxylase, β subunit |
| Rv3667 | <i>acs</i> | acetyl-CoA synthase |
| Rv3409c | <i>choD</i> | cholesterol oxidase |
| Rv0222 | <i>echA1</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0456c | <i>echA2</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0632c | <i>echA3</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0673 | <i>echA4</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0675 | <i>echA5</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0905 | <i>echA6</i> | enoyl-CoA hydratase/isomerase superfamily (aka <i>ecchH</i>) |
| Rv0971c | <i>echA7</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1070c | <i>echA8</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1071c | <i>echA9</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1142c | <i>echA10</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1141c | <i>echA11</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1472 | <i>echA12</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv1935c | <i>echA13</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv2486 | <i>echA14</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv2679 | <i>echA15</i> | enoyl-CoA hydratase/isomerase superfamily |

| | | |
|---------|----------------|---|
| Rv2831 | <i>echA16</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv3039c | <i>echA17</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv3373 | <i>echA18</i> | enoyl-CoA hydratase/isomerase superfamily, N-term |
| Rv3374 | <i>echA18'</i> | enoyl-CoA hydratase/isomerase superfamily, C-term |
| Rv3516 | <i>echA19</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv3550 | <i>echA20</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv3774 | <i>echA21</i> | enoyl-CoA hydratase/isomerase superfamily |
| Rv0859 | <i>fadA</i> | β oxidation complex, β subunit (acetyl-CoA C-acetyltransferase) |
| Rv0243 | <i>fadA2</i> | acetyl-CoA C-acetyltransferase |
| Rv1074c | <i>fadA3</i> | acetyl-CoA C-acetyltransferase |
| Rv1323 | <i>fadA4</i> | acetyl-CoA C-acetyltransferase (aka <i>thiL</i>) |
| Rv3546 | <i>fadA5</i> | acetyl-CoA C-acetyltransferase |
| Rv3556c | <i>fadA6</i> | acetyl-CoA C-acetyltransferase |
| Rv0860 | <i>fadB</i> | β oxidation complex, α subunit (multiple activities) |
| Rv0468 | <i>fadB2</i> | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1715 | <i>fadB3</i> | 3-hydroxyacyl-CoA dehydrogenase |
| Rv3141 | <i>fadB4</i> | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1912c | <i>fadB5</i> | 3-hydroxyacyl-CoA dehydrogenase |
| Rv1750c | <i>fadD1</i> | acyl-CoA synthase |
| Rv0270 | <i>fadD2</i> | acyl-CoA synthase |
| Rv3561 | <i>fadD3</i> | acyl-CoA synthase |
| Rv0214 | <i>fadD4</i> | acyl-CoA synthase |
| Rv0166 | <i>fadD5</i> | acyl-CoA synthase |
| Rv1206 | <i>fadD6</i> | acyl-CoA synthase |
| Rv0119 | <i>fadD7</i> | acyl-CoA synthase |
| Rv0551c | <i>fadD8</i> | acyl-CoA synthase |
| Rv2590 | <i>fadD9</i> | acyl-CoA synthase |
| Rv0099 | <i>fadD10</i> | acyl-CoA synthase |
| Rv1550 | <i>fadD11</i> | acyl-CoA synthase, N-term |
| Rv1549 | <i>fadD11'</i> | acyl-CoA synthase, C-term |
| Rv1427c | <i>fadD12</i> | acyl-CoA synthase |
| Rv3089 | <i>fadD13</i> | acyl-CoA synthase |
| Rv1058 | <i>fadD14</i> | acyl-CoA synthase |
| Rv2187 | <i>fadD15</i> | acyl-CoA synthase |
| Rv0852 | <i>fadD16</i> | acyl-CoA synthase |
| Rv3506 | <i>fadD17</i> | acyl-CoA synthase |
| Rv3513c | <i>fadD18</i> | acyl-CoA synthase |
| Rv3515c | <i>fadD19</i> | acyl-CoA synthase |
| Rv1185c | <i>fadD21</i> | acyl-CoA synthase |
| Rv2948c | <i>fadD22</i> | acyl-CoA synthase |
| Rv3826 | <i>fadD23</i> | acyl-CoA synthase |
| Rv1529 | <i>fadD24</i> | acyl-CoA synthase |
| Rv1521 | <i>fadD25</i> | acyl-CoA synthase |
| Rv2930 | <i>fadD26</i> | acyl-CoA synthase |
| Rv0275c | <i>fadD27</i> | acyl-CoA synthase |
| Rv2941 | <i>fadD28</i> | acyl-CoA synthase |
| Rv2950c | <i>fadD29</i> | acyl-CoA synthase |
| Rv0404 | <i>fadD30</i> | acyl-CoA synthase |
| Rv1925 | <i>fadD31</i> | acyl-CoA synthase |
| Rv3801c | <i>fadD32</i> | acyl-CoA synthase |
| Rv1345 | <i>fadD33</i> | acyl-CoA synthase |
| Rv0035 | <i>fadD34</i> | acyl-CoA synthase |
| Rv2505c | <i>fadD35</i> | acyl-CoA synthase |
| Rv1193 | <i>fadD36</i> | acyl-CoA synthase |
| Rv0131c | <i>fadE1</i> | acyl-CoA dehydrogenase |
| Rv0154c | <i>fadE2</i> | acyl-CoA dehydrogenase |
| Rv0215c | <i>fadE3</i> | acyl-CoA dehydrogenase |
| Rv0231 | <i>fadE4</i> | acyl-CoA dehydrogenase |
| Rv0244c | <i>fadE5</i> | acyl-CoA dehydrogenase |
| Rv0271c | <i>fadE6</i> | acyl-CoA dehydrogenase |
| Rv0400c | <i>fadE7</i> | acyl-CoA dehydrogenase |
| Rv0672 | <i>fadE8</i> | acyl-CoA dehydrogenase (aka <i>aidB</i>) |
| Rv0752c | <i>fadE9</i> | acyl-CoA dehydrogenase |
| Rv0873 | <i>fadE10</i> | acyl-CoA dehydrogenase |
| Rv0972c | <i>fadE12</i> | acyl-CoA dehydrogenase |
| Rv0975c | <i>fadE13</i> | acyl-CoA dehydrogenase |
| Rv1346 | <i>fadE14</i> | acyl-CoA dehydrogenase |
| Rv1467c | <i>fadE15</i> | acyl-CoA dehydrogenase |
| Rv1679 | <i>fadE16</i> | acyl-CoA dehydrogenase |
| Rv1934c | <i>fadE17</i> | acyl-CoA dehydrogenase |
| Rv1933c | <i>fadE18</i> | acyl-CoA dehydrogenase |
| Rv2500c | <i>fadE19</i> | acyl-CoA dehydrogenase (aka <i>mmgC</i>) |
| Rv2724c | <i>fadE20</i> | acyl-CoA dehydrogenase |
| Rv2789c | <i>fadE21</i> | acyl-CoA dehydrogenase |
| Rv3061c | <i>fadE22</i> | acyl-CoA dehydrogenase |
| Rv3140 | <i>fadE23</i> | acyl-CoA dehydrogenase |
| Rv3139 | <i>fadE24</i> | acyl-CoA dehydrogenase |
| Rv3274c | <i>fadE25</i> | acyl-CoA dehydrogenase |
| Rv3504 | <i>fadE26</i> | acyl-CoA dehydrogenase |
| Rv3505 | <i>fadE27</i> | acyl-CoA dehydrogenase |
| Rv3544c | <i>fadE28</i> | acyl-CoA dehydrogenase |

| | | |
|---------|---------------|---------------------------------------|
| Rv3543c | <i>fadE29</i> | acyl-CoA dehydrogenase |
| Rv3560c | <i>fadE30</i> | acyl-CoA dehydrogenase |
| Rv3562 | <i>fadE31</i> | acyl-CoA dehydrogenase |
| Rv3563 | <i>fadE32</i> | acyl-CoA dehydrogenase |
| Rv3564 | <i>fadE33</i> | acyl-CoA dehydrogenase |
| Rv3573c | <i>fadE34</i> | acyl-CoA dehydrogenase |
| Rv3797 | <i>fadE35</i> | acyl-CoA dehydrogenase |
| Rv3761c | <i>fadE36</i> | acyl-CoA dehydrogenase |
| Rv1175c | <i>fadH</i> | 2,4-Dienoyl-CoA Reductase |
| Rv0855 | <i>far</i> | fatty acyl-CoA racemase |
| Rv1143 | <i>mcr</i> | α-methyl acyl-CoA racemase |
| Rv1492 | <i>mutA</i> | methylmalonyl-CoA mutase, β subunit |
| Rv1493 | <i>mutB</i> | methylmalonyl-CoA mutase, α subunit |
| Rv2504c | <i>scoA</i> | 3-oxo acid:CoA transferase, α subunit |
| Rv2503c | <i>scoB</i> | 3-oxo acid:CoA transferase, β subunit |
| Rv1136 | - | probable carnitine racemase |
| Rv1683 | - | possible acyl-CoA synthase |

4. Phosphorous compounds

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

B. Energy metabolism

1. Glycolysis

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

2. Pyruvate dehydrogenase

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

3. TCA cycle

| | | |
|---------|---------------|--------------------------------------|
| Rv1475c | <i>acn</i> | aconitate hydratase |
| Rv0889c | <i>citA</i> | citrate synthase 2 |
| Rv2498c | <i>citE</i> | citrate lyase β chain |
| Rv1098c | <i>fum</i> | fumarase |
| Rv1131 | <i>glitA1</i> | citrate synthase 3 |
| Rv0896 | <i>glitA2</i> | citrate synthase 1 |
| Rv3339c | <i>icd1</i> | isocitrate dehydrogenase |
| Rv0066c | <i>icd2</i> | isocitrate dehydrogenase |
| Rv0794c | <i>lpdB</i> | dihydrolipoamide dehydrogenase |
| Rv1240 | <i>mdh</i> | malate dehydrogenase |
| Rv2967c | <i>pca</i> | pyruvate carboxylase |
| Rv3318 | <i>sdhA</i> | succinate dehydrogenase A |
| Rv3319 | <i>sdhB</i> | succinate dehydrogenase B |
| Rv3316 | <i>sdhC</i> | succinate dehydrogenase C subunit |
| Rv3317 | <i>sdhD</i> | succinate dehydrogenase D subunit |
| Rv1248c | <i>sucA</i> | 2-oxoglutarate dehydrogenase |
| Rv2215 | <i>sucB</i> | dihydrolipoamide succinyltransferase |
| Rv0951 | <i>sucC</i> | succinyl-CoA synthase β chain |
| Rv0952 | <i>sucD</i> | succinyl-CoA synthase α chain |

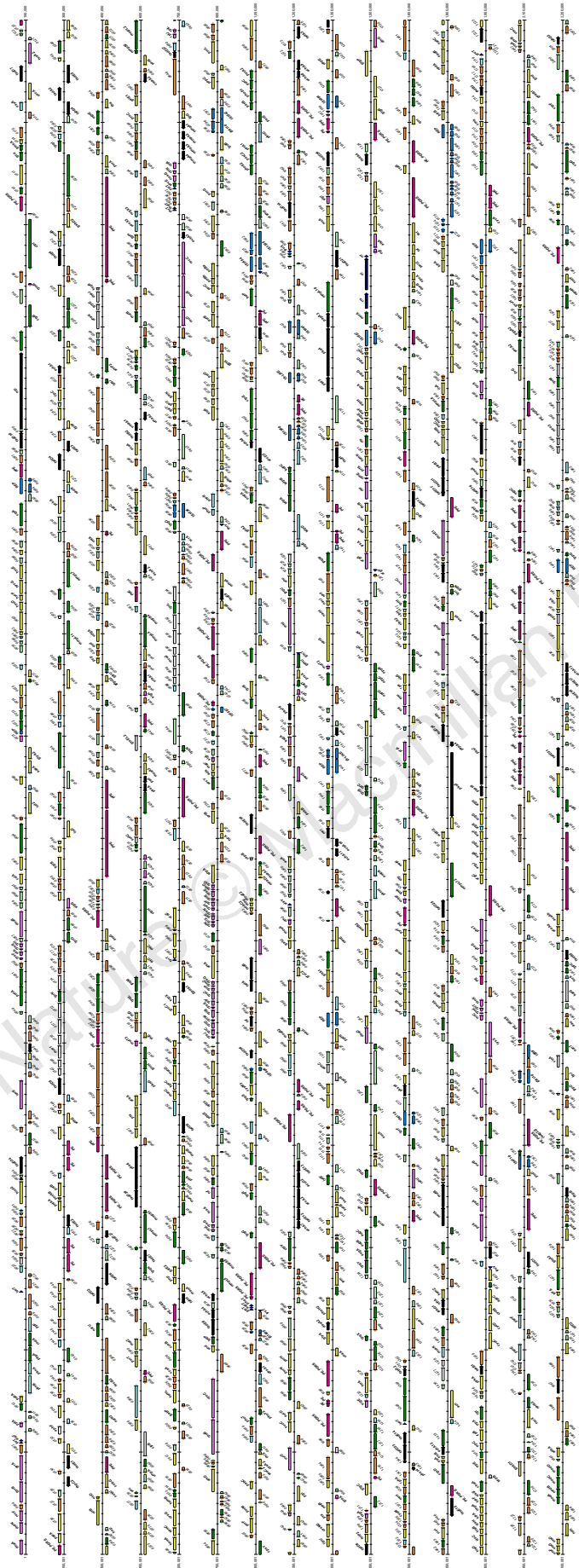
4. Glyoxylate bypass

| | | |
|---------|--------------|------------------------------|
| Rv0467 | <i>aceA</i> | isocitrate lyase |
| Rv1915 | <i>aceAa</i> | isocitrate lyase, α module |
| Rv1916 | <i>aceAb</i> | isocitrate lyase, β module |
| Rv1837c | <i>glcB</i> | malate synthase |
| Rv3233c | <i>gphA</i> | phosphoglycolate phosphatase |

5. Pentose phosphate pathway

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

| | | | | | | | | |
|---|--------------|--|---|--------------|--|---|--------------|---|
| Rv1605 | <i>hisF</i> | imidazole glycerol-phosphate synthase | Rv3048c | <i>nrdG</i> | subunit ribonucleoside-diphosphate small subunit | Rv3119 | <i>moaE</i> | subunit 1 molybdopterin-converting factor |
| Rv2121c | <i>hisG</i> | ATP phosphoribosyltransferase | Rv3053c | <i>nrdH</i> | glutaredoxin electron transport component of NrdEF system | Rv0866 | <i>moaE2</i> | molybdopterin-converting factor subunit 2 |
| Rv1602 | <i>hisH</i> | amidotransferase | Rv3052c | <i>nrdI</i> | NrdI/YgaO/YmaA family thymidylate kinase | Rv3322c | <i>moaE3</i> | molybdopterin-converting factor subunit 2 |
| Rv2122c | <i>hisI</i> | phosphoribosyl-AMP cyclohydro-lase | Rv3247c | <i>tmk</i> | thymidylate kinase | Rv0994 | <i>moaA</i> | molybdopterin biosynthesis |
| Rv1606 | <i>hisI2</i> | probable phosphoribosyl-AMP 1,6 cyclohydrolyase | Rv2764c | <i>thyA</i> | thymidylate synthase | Rv3116 | <i>moaB</i> | molybdopterin biosynthesis |
| Rv0114 | - | similar to HisB | Rv0570 | <i>nrdZ</i> | ribonucleotide reductase, class II | Rv2338c | <i>moaW</i> | molybdopterin biosynthesis |
| 6. Pyruvate family | | | Rv3752c | - | probable cytidine/deoxycytidylate deaminase | Rv1681 | <i>moaX</i> | weak similarity to <i>E. coli</i> MoaA |
| Rv3423c | <i>alr</i> | alanine racemase | 4. Salvage of nucleosides and nucleotides | | | Rv1355c | <i>moaY</i> | weak similarity to <i>E. coli</i> MoaB |
| 7. Branched amino acid family | | | Rv3313c | <i>add</i> | probable adenosine deaminase | Rv3206c | <i>moaZ</i> | probably involved in molybdopterin biosynthesis |
| Rv1559 | <i>ilvA</i> | threonine deaminase | Rv2584c | <i>apt</i> | adenine phosphoribosyltransferase | Rv0865 | <i>mog</i> | molybdopterin biosynthesis |
| Rv3003c | <i>ilvB</i> | acetolactate synthase I large subunit | Rv3315c | <i>cdd</i> | probable cytidine deaminase | 5. Pantothenate | | |
| Rv3470c | <i>ilvB2</i> | acetolactate synthase large subunit | Rv3314c | <i>deoA</i> | thymidine phosphorylase | Rv1092c | <i>coaA</i> | pantothenate kinase |
| Rv3001c | <i>ilvC</i> | ketol-acid reductoisomerase | Rv0478 | <i>deoC</i> | deoxyribose-phosphate aldolase | Rv2225 | <i>panB</i> | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| Rv0189c | <i>ilvD</i> | dihydroxy-acid dehydratase | Rv3307 | <i>deoD</i> | probable purine nucleoside phosphorylase | Rv3602c | <i>panC</i> | 3-methyl-2-oxobutanoate hydroxymethyltransferase |
| Rv2210c | <i>ilvE</i> | branched-chain-amino-acid transaminase | Rv3624c | <i>hpt</i> | probable hypoxanthine-guanine phosphoribosyltransferase | Rv3601c | <i>panD</i> | aspartate 1-decarboxylase |
| Rv1820 | <i>ilvG</i> | acetolactate synthase II | Rv3393 | <i>iunH</i> | probable inosine-uridine preferring nucleoside hydrolase | 6. Pyridoxine | | |
| Rv3002c | <i>ilvN</i> | acetolactate synthase I small subunit | Rv0535 | <i>pnp</i> | phosphorylase from Pnp/MtaP family 2 | Rv2607 | <i>pdxH</i> | pyridoxamine 5'-phosphate oxidase |
| Rv3509c | <i>ilvX</i> | probable acetohydroxyacid synthase I large subunit | Rv3309c | <i>upp</i> | uracil phosphoribosyltransferase | 7. Pyridine nucleotide | | |
| Rv3710 | <i>leuA</i> | α -isopropyl malate synthase | 5. Miscellaneous nucleoside/nucleotide reactions | | | Rv1594 | <i>nadA</i> | quinolinate synthase |
| Rv2995c | <i>leuB</i> | 3-isopropylmalate dehydrogenase | Rv0733 | <i>adk</i> | probable adenylate kinase | Rv1595 | <i>nadB</i> | L-aspartate oxidase |
| Rv2988c | <i>leuC</i> | 3-isopropylmalate dehydratase large subunit | Rv2364c | <i>bex</i> | GTP-binding protein of Era/ThdF family | Rv1596 | <i>nadC</i> | nicotinate-nucleotide pyrophosphatase |
| Rv2987c | <i>leuD</i> | 3-isopropylmalate dehydratase small subunit | Rv1712 | <i>cmk</i> | cytidylate kinase | Rv0423c | <i>thiC</i> | thiamine synthesis, pyrimidine moiety |
| Rv2344c | <i>dgt</i> | probable deoxyguanosine triphosphate hydrolase | Rv2404c | <i>lepA</i> | GTP-binding protein LepA | 8. Thiamine | | |
| Rv2727c | <i>miaA</i> | tRNA δ (2)-isopentenylpyrophosphate transferase | Rv2727c | <i>miaA</i> | tRNA δ (2)-isopentenylpyrophosphate transferase | Rv0422c | <i>thiD</i> | phosphomethylpyrimidine kinase |
| Rv2445c | <i>ndkA</i> | nucleoside diphosphate kinase | Rv2445c | <i>ndkA</i> | nucleoside diphosphate kinase | Rv0414c | <i>thiE</i> | thiamine synthesis, thiazole moiety |
| Rv2440c | <i>obg</i> | Obg GTP-binding protein | Rv2583c | <i>relA</i> | Obg GTP-binding protein | Rv0417 | <i>thiG</i> | thiamine synthesis, thiazole moiety |
| Rv2583c | <i>relA</i> | (p)ppGpp synthase I | | | | Rv2977c | <i>thiL</i> | probable thiamine-monophosphate kinase |
| <i>E. Polyamine synthesis</i> | | | <i>G. Biosynthesis of cofactors, prosthetic groups and carriers</i> | | | | | |
| Rv2601 | <i>speE</i> | spermidine synthase | 1. Biotin | | | 9. Riboflavin | | |
| <i>F. Purines, pyrimidines, nucleosides and nucleotides</i> | | | Rv1568 | <i>bioA</i> | adenosylmethionine-8-amino-7-oxononanoate aminotransferase | Rv1940 | <i>ribA</i> | GTP cyclohydrolase II |
| 1. Purine ribonucleotide biosynthesis | | | Rv1589 | <i>bioB</i> | oxiotin synthase | Rv1415 | <i>ribA2</i> | probable GTP cyclohydrolase II |
| Rv1389 | <i>gmk</i> | putative guanylate kinase | Rv1570 | <i>bioD</i> | dethiobiotin synthase | Rv1412 | <i>ribC</i> | riboflavin synthase α chain |
| Rv3396c | <i>guaA</i> | GMP synthase | Rv1569 | <i>bioF</i> | 8-amino-7-oxononanoate synthase | Rv2671 | <i>ribD</i> | probable riboflavin deaminase |
| Rv1843c | <i>guaB1</i> | inosine-5'-monophosphate dehydrogenase | Rv0032 | <i>bioF2</i> | C-terminal similar to <i>B. subtilis</i> BioF | Rv2786c | <i>ribF</i> | riboflavin kinase |
| Rv3411c | <i>guaB2</i> | inosine-5'-monophosphate dehydrogenase | Rv3279c | <i>birA</i> | biotin apo-protein ligase | Rv1409 | <i>ribG</i> | riboflavin biosynthesis |
| Rv3410c | <i>guaB3</i> | inosine-5'-monophosphate dehydrogenase | Rv1442 | <i>bisC</i> | biotin sulfoxide reductase | Rv1416 | <i>ribH</i> | riboflavin synthase β chain |
| Rv1017c | <i>prsA</i> | ribose-phosphate pyrophosphokinase | Rv0089 | - | possible <i>bioC</i> biotin synthesis gene | Rv3300c | - | probable deaminase, riboflavin synthesis |
| Rv0357c | <i>purA</i> | adenylosuccinate synthase | 2. Folic acid | | | 10. Thioredoxin, glutaredoxin and mycothiol | | |
| Rv0777 | <i>purB</i> | adenylosuccinate lyase | Rv2763c | <i>dfrA</i> | dihydrofolate reductase | Rv0773c | <i>ggtA</i> | putative γ -glutamyl transpeptidase |
| Rv0780 | <i>purC</i> | phosphoribosylaminoimidazole-succinocarboxamide synthase | Rv2447c | <i>folC</i> | folypolyglutamate synthase | Rv2394 | <i>ggtB</i> | γ -glutamyltranspeptidase precursor |
| Rv0772 | <i>purD</i> | phosphoribosylamine-glycine ligase | Rv3356c | <i>folD</i> | methylene tetrahydrofolate dehydrogenase | Rv2855 | <i>gorA</i> | glutathione reductase homologue |
| Rv3275c | <i>purE</i> | phosphoribosylaminoimidazole carboxylase | Rv3609c | <i>folE</i> | GTP cyclohydrolase I | Rv0816c | <i>thiX</i> | equivalent to <i>M. leprae</i> ThiX |
| Rv0808 | <i>purF</i> | amidophosphoribosyltransferase | Rv3606c | <i>folK</i> | 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase | Rv1470 | <i>trxA</i> | thioredoxin |
| Rv0957 | <i>purH</i> | phosphoribosylaminoimidazole-carboxamide formyltransferase | Rv3608c | <i>folP</i> | dihydropterate synthase | Rv1471 | <i>trxB</i> | thioredoxin reductase |
| Rv3276c | <i>purK</i> | phosphoribosylaminoimidazole carboxylase ATPase subunit | Rv1207 | <i>folP2</i> | dihydropterate synthase | Rv3913 | <i>trxB2</i> | thioredoxin reductase |
| Rv0803 | <i>purL</i> | phosphoribosylformylglycinamide synthase II | Rv3607c | <i>folX</i> | may be involved in folate biosynthesis | Rv3914 | <i>trxC</i> | thioredoxin |
| Rv0809 | <i>purM</i> | 5'-phosphoribosyl-5-aminoimidazole synthase | Rv0013 | <i>pabA</i> | <i>p</i> -aminobenzoate synthase | 11. Menaquinone, PQQ, ubiquinone and other terpenoids | | |
| Rv0956 | <i>purN</i> | phosphoribosylglycinamide formyltransferase I | Rv1005c | <i>pabB</i> | <i>p</i> -aminobenzoate synthase | Rv2682c | <i>dxs</i> | 1-deoxy-D-xylulose 5-phosphate synthase |
| Rv0788 | <i>purQ</i> | phosphoribosylformylglycinamide synthase I | Rv0812 | <i>pabC</i> | aminodeoxychorismate lyase | Rv0562 | <i>grcC1</i> | heptaprenyl diphosphate synthase II |
| Rv0389 | <i>purT</i> | phosphoribosylglycinamide formyltransferase II | | | | Rv0989c | <i>grcC2</i> | heptaprenyl diphosphate synthase II |
| Rv2964 | <i>purU</i> | formyltetrahydrofolate deformylase | | | | Rv3398c | <i>idsA</i> | geranylgeranyl pyrophosphate synthase |
| 2. Pyrimidine ribonucleotide biosynthesis | | | 3. Lipoate | | | Rv2173 | <i>idsA2</i> | geranylgeranyl pyrophosphate synthase |
| Rv1383 | <i>carA</i> | carbamoyl-phosphate synthase subunit | Rv2218 | <i>lipA</i> | lipoate biosynthesis protein A | Rv3383c | <i>idsB</i> | transfergeranyl, similar geranyl pyrophosphate synthase |
| Rv1384 | <i>carB</i> | carbamoyl-phosphate synthase subunit | Rv2217 | <i>lipB</i> | lipoate biosynthesis protein B | Rv0534c | <i>menA</i> | pyrophosphate synthase |
| Rv1380 | <i>pyrB</i> | aspartate carbamoyltransferase | 4. Molybdopterin | | | Rv0548c | <i>menB</i> | 4-dihydroxy-2-naphthoate octaprenyltransferase |
| Rv1381 | <i>pyrC</i> | dihydroorotase | Rv3109 | <i>moaA</i> | molybdenum cofactor biosynthesis, protein A | Rv0553 | <i>menC</i> | naphthoate synthase |
| Rv2139 | <i>pyrD</i> | dihydroorotase dehydrogenase | Rv0869c | <i>moaA2</i> | molybdenum cofactor biosynthesis, protein A | Rv0555 | <i>menD</i> | <i>o</i> -succinylbenzoate-CoA synthase |
| Rv1385 | <i>pyrF</i> | orotidine 5'-phosphate decarboxylase | Rv0438c | <i>moaA3</i> | molybdenum cofactor biosynthesis, protein A | Rv0542c | <i>menE</i> | 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase |
| Rv1699 | <i>pyrG</i> | CTP synthase | Rv3110 | <i>moaB</i> | molybdenum cofactor biosynthesis, protein B | Rv3853 | <i>menG</i> | <i>o</i> -succinylbenzoic acid-CoA ligase |
| Rv2883c | <i>pyrH</i> | uridylylase | Rv0984 | <i>moaB2</i> | molybdenum cofactor biosynthesis, protein B | Rv3397c | <i>phyA</i> | <i>S</i> -adenosylmethionine: 2-demethylmenaquinone phytoene synthase |
| Rv0382c | <i>umpA</i> | probable uridine 5'-monophosphate synthase | Rv3111 | <i>moaC</i> | molybdenum cofactor biosynthesis, protein C | Rv0693 | <i>pqqE</i> | coenzyme PQQ synthesis protein E |
| 3. 2'-deoxyribonucleotide metabolism | | | Rv0864 | <i>moaC2</i> | molybdenum cofactor biosynthesis, protein C | Rv0558 | <i>ubiE</i> | ubiquinone/menaquinone biosynthesis methyltransferase |
| Rv0321 | <i>dcd</i> | deoxycytidine triphosphate deaminase | Rv3324c | <i>moaC3</i> | molybdenum cofactor biosynthesis, protein C | 12. Heme and porphyrin | | |
| Rv2697c | <i>dut</i> | deoxyuridine triphosphatase | Rv3112 | <i>moaD</i> | molybdopterin converting factor subunit 1 | Rv0509 | <i>hema</i> | glutamyl-tRNA reductase |
| Rv0233 | <i>nrdB</i> | ribonucleoside-diphosphate reductase B2 (eukaryotic-like) | Rv0868c | <i>moaD2</i> | molybdopterin converting factor | Rv0512 | <i>hemB</i> | δ -aminolevulinic acid dehydratase |
| Rv3051c | <i>nrdE</i> | ribonucleoside diphosphate reductase α chain | | | | Rv0510 | <i>hemC</i> | porphobilinogen deaminase |
| Rv1981c | <i>nrdF</i> | ribonucleotide reductase small | | | | Rv2678c | <i>hemE</i> | uroporphyrinogen decarboxylase |





| | | | | | | | | |
|---------|---|---|--|--------------|--|---|-------------|---|
| Rv0823c | - | family) transcriptional regulator (NifR3/Smm1 family) | Rv3160c | - | putative transcriptional regulator | Rv0018c | <i>ppp</i> | truncated putative phosphoprotein phosphatase |
| Rv0827c | - | transcriptional regulator (ArsR family) | Rv3167c | - | putative transcriptional regulator | Rv2234 | <i>ptpA</i> | low molecular weight protein-tyrosine-phosphatase |
| Rv0890c | - | transcriptional regulator (LuxR/UhpA family) | Rv3173c | - | transcriptional regulator (TetR/AcrR family) | Rv0153c | - | putative protein-tyrosine-phosphatase |
| Rv0891c | - | putative transcriptional regulator | Rv3183 | - | putative transcriptional regulator | II. Macromolecule metabolism | | |
| Rv0894 | - | putative transcriptional regulator | Rv3208 | - | transcriptional regulator (TetR/AcrR family) | | | |
| Rv1019 | - | transcriptional regulator (TetR/AcrR family) | Rv3249c | - | transcriptional regulator (TetR/AcrR family) | 1. Ribosomal protein synthesis and modification | | |
| Rv1049 | - | transcriptional regulator (MarR family) | Rv3291c | - | transcriptional regulator (Lrp/AsnC family) | | | |
| Rv1129c | - | transcriptional regulator (PbsX/Xre family) | Rv3295 | - | transcriptional regulator (TetR/AcrR family) | ribosomal protein S18 acetyl transferase | | |
| Rv1151c | - | putative transcriptional regulator | Rv3334 | - | transcriptional regulator (MerR family) | | | |
| Rv1152 | - | transcriptional regulator (GntR family) | Rv3405c | - | putative transcriptional regulator | 50S ribosomal protein L1 | | |
| Rv1167c | - | putative transcriptional regulator | Rv3522 | - | putative transcriptional regulator | | | |
| Rv1219c | - | putative transcriptional regulator | Rv3557c | - | transcriptional regulator (TetR/AcrR family) | 50S ribosomal protein L3 | | |
| Rv1255c | - | transcriptional regulator (TetR/AcrR family) | Rv3574 | - | transcriptional regulator (TetR/AcrR family) | | | |
| Rv1332 | - | putative transcriptional regulator | Rv3575c | - | transcriptional regulator (LacI family) | 50S ribosomal protein L5 | | |
| Rv1353c | - | transcriptional regulator (TetR/AcrR family) | Rv3583c | - | putative transcriptional regulator | | | |
| Rv1358 | - | transcriptional regulator (LuxR/UhpA family) | Rv3676 | - | transcriptional regulator (Crp/Fnr family) | 50S ribosomal protein L9 | | |
| Rv1359 | - | putative transcriptional regulator | Rv3678c | - | transcriptional regulator (LysR family) | | | |
| Rv1395 | - | transcriptional regulator (AraC/XylS family) | Rv3736 | - | transcriptional regulator (AraC/XylS family) | 50S ribosomal protein L11 | | |
| Rv1404 | - | transcriptional regulator (MarR family) | Rv3744 | - | transcriptional regulator (ArsR family) | | | |
| Rv1423 | - | putative transcriptional regulator | Rv3830c | - | transcriptional regulator (TetR/AcrR family) | 50S ribosomal protein L7/L13 | | |
| Rv1460 | - | putative transcriptional regulator | Rv3833 | - | transcriptional regulator (AraC/XylS family) | | | |
| Rv1474c | - | transcriptional regulator (TetR/AcrR family) | Rv3840 | - | putative transcriptional regulator | 50S ribosomal protein L15 | | |
| Rv1534 | - | transcriptional regulator (TetR/AcrR family) | Rv3855 | - | putative transcriptional regulator | | | |
| Rv1556 | - | putative transcriptional regulator | 2. Two component systems | | | 50S ribosomal protein L17 | | |
| Rv1674c | - | putative transcriptional regulator | Rv1028c | <i>kdpD</i> | sensor histidine kinase | | | |
| Rv1675c | - | putative transcriptional regulator | Rv1027c | <i>kdpE</i> | two-component response regulator | 50S ribosomal protein L19 | | |
| Rv1719 | - | transcriptional regulator (IclR family) | Rv3246c | <i>mtrA</i> | two-component response regulator | | | |
| Rv1773c | - | transcriptional regulator (IclR family) | Rv3245c | <i>mtrB</i> | sensor histidine kinase | 50S ribosomal protein L21 | | |
| Rv1776c | - | putative transcriptional regulator | Rv0844c | <i>narL</i> | two-component response regulator | | | |
| Rv1816 | - | putative transcriptional regulator | Rv0757 | <i>phoP</i> | two-component response regulator | 50S ribosomal protein L23 | | |
| Rv1846c | - | putative transcriptional regulator | Rv0758 | <i>phoR</i> | sensor histidine kinase | | | |
| Rv1931c | - | transcriptional regulator (AraC/XylS family) | Rv0491 | <i>regX3</i> | two-component response regulator | 50S ribosomal protein L25 | | |
| Rv1956 | - | putative transcriptional regulator | Rv0490 | <i>senX3</i> | sensor histidine kinase | | | |
| Rv1963c | - | putative transcriptional regulator | Rv0602c | <i>trcA</i> | two-component response regulator | 50S ribosomal protein L28 | | |
| Rv1985c | - | transcriptional regulator (LysR family) | Rv0260c | - | two-component response regulator | | | |
| Rv1990c | - | putative transcriptional regulator | Rv0600c | - | sensor histidine kinase | 50S ribosomal protein L30 | | |
| Rv1994c | - | transcriptional regulator (MerR family) | Rv0601c | - | sensor histidine kinase | | | |
| Rv2017 | - | putative transcriptional regulator (PbsX/Xre family) | Rv0818 | - | two-component response regulator | 50S ribosomal protein L33 | | |
| Rv2021c | - | putative transcriptional regulator | Rv0845 | - | sensor histidine kinase | | | |
| Rv2034 | - | transcriptional regulator (ArsR family) | Rv0902c | - | sensor histidine kinase | 50S ribosomal protein L35 | | |
| Rv2175c | - | putative transcriptional regulator | Rv0903c | - | two-component response regulator | | | |
| Rv2250c | - | putative transcriptional regulator | Rv0981 | - | two-component response regulator | 30S ribosomal protein S1 | | |
| Rv2258c | - | putative transcriptional regulator | Rv0982 | - | sensor histidine kinase | | | |
| Rv2282c | - | transcriptional regulator (LysR family) | Rv1032c | - | sensor histidine kinase | 30S ribosomal protein S3 | | |
| Rv2308 | - | putative transcriptional regulator | Rv1033c | - | two-component response regulator | | | |
| Rv2324 | - | transcriptional regulator (Lrp/AsnC family) | Rv1626 | - | two-component response regulator | 30S ribosomal protein S5 | | |
| Rv2358 | - | transcriptional regulator (ArsR family) | Rv2027c | - | sensor histidine kinase | | | |
| Rv2488c | - | transcriptional regulator (LuxR/UhpA family) | Rv2884 | - | two-component response regulator | 30S ribosomal protein S7 | | |
| Rv2506 | - | transcriptional regulator (TetR/AcrR family) | Rv3132c | - | sensor histidine kinase | | | |
| Rv2621c | - | putative transcriptional regulator | Rv3133c | - | two-component response regulator | 30S ribosomal protein S9 | | |
| Rv2640c | - | transcriptional regulator (ArsR family) | Rv3143 | - | putative sensory transduction protein | | | |
| Rv2642 | - | transcriptional regulator (ArsR family) | Rv3220c | - | sensor histidine kinase | 30S ribosomal protein S11 | | |
| Rv2669 | - | putative transcriptional regulator | Rv3764c | - | sensor histidine kinase | | | |
| Rv2745c | - | putative transcriptional regulator | Rv3765c | - | two-component response regulator | 30S ribosomal protein S13 | | |
| Rv2779c | - | transcriptional regulator (Lrp/AsnC family) | 3. Serine-threonine protein kinases and phosphoprotein phosphatases | | | | | |
| Rv2887 | - | transcriptional regulator (MarR family) | Rv0015c | <i>pknA</i> | serine-threonine protein kinase | 30S ribosomal protein S15 | | |
| Rv2912c | - | transcriptional regulator (TetR/AcrR family) | Rv0014c | <i>pknB</i> | serine-threonine protein kinase | | | |
| Rv2989 | - | transcriptional regulator (IclR family) | Rv0931c | <i>pknD</i> | serine-threonine protein kinase | 30S ribosomal protein S17 | | |
| Rv3050c | - | putative transcriptional regulator | Rv1743 | <i>pknE</i> | serine-threonine protein kinase | | | |
| Rv3055 | - | putative transcriptional regulator | Rv1746 | <i>pknF</i> | serine-threonine protein kinase | 30S ribosomal protein S19 | | |
| Rv3058c | - | putative transcriptional regulator | Rv0410c | <i>pknG</i> | serine-threonine protein kinase | | | |
| Rv3060c | - | transcriptional regulator (GntR family) | Rv1266c | <i>pknH</i> | serine-threonine protein kinase | member of S30AE ribosomal protein family | | |
| Rv3066 | - | putative transcriptional regulator | Rv2914c | <i>pknI</i> | serine-threonine protein kinase | | | |
| Rv3095 | - | putative transcriptional regulator | Rv2088 | <i>pknJ</i> | serine-threonine protein kinase | Rv1010 <i>ksgA</i> | | |
| Rv3124 | - | transcriptional regulator (AisR/DndI/RedD family) | Rv3080c | <i>pknK</i> | serine-threonine protein kinase | | | |
| | | | Rv2176 | <i>pknL</i> | serine-threonine protein kinase, | Rv2838c <i>rbfA</i> | | |
| | | | | | | | | |
| | | | | | | 3. Aminoacyl tRNA synthetases and their modification | | |
| | | | | | | | | |
| | | | | | | alanyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | arginyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | aspartyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | cysteinyI-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | cysteinyI-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | methionyl-tRNA formyltransferase | | |
| | | | | | | | | |
| | | | | | | glu-tRNA-gln amidotransferase, subunit B | | |
| | | | | | | | | |
| | | | | | | glu-tRNA-gln amidotransferase, subunit B | | |
| | | | | | | | | |
| | | | | | | glu-tRNA-gln amidotransferase, subunit C | | |
| | | | | | | | | |
| | | | | | | glutamyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | glycyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | histidyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | isoleucyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | leucyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | lysyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | C-term lysyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | methionyl-tRNA synthase | | |
| | | | | | | | | |
| | | | | | | phenylalanyl-tRNA synthase α subunit | | |

| | | | | | | | |
|--|--------------|--|--|--------------|---|---|--|
| Rv1650 | <i>pheT</i> | phenylalanyl-tRNA synthase β subunit | Rv2090 | - | partially similar to DNA polymerase I | 2. DNA | |
| Rv2845c | <i>proS</i> | prolyl-tRNA synthase | Rv2191 | - | similar to both PolC and UvrC proteins | Rv0670 | <i>end</i> endonuclease IV (apurinase) |
| Rv3834c | <i>serS</i> | seryl-tRNA synthase | | | | Rv1108c | <i>xseA</i> exonuclease VII large subunit |
| Rv2614c | <i>thrS</i> | threonyl-tRNA synthase | Rv2464c | - | probable DNA glycosylase, endonuclease VIII | Rv1107c | <i>xseB</i> exonuclease VII small subunit |
| Rv2906c | <i>trmD</i> | tRNA (guanine-N1)-methyltransferase | Rv3201c | - | probable ATP-dependent DNA helicase | 3. Proteins, peptides and glycopeptides | |
| Rv3336c | <i>trpS</i> | tryptophanyl tRNA synthase | Rv3202c | - | similar to UvrD proteins | Rv3305c | <i>amiA</i> probable aminohydrolase |
| Rv1689 | <i>tyrS</i> | tyrosyl-tRNA synthase | Rv3263 | - | probable DNA methylase | Rv3306c | <i>amiB</i> probable aminohydrolase |
| Rv2448c | <i>valS</i> | valyl-tRNA synthase | Rv3644c | - | similar in N-term to DNA polymerase III | Rv3596c | <i>clpC</i> ATP-dependent Clp protease |
| | | | | | | Rv2461c | <i>clpP</i> ATP-dependent Clp protease proteolytic subunit |
| 4. Nucleoproteins | | | | | | Rv2460c | <i>clpP2</i> ATP-dependent Clp protease proteolytic subunit |
| Rv1407 | <i>fmu</i> | similar to Fmu protein | | | | Rv2457c | <i>clpX</i> ATP-dependent Clp protease |
| Rv3852 | <i>hns</i> | HU-histone protein | 6. Protein translation and modification | | | | ATP-binding subunit ClpX |
| Rv2986c | <i>hupB</i> | DNA-binding protein II | Rv0429c | <i>def</i> | polypeptide deformylase | Rv2667 | <i>clpX'</i> similar to ClpC from <i>M. leprae</i> but shorter |
| Rv1388 | <i>mIHF</i> | integration host factor | Rv2534c | <i>efp</i> | elongation factor P | | |
| 5. DNA replication, repair, recombination and restriction/modification | | | Rv2882c | <i>frr</i> | ribosome recycling factor | Rv3419c | <i>gcp</i> glycoprotease |
| Rv1317c | <i>alkA</i> | DNA-3-methyladenine glycosidase II | Rv0684 | <i>fusA</i> | elongation factor G | Rv2725c | <i>hflX</i> GTP-binding protein |
| Rv2836c | <i>dinF</i> | DNA-damage-inducible protein F | Rv0120c | <i>fusA2</i> | elongation factor G | Rv1223 | <i>htrA</i> serine protease |
| Rv1329c | <i>dinG</i> | probable ATP-dependent helicase | Rv1080c | <i>greA</i> | transcription elongation factor G | Rv2861c | <i>map</i> probable methionine aminopeptidase |
| Rv3056 | <i>dinP</i> | DNA-damage-inducible protein | Rv3462c | <i>infA</i> | initiation factor IF-1 | Rv0734 | <i>map'</i> probable methionine aminopeptidase |
| Rv1537 | <i>dinX</i> | probable DNA-damage-inducible protein | Rv2839c | <i>infB</i> | initiation factor IF-2 | | |
| Rv0001 | <i>dnaA</i> | chromosomal replication initiator protein | Rv1641 | <i>infC</i> | initiation factor IF-3 | Rv0319 | <i>pcp</i> pyrrolidone-carboxylate peptidase |
| Rv0058 | <i>dnaB</i> | DNA helicase (contains intein) | Rv2582 | <i>ppiA</i> | peptidyl-prolyl <i>cis-trans</i> isomerase | Rv0125 | <i>pepA</i> probable serine protease |
| Rv1547 | <i>dnaE1</i> | DNA polymerase III, α subunit | Rv1299 | <i>prfA</i> | peptide chain release factor 1 | Rv2213 | <i>pepB</i> aminopeptidase A/I |
| Rv3370c | <i>dnaE2</i> | DNA polymerase III α chain | Rv3105c | <i>prfB</i> | peptide chain release factor 2 | Rv0800 | <i>pepC</i> aminopeptidase I |
| Rv2343c | <i>dnaG</i> | DNA primase | Rv2889c | <i>tsf</i> | elongation factor EF-Ts | Rv2467 | <i>pepD</i> probable aminopeptidase |
| Rv0002 | <i>dnaN</i> | DNA polymerase III, β subunit | Rv0685 | <i>tuf</i> | elongation factor EF-Tu | Rv2089c | <i>pepE</i> cytoplasmic peptidase |
| Rv3711c | <i>dnaQ</i> | DNA polymerase III ϵ chain | 7. RNA synthesis, RNA modification and DNA transcription | | | Rv2535c | <i>pepQ</i> cytoplasmic peptidase |
| Rv3721c | <i>dnaX</i> | DNA polymerase III, γ (dnaZ) and τ (dnaX) | Rv1253 | <i>deaD</i> | ATP-dependent DNA/RNA helicase | Rv2782c | <i>pepR</i> protease/peptidase, M16 family (insulinase) |
| Rv2924c | <i>fpg</i> | formamidopyrimidine-DNA glycosylase | Rv2783c | <i>gpsI</i> | pppGpp synthase and polynucleotide phosphorylase | Rv2109c | <i>prcA</i> proteasome α -type subunit 1 |
| Rv0006 | <i>gyrA</i> | DNA gyrase subunit A | Rv2841c | <i>nusA</i> | transcription termination factor | Rv2110c | <i>prcB</i> proteasome β -type subunit 2 |
| Rv0005 | <i>gyrB</i> | DNA gyrase subunit B | Rv2533c | <i>nusB</i> | N-utilization substance protein B | Rv0782 | <i>ptrBa</i> protease II, α subunit |
| Rv2092c | <i>heliY</i> | probable helicase, Ski2 subfamily | Rv0639 | <i>nusG</i> | transcription antitermination protein | Rv0781 | <i>ptrBb</i> protease II, β subunit |
| Rv2101 | <i>heliZ</i> | probable helicase, Snf2/Rad54 family | Rv3907c | <i>pcnA</i> | polynucleotide polymerase | Rv0724 | <i>sppA</i> protease IV, signal peptide peptidase |
| Rv2756c | <i>hsdM</i> | type I restriction/modification system DNA methylase | Rv3232c | <i>pvdS</i> | alternative sigma factor for siderophore production | Rv0198c | - probable zinc metalloprotease |
| Rv2755c | <i>hsdS'</i> | type I restriction/modification system specificity determinant | Rv3211 | <i>rhlE</i> | probable ATP-dependent RNA helicase | Rv0457c | - probable peptidase |
| Rv3296 | <i>lhr</i> | ATP-dependent helicase | Rv1297 | <i>rho</i> | transcription termination factor rho | Rv0840c | - probable proline iminopeptidase |
| Rv3014c | <i>ligA</i> | DNA ligase | Rv3457c | <i>rpoA</i> | α subunit of RNA polymerase | Rv0983 | - probable serine protease |
| Rv3062 | <i>ligB</i> | DNA ligase | Rv0667 | <i>rpoB</i> | β subunit of RNA polymerase | Rv1977 | - probable zinc metallopeptidase |
| Rv3731 | <i>ligC</i> | probable DNA ligase | Rv0668 | <i>rpoC</i> | β' subunit of RNA polymerase | Rv3668c | - probable alkaline serine protease |
| Rv1020 | <i>mfd</i> | transcription-repair coupling factor | Rv1364c | <i>rsbU</i> | SigB regulation protein | Rv3671c | - probable serine protease |
| Rv2528c | <i>mrr</i> | restriction system protein | Rv3287c | <i>rsbW</i> | anti-sigma B factor | Rv3883c | - probable secreted protease |
| Rv2985 | <i>mutT1</i> | MutT homologue | Rv2703 | <i>sigA</i> | RNA polymerase sigma factor (aka MysA, RpoV) | Rv3886c | - protease |
| Rv1160 | <i>mutT2</i> | MutT homologue | Rv2710 | <i>sigB</i> | RNA polymerase sigma factor (aka MysB) | 4. Polysaccharides, lipopolysaccharides and phospholipids | |
| Rv0413 | <i>mutT3</i> | MutT homologue | Rv2069 | <i>sigC</i> | ECF subfamily sigma subunit | Rv0062 | <i>celA</i> cellulase/endoglucanase |
| Rv3589 | <i>mutY</i> | probable DNA glycosylase | Rv3414c | <i>sigD</i> | ECF subfamily sigma subunit | Rv3915 | <i>cwIM</i> hydrolase |
| Rv3297 | <i>nei</i> | probable endonuclease VIII | Rv1221 | <i>sigE</i> | ECF subfamily sigma subunit | Rv0315 | - probable β -1,3-glucanase |
| Rv3674c | <i>nth</i> | probable endonuclease III | Rv3286c | <i>sigF</i> | ECF subfamily sigma subunit | Rv1090 | - probable inactivated cellulase/endoglucanase |
| Rv1316c | <i>ogt</i> | methylated-DNA-protein-cysteine methyltransferase | Rv0182c | <i>sigG</i> | sigma-70 factors ECF subfamily | Rv1327c | - probable glycosyl hydrolase, α -amylase family |
| Rv1629 | <i>polA</i> | DNA polymerase I | Rv3223c | <i>sigH</i> | ECF subfamily sigma subunit | Rv1333 | - probable hydrolase |
| Rv1402 | <i>priA</i> | putative primosomal protein n' (replication factor Y) | Rv1189 | <i>sigI</i> | ECF family sigma factor | Rv3463 | - probable neuroaminidase |
| Rv3585 | <i>radA</i> | probable DNA repair RadA homologue | Rv3328c | <i>sigJ</i> | similar to SigI, ECF family | Rv3717 | - possible N-acetylmuramoyl-L-alanine amidase |
| Rv2737c | <i>recA</i> | recombinase (contains intein) | Rv0445c | <i>sigK</i> | ECF-type sigma factor | 5. Esterases and lipases | |
| Rv0630c | <i>recB</i> | exodeoxyribonuclease V | Rv0735 | <i>sigL</i> | sigma-70 factors ECF subfamily | Rv0220 | <i>lipC</i> probable esterase |
| Rv0631c | <i>recC</i> | exodeoxyribonuclease V | Rv3911 | <i>sigM</i> | probable sigma factor, similar to SigE | Rv1923 | <i>lipD</i> probable esterase |
| Rv0629c | <i>recD</i> | exodeoxyribonuclease V | Rv3366 | <i>spoU</i> | probable rRNA methylase | Rv3775 | <i>lipE</i> probable hydrolase |
| Rv0003 | <i>recF</i> | DNA replication and SOS induction | Rv3455c | <i>truA</i> | probable pseudouridylylase | Rv3487c | <i>lipF</i> probable esterase |
| Rv2973c | <i>recG</i> | ATP-dependent DNA helicase | Rv2793c | <i>truB</i> | tRNA pseudouridine 55 synthase | Rv0646c | <i>lipG</i> probable hydrolase |
| Rv1696 | <i>recN</i> | recombination and DNA repair | Rv1644 | <i>tsnR</i> | putative 23S rRNA methyltransferase | Rv1399c | <i>lipH</i> probable lipase |
| Rv3715c | <i>recR</i> | RecBC-Independent process of DNA repair | Rv3649 | - | ATP-dependent DNA/RNA helicase | Rv1400c | <i>lipI</i> probable lipase |
| Rv2736c | <i>recX</i> | regulatory protein for RecA | 8. Polysaccharides (cytoplasmic) | | | Rv1900c | <i>lipJ</i> probable esterase |
| Rv2593c | <i>ruvA</i> | Holliday junction binding protein, DNA helicase | Rv1326c | <i>glgB</i> | 1,4- α -glucan branching enzyme | Rv2385 | <i>lipK</i> probable acetyl-hydrolase |
| Rv2592c | <i>ruvB</i> | Holliday junction binding protein | Rv1328 | <i>glgP</i> | probable glycogen phosphorylase | Rv1497 | <i>lipL</i> esterase |
| Rv2594c | <i>ruvC</i> | Holliday junction resolvase, endodeoxyribonuclease | Rv1564c | <i>glgX</i> | probable glycogen debranching enzyme | Rv2284 | <i>lipM</i> probable esterase |
| Rv0054 | <i>ssb</i> | single strand binding protein | Rv1563c | <i>glgY</i> | putative α -amylase | Rv2970c | <i>lipN</i> probable lipase/esterase |
| Rv1210 | <i>tagA</i> | DNA-3-methyladenine glycosidase I | Rv1562c | <i>glgZ</i> | maltooligosyltrehalose trehalohydrolase | Rv1426c | <i>lipO</i> probable esterase |
| Rv3646c | <i>topA</i> | DNA topoisomerase | Rv0126 | - | probable glycosyl hydrolase | Rv2463 | <i>lipP</i> probable esterase |
| Rv2976c | <i>ung</i> | uracil-DNA glycosylase | Rv1781c | - | probable 4- α -glucanotransferase | Rv2485c | <i>lipQ</i> probable carboxylesterase |
| Rv1638 | <i>uvrA</i> | excinuclease ABC subunit A | Rv2471 | - | probable maltase α -glucosidase | Rv3084 | <i>lipR</i> probable acetyl-hydrolase |
| Rv1633 | <i>uvrB</i> | excinuclease ABC subunit B | | | | Rv3176c | <i>lipS</i> probable esterase/lipase |
| Rv1420 | <i>uvrC</i> | excinuclease ABC subunit C | | | | Rv2045c | <i>lipT</i> probable carboxylesterase |
| Rv0949 | <i>uvrD</i> | DNA-dependent ATPase I and helicase II | | | | Rv1076 | <i>lipU</i> probable esterase |
| Rv3198c | <i>uvrD2</i> | putative UvrD | | | | Rv3203 | <i>lipV</i> probable lipase |
| Rv0427c | <i>xthA</i> | exodeoxyribonuclease III | | | | Rv0217c | <i>lipW</i> probable esterase |
| Rv0071 | - | group II intron maturase | | | | Rv2351c | <i>plcA</i> phospholipase C precursor |
| Rv0861c | - | probable DNA helicase | | | | Rv2350c | <i>plcB</i> phospholipase C precursor |
| Rv0944 | - | possible formamidopyrimidine-DNA glycosylase | | | | Rv2349c | <i>plcC</i> phospholipase C precursor |
| Rv1688 | - | probable 3-methylpurine DNA glycosylase | | | | Rv1755c | <i>plcD</i> partial CDS for phospholipase C |
| | | | | | | Rv1104 | - probable esterase pseudogene |
| | | | | | | Rv1105 | - probable esterase pseudogene |
| | | | | | | 6. Aromatic hydrocarbons | |
| | | | | | | Rv3469c | <i>mhpE</i> probable 4-hydroxy-2-oxovalerate aldolase |
| | | | | | | Rv0316 | - probable muconolactone isomerase |
| | | | | | | Rv0771 | - probable 4-carboxymuconolactone decarboxylase |
| | | | | | | Rv0939 | - probable dehydrase |
| | | | | | | Rv1723 | - 6-aminohexanoate-dimer hydro- |

| | | | | | | | | |
|---|--------------|--|---------------------------------------|---------------|---|---|--------------|---|
| Rv2715 | - | lase 2-hydroxymuconic semialdehyde hydrolase | Rv1367c | - | probable penicillin binding protein | Rv1030 | <i>kdpB</i> | potassium-transporting ATPase B chain |
| Rv3530c | - | probable <i>cis</i> -diol dehydrogenase | Rv1730c | - | probable penicillin binding protein | Rv1031 | <i>kdpC</i> | potassium-transporting ATPase C chain |
| Rv3534c | - | 4-hydroxy-2-oxovalerate aldolase | Rv1922 | - | probable penicillin binding protein | Rv3236c | <i>kefB</i> | probable glutathione-regulated potassium-efflux protein |
| Rv3536c | - | aromatic hydrocarbon degradation | Rv2864c | - | probable penicillin binding protein | Rv2877c | <i>merT</i> | possible mercury resistance transport system |
| C. Cell envelope | | | 4. Conserved membrane proteins | | | Rv1811 | <i>mgtC</i> | probable magnesium transport ATPase protein C |
| 1. Lipoproteins (<i>lppA-lppO</i>) 65 | | | Rv0402c | <i>mmpL1</i> | conserved large membrane protein | Rv0362 | <i>mgtE</i> | putative magnesium ion transporter |
| 2. Surface polysaccharides, lipopolysaccharides, proteins and antigens | | | Rv0507 | <i>mmpL2</i> | conserved large membrane protein | Rv2856 | <i>nicT</i> | probable nickel transport protein |
| Rv0806c | <i>cpsY</i> | probable UDP-glucose-4-epimerase | Rv0206c | <i>mmpL3</i> | conserved large membrane protein | Rv0924c | <i>nramp</i> | transmembrane protein belonging to Nramp family |
| Rv3811 | <i>csp</i> | secreted protein | Rv0450c | <i>mmpL4</i> | conserved large membrane protein | Rv2691 | <i>trkA</i> | probable potassium uptake protein |
| Rv1677 | <i>dsbF</i> | highly similar to C-term Mpt53 | Rv0676c | <i>mmpL5</i> | conserved large membrane protein | Rv2692 | <i>trkB</i> | probable potassium uptake protein |
| Rv3794 | <i>embA</i> | involved in arabinogalactan synthesis | Rv1557 | <i>mmpL6</i> | conserved large membrane protein | Rv2287 | <i>yjcE</i> | probable Na ⁺ /H ⁺ exchanger |
| Rv3795 | <i>embB</i> | involved in arabinogalactan synthesis | Rv2942 | <i>mmpL7</i> | conserved large membrane protein | Rv2723 | - | probable membrane protein, tellurium resistance |
| Rv3793 | <i>embC</i> | involved in arabinogalactan synthesis | Rv3823c | <i>mmpL8</i> | conserved large membrane protein | Rv3162c | - | probable membrane protein |
| Rv3875 | <i>esat6</i> | early secretory antigen target | Rv2339 | <i>mmpL9</i> | conserved large membrane protein | Rv3237c | - | possible potassium channel protein |
| Rv0112 | <i>gca</i> | probable GDP-mannose dehydratase | Rv1183 | <i>mmpL10</i> | conserved large membrane protein | Rv3743c | - | probable cation-transporting ATPase |
| Rv0113 | <i>gmhA</i> | phosphoheptose isomerase | Rv0202c | <i>mmpL11</i> | conserved large membrane protein | 3. Carbohydrates, organic acids and alcohols | | |
| Rv2965c | <i>kdtB</i> | lipopolysaccharide core biosynthesis protein | Rv1522c | <i>mmpL12</i> | conserved large membrane protein | Rv2443 | <i>dctA</i> | C4-dicarboxylate transport protein |
| Rv2878c | <i>mpt53</i> | secreted protein Mpt53 | Rv0403c | <i>mmpS1</i> | conserved small membrane protein | Rv3476c | <i>kgtP</i> | sugar transport protein |
| Rv1980c | <i>mpt64</i> | secreted immunogenic protein Mpb64/Mpt64 | Rv0506 | <i>mmpS2</i> | conserved small membrane protein | Rv1902c | <i>nanT</i> | probable salic acid transporter |
| Rv2875 | <i>mpt70</i> | major secreted immunogenic protein Mpt70 precursor | Rv2198c | <i>mmpS3</i> | conserved small membrane protein | Rv1236 | <i>sugA</i> | membrane protein probably involved in sugar transport |
| Rv2873 | <i>mpt83</i> | surface lipoprotein Mpt83 | Rv0451c | <i>mmpS4</i> | conserved small membrane protein | Rv1237 | <i>sugB</i> | sugar transport protein |
| Rv0899 | <i>ompA</i> | member of OmpA family | Rv0677c | <i>mmpS5</i> | conserved small membrane protein | Rv1238 | <i>sugC</i> | ABC transporter component of sugar uptake system |
| Rv3810 | <i>pirG</i> | cell surface protein precursor (Erp protein) | 5. Other membrane proteins 211 | | | Rv3331 | <i>sugI</i> | probable sugar transport protein |
| Rv3782 | <i>rfeE</i> | similar to rhamnosyl transferase | III. Cell processes | | | Rv2835c | <i>ugpA</i> | sn-glycerol-3-phosphate permease |
| Rv1302 | <i>rfe</i> | undecaprenyl-phosphate α -N-acetylglucosaminyltransferase | A. Transport/binding proteins | | | Rv2833c | <i>ugpB</i> | sn-glycerol-3-phosphate-binding periplasmic lipoprotein |
| Rv2145c | <i>wag31</i> | antigen 84 (aka wag31) | 1. Amino acids | | | Rv2832c | <i>ugpC</i> | sn-glycerol-3-phosphate transport ATP-binding protein |
| Rv0431 | - | tuberculin related peptide (AT103) | 1. Amino acids | | | Rv2834c | <i>ugpE</i> | sn-glycerol-3-phosphate transport system protein |
| Rv0954 | - | cell envelope antigen | Rv2127 | <i>ansP</i> | L-asparagine permease | Rv2316 | <i>uspA</i> | sugar transport protein |
| Rv1514c | - | involved in polysaccharide synthesis | Rv0346c | <i>aroP2</i> | probable aromatic amino acid permease | Rv2318 | <i>uspC</i> | sugar transport protein |
| Rv1518 | - | involved in exopolysaccharide synthesis | Rv0917 | <i>betP</i> | glycine betaine transport | Rv2317 | <i>uspE</i> | sugar transport protein |
| Rv1758 | - | partial cutinase | Rv1704c | <i>cycA</i> | transport of D-alanine, D-serine and glycine | Rv1200 | - | probable sugar transporter |
| Rv1910c | - | probable secreted protein | Rv3666c | <i>dppA</i> | probable peptide transport system permease | Rv2038c | - | probable ABC sugar transporter |
| Rv1919c | - | weak similarity to pollen antigens | Rv3665c | <i>dppB</i> | probable peptide transport system permease | Rv2039c | - | probable sugar transporter |
| Rv1984c | - | probable secreted protein | Rv3664c | <i>dppC</i> | probable peptide transport system permease | Rv2040c | - | probable sugar transporter |
| Rv1987 | - | probable secreted protein | Rv3663c | <i>dppD</i> | probable ABC-transporter | Rv2041c | - | probable sugar transporter |
| Rv2223c | - | probable exported protease | Rv0522 | <i>gabP</i> | probable 4-amino butyrate transporter | 4. Anions | | |
| Rv2224c | - | probable exported protease | Rv0411c | <i>glnH</i> | putative glutamine binding protein | Rv2684 | <i>arsA</i> | probable arsenical pump |
| Rv2301 | - | probable cutinase | Rv2564 | <i>glnQ</i> | probable ATP-binding transport protein | Rv2685 | <i>arsB</i> | probable arsenical pump |
| Rv2345 | - | precursor of probable membrane protein | Rv1280c | <i>oppA</i> | probable oligopeptide transport protein | Rv3578 | <i>arsB2</i> | probable arsenical pump |
| Rv2672 | - | putative exported protease | Rv1283c | <i>oppB</i> | oligopeptide transport protein | Rv2643 | <i>arsC</i> | probable arsenical pump |
| Rv3019c | - | similar to Esat6 | Rv1282c | <i>oppC</i> | oligopeptide transport system permease | Rv2397c | <i>cysA</i> | sulphate transport ATP-binding protein |
| Rv3036c | - | probable secreted protein | Rv1281c | <i>oppD</i> | probable peptide transport protein | Rv2399c | <i>cysT</i> | sulphate transport system permease protein |
| Rv3449 | - | probable precursor of serine protease | Rv2320c | <i>rocE</i> | probable cationic amino acid transport | Rv2398c | <i>cysW</i> | sulphate transport system permease protein |
| Rv3451 | - | probable cutinase | Rv3253c | <i>rocE</i> | probable cationic amino acid transport | Rv1857 | <i>modA</i> | molybdate binding protein |
| Rv3452 | - | probable cutinase precursor | Rv3454 | - | possible proline permease | Rv1858 | <i>modB</i> | transport system permease, molybdate uptake |
| Rv3724 | - | probable cutinase precursor | 2. Cations | | | Rv1859 | <i>modC</i> | molybdate uptake ABC-transporter |
| 3. Murein sacculus and peptidoglycan | | | Rv2920c | <i>amt</i> | putative ammonium transporter | Rv1860 | <i>modD</i> | precursor of Apa (45/47 kD secreted protein) |
| Rv2911 | <i>dacB</i> | penicillin binding protein | Rv1607 | <i>chaA</i> | putative calcium/proton antiporter | Rv2329c | <i>narK1</i> | probable nitrite extrusion protein |
| Rv2981c | <i>ddlA</i> | D-alanine-D-alanine ligase A | Rv1239c | <i>corA</i> | probable magnesium and cobalt transport protein | Rv1737c | <i>narK2</i> | nitrite extrusion protein |
| Rv3809c | <i>glf</i> | UDP-galactopyranose mutase | Rv0092 | <i>ctpA</i> | cation-transporting ATPase | Rv0261c | <i>narK3</i> | nitrite extrusion protein1 |
| Rv1018c | <i>glmU</i> | UDP-N-acetylglucosamine pyrophosphorylase | Rv0103c | <i>ctpB</i> | cation transport ATPase | Rv0267 | <i>narU</i> | similar to nitrite extrusion protein 2 |
| Rv3382c | <i>lytB</i> | LytB protein homologue | Rv3270 | <i>ctpC</i> | cation transport ATPase | Rv0934 | <i>phoS1</i> | PstS component of phosphate uptake |
| Rv1110 | <i>lytB'</i> | very similar to LytB | Rv1469 | <i>ctpD</i> | probable cadmium-transporting ATPase | Rv0928 | <i>phoS2</i> | PstS component of phosphate uptake |
| Rv1315 | <i>murA</i> | UDP-N-acetylglucosamine-1-carboxyvinyltransferase | Rv0908 | <i>ctpE</i> | probable cation transport ATPase | Rv0820 | <i>phoT</i> | phosphate transport system ABC transporter |
| Rv0482 | <i>murB</i> | UDP-N-acetylenolpyruvoylglucosamine reductase | Rv1997 | <i>ctpF</i> | probable cation transport ATPase | Rv3301c | <i>phoY1</i> | phosphate transport system regulator |
| Rv2152c | <i>murC</i> | UDP-N-acetyl-muramate-alanine ligase | Rv1992c | <i>ctpG</i> | probable cation transport ATPase | Rv0821c | <i>phoY2</i> | phosphate transport system regulator |
| Rv2155c | <i>murD</i> | UDP-N-acetylmuramoylalanine-D-glutamate ligase | Rv0425c | <i>ctpH</i> | C-terminal region putative cation-transporting ATPase | Rv0545c | <i>pitA</i> | low-affinity inorganic phosphate transporter |
| Rv2158c | <i>murE</i> | meso-diaminopimelate-adding enzyme | Rv0107c | <i>ctpl</i> | probable magnesium transport ATPase | Rv2281 | <i>pitB</i> | phosphate permease |
| Rv2157c | <i>murF</i> | D-alanine:D-alanine-adding enzyme | Rv0969 | <i>ctpV</i> | cation transport ATPase | Rv0930 | <i>pstA1</i> | PstA component of phosphate uptake |
| Rv2153c | <i>murG</i> | transferase in peptidoglycan synthesis | Rv3044 | <i>fecB</i> | putative FcII-dicitrate transporter | Rv0936 | <i>pstA2</i> | PstA component of phosphate uptake |
| Rv1338 | <i>murI</i> | glutamate racemase | Rv0265c | <i>fecB2</i> | iron transport protein FeIII dicitrate transporter | Rv0933 | <i>pstB</i> | ABC transport component of phosphate uptake |
| Rv2156c | <i>murX</i> | phospho-N-acetylmuramoyl-pentapeptide transferase | Rv1029 | <i>kdpA</i> | potassium-transporting ATPase A chain | Rv0935 | <i>pstC</i> | PstC component of phosphate uptake |
| Rv3332 | <i>nagA</i> | N-acetylglucosamine-6-P-deacetylase | | | | Rv0929 | <i>pstC2</i> | membrane-bound component of |
| Rv0016c | <i>pbpA</i> | penicillin-binding protein | | | | | | |
| Rv2163c | <i>pbpB</i> | penicillin-binding protein 2 | | | | | | |
| Rv0050 | <i>ponA</i> | penicillin-binding protein class A penicillin binding protein | | | | | | |
| Rv3682 | <i>ponA'</i> | FtsW/RodA/SpovE family | | | | | | |
| Rv0017c | <i>rodA</i> | probable penicillin binding protein | | | | | | |
| Rv0907 | - | | | | | | | |

Rv0932c *pstS* phosphate transport system
PstS component of phosphate uptake

Rv2400c *subI* sulphate binding precursor

Rv0143c - probable chloride channel

Rv17107 - probable sulphate permease

Rv1739c - possible sulphate transporter

Rv3679 - possible anion transporter

Rv3680 - probable anion transporter

5. Fatty acid transport

Rv2790c *lip1* non-specific lipid transport protein

Rv3540c *lip2* non-specific lipid transport protein

6. Efflux proteins

Rv2936 *draA* similar daunorubicin resistance ABC-transporter

Rv2937 *draB* similar daunorubicin resistance transmembrane protein

Rv2938 *draC* 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 *shhB* 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 diacylglycerol transferase

Rv1539 *lspA* lipoprotein signal peptidase

Rv0379 *sec* probable transport protein SecE/Sec61- γ family

Rv3240c *secA* SecA, preprotein translocase sub-

Rv1821 *secA2* unit 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-phosphohydrolyase

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 *smcB* 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

IS6110 16 copies

IS1081 6 copies

Others 37 copies

2. REP13E12 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'-phosphotransferase

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

link to Nature.com