1   Genome-wide assessment of *Mycobacterium tuberculosis* conditionally essential

2   metabolic pathways

3

4   Yusuke Minato[a], Daryl M Gohl [b], Joshua M. Thiede[a], Jeremy M. Chacón[c], William R.

5   Harcombe[c], Fumito Maruyama[d,e,f], and Anthony D. Baughn [a#]

6

7   [a]Department of Microbiology and Immunology, University of Minnesota Medical School,

8   Minneapolis, Minnesota, USA.

9   [b]University of Minnesota Genomics Center, Minneapolis, Minnesota, USA.

10   [c]Biotechnology Institute and Department of Ecology, Evolution and Behavior, University

11   of Minnesota, Minnesota, USA.

12   [d]Department of Microbiology, Graduate School of Medicine, Kyoto University, Kyoto,

13   Japan.

14   [e]The Japan Science and Technology Agency/Japan International Cooperation Agency,

15   Science and Technology Research Partnership for Sustainable Development (JST/JICA,

16   SATREPS), Tokyo, Japan.

17   [f]Scientific and Technological Bioresource Nucleus, Universidad de La Frontera, Temuco,

18   Chile

19

20

21   #Address correspondence to Anthony D. Baughn, abaughn@umn.edu

22

23

24

Abstract

Better understanding of essential cellular functions in pathogenic bacteria is important for the development of more effective antimicrobial agents. We performed a comprehensive identification of essential genes in *Mycobacterium tuberculosis*, the major causative agent of tuberculosis, using a combination of transposon insertion sequencing (Tn-seq) and comparative genomic analysis. To identify conditionally essential genes by Tn-seq, we used media with different nutrient composition. Although many conditional gene essentialities were affected by the presence of relevant nutrient sources, we also found that the essentiality of genes in a subset of metabolic pathways was unaffected by metabolite availability. Comparative genomic analysis revealed that not all essential genes identified by Tn-seq were fully conserved within the *M. tuberculosis* complex including some existing anti-tubercular drug target genes. In addition, we utilized an available *M. tuberculosis* genome-scale metabolic model, iSM810, to predict *M. tuberculosis* gene essentiality *in silico*. Comparing the sets of essential genes experimentally identified by Tn-seq to those predicted *in silico* reveals the capabilities and limitations of gene essentiality predictions highlighting the complexity of *M. tuberculosis* essential metabolic functions. This study provides a promising platform to study essential cellular functions in *M. tuberculosis*.

43

Author Summary

Mycobacterium tuberculosis causes 10 million cases of tuberculosis (TB) resulting in over one million deaths each year. TB therapy is challenging because it requires a

2

47   minimum of six months of treatment with multiple drugs. Protracted treatment times and

48   the emergent spread of drug resistant *M. tuberculosis* necessitate the identification of

49   novel targets for drug discovery to curb this global health threat. Essential functions,

50   defined as those indispensable for growth and/or survival, are potential targets for new

51   antimicrobial drugs. In this study, we aimed to define gene essentialities of *M.*

52   *tuberculosis* on a genome-wide scale to comprehensively identify potential targets for

53   drug discovery. We utilized a combination of experimental (functional genomics) and *in*

54   *silico* approaches (comparative genomics and flux balance analysis). Our functional

55   genomics approach identified sets of genes whose essentiality was affected by nutrient

56   availability. Comparative genomics revealed that not all essential genes were fully

57   conserved within the *M. tuberculosis* complex. Comparing sets of essential genes

58   identified by functional genomics to those predicted by flux balance analysis highlighted

59   gaps in current knowledge regarding *M. tuberculosis* metabolic capabilities. Thus, our

60   study identifies numerous potential anti-tubercular drug targets and provides a

61   comprehensive picture of the complexity of *M. tuberculosis* essential cellular functions.

62

63   **Introduction**

64   *Mycobacterium tuberculosis* is responsible for approximately 10.4 million new

65   cases of active tuberculosis (TB) infection and 1.4 million deaths annually (1). While TB

66   chemotherapy has a high success rate in curing drug susceptible TB infections, it is

67   challenging, in part because it requires a minimum of 6 months of treatment with drugs

68   associated with adverse reactions. Thus, finding targets for new TB drugs that are more

69   potent than existing drugs is needed (2).

70   Essential genes, defined as genes indispensable for growth and/or survival, are

71   potential targets for new types of antimicrobial drugs. Gene essentiality can be

72   assessed by targeted gene-disruptions, where genes that cannot be disrupted are

73   typically categorized as being essential. However, such traditional genetic approaches

74   are labor-intensive and not easily adaptable to genome-scale screening. Recent

75   advances in Next-Generation Sequencing (NGS)-based approaches have transformed

76   our ability to examine gene functions in a genome-wide manner. Transposon insertion

77   sequencing (Tn-seq) has been widely used to conduct fitness profiling of gene functions

78   in many bacterial species, including *M. tuberculosis* (3-10). In addition to fitness profiling,

79   a lack of representation of specific transposon insertions within a saturated transposon

80   library has been used to identify essential genes in genome-wide screens (3, 9, 10).

81   In *M. tuberculosis*, several systematic genome-wide studies, including studies

82   using Tn-seq, have been performed to identify essential genes *in vitro* (7-16) and *in vivo*

83   (4). The gene essentialities determined by Tn-seq studies are accessible through

84   publicly available databases, such as Tuberculist (17), Biocyc (18), and OGEE (19).

85   Most of these gene essentiality data were obtained from Tn-seq studies that were

86   carried out using a defined growth medium that was supplemented with a limited

87   number of nutrients (7-10, 13, 15). Under this growth condition, genes for numerous

88   essential central metabolic pathways can be rendered dispensable through

89   supplementation. Thus, these previous studies have likely miscategorized a large set of

90   genes as essential rather than conditionally essential. For example, pantothenate, an

91   essential precursor in coenzyme A biosynthesis, is not supplemented in most of the

92   commonly used media for *M. tuberculosis*. As a result, *panC*, encoding pantothenate

4

93   synthetase, the enzyme catalyzing the last step in pantothenate biosynthesis, was

94   defined as an essential gene in these three databases (17-19) despite a previous study

95   showing that the *panCD* double deletion mutant of *M. tuberculosis* H37Rv strain can

96   grow in the presence of supplemental pantothenate (20). Thus, more precise definition

97   of *M. tuberculosis* essential genes on a genome-wide scale, and annotation of genes

98   which are conditionally essential, would improve the usefulness of these databases.

99   In this study, we developed a defined-nutrient rich medium (Mtb YM rich) that

100 included a variety of nutrient sources for *M. tuberculosis* to ease identification of

101 conditionally essential genes. We used Tn-seq to identify essential genes in Mtb YM

102 rich and minimal medium (Mtb minimal). As expected, essentialities of many genes

103 involved in metabolite biosynthesis pathways were affected by the supplemental

104 nutrients in Mtb YM rich medium. However, we found that essentialities of certain

105 metabolic pathways were unaffected by the absence or presence of relevant nutrient

106 sources. We also found that both growth conditions had some essential genes that were

107 unique to each environment. In addition, we compared essential genes identified by Tn-

108 seq with highly conserved genes that were identified by a comparative genomics

109 analysis and a modified *in silico* metabolic model. These comparisons indicated that

110 essential genes were highly enriched amongst the conserved core genome and that

111 such gene essentiality measurements can be used to refine metabolic models.

112

113 **Results and Discussions**

114   **Identification of *M. tuberculosis* essential genes in a chemically defined**

115 **nutrient-rich media.** We developed a defined nutrient-rich medium for *M. tuberculosis*

116   (Mtb YM rich medium) (Fig. 1A, Table S1) by supplementing numerous nutrients into a

117   minimal *M. tuberculosis* growth medium (Mtb minimal medium) (21). We chose

118   supplements based on their known use by various bacterial species including *M.*

119   *tuberculosis* (7, 12, 20, 22-24). The supplements included several carbon sources,

120   nitrogen sources, co-factors, amino acids, nucleotide bases and other nutrients. Four of

121   the supplements, lipoic acid, nicotinamide, hemin, and ribose, inhibited *M. tuberculosis*

122   H37Rv growth at higher concentrations. Thus, we utilized concentrations of these

123   supplements that did not impair growth. The Mtb YM rich medium supported the *M.*

124   *tuberculosis* H37Rv growth similar to the commonly used 7H9 medium that contains

125   essential salts and relatively few nutrients (Fig. 1B).

126   We next generated a library of *M. tuberculosis* H37Rv transposon insertion

127   mutants on Mtb YM rich medium plates. As the *M. tuberculosis* H37Rv genome contains

128   74,602 TA dinucleotides, we collected at least 150,000 colonies to approach saturation

129   of *himar1* transposon insertion sites in the library. Genomic DNA was isolated from the

130   library. Next, the DNA was sheared, end repaired, sequencing adapters were added,

131   and the transposon adjacent regions were enriched by PCR before massively parallel

132   sequencing. The resultant sequencing data was analyzed using TRANSIT, a recently

133   developed software package for analyzing Tn-seq data (14). TRANSIT contains two

134   statistical methods, Bayesian/Gumbel Method and Hidden Markov Model (HMM), to

135   identify essential genes or essential genomic regions, respectively, in a single growth

136   condition. By using the Bayesian/Gumbel Method, we found that 542 genes were

137   essential for *M. tuberculosis* growth in the Mtb YM rich medium (Fig. 1C, Dataset S1).

138   The Bayesian/Gumbel method assesses gene essentiality based on consecutive

139 sequence of TA sites lacking insertion within a gene. When the analysis did not exceed

140 the significance thresholds, genes were called either short or uncertain by TRANSIT. As

141 a result, 231 genes were called short and 115 genes were called uncertain by this

142 analysis (Fig. 1C). To overcome this issue, we also identified essential genomic regions

143 using an HMM (14). The HMM is based on the read count at a given site and the

144 distribution over the surrounding sites. This analysis identified 13.3 % of TA sites as

145 essential and 2.8 % of TA sites as growth defective (Fig. 1D, Dataset S1). These

146 essential and growth defective regions included 17 short genes (14 essential and 3

147 growth defective genes) and 21 uncertain genes (10 essential and 11 growth defective

148 genes) whose essentiality could not be assessed by the Bayesian/Gumbel Method. In

149 addition, HMM identified 5 short genes and 16 uncertain genes that contained both

150 essential and non-essential TA sites. These genes are listed in Table S2. In total, we

151 identified 601 genes (542 genes by the Bayesian/Gumbel Method and additional 59

152 genes by the HMM) as essential genes for *M. tuberculosis* survival in the Mtb YM rich

153 medium. A list of these genes is available through the BioCyc smart table format

154 (https://biocyc.org/group?id=biocyc14-7907-3764257976) (25). These essential genes

155 included known targets for existing anti-tubercular drugs further validating that the Tn-

156 seq assay successfully identified *M. tuberculosis* essential genes (Table 1).

157 The list of essential genes identified in this study was also compared with the

158 essential genes identified by the past Tn-seq studies, Griffin et al (8) and DeJesus et al

159 (9) (Fig. 1E and Table S3). In general, our result was largely consistent with the past

160 studies and a total of 458 genes were identified as essential in all three studies (Fig. 1E)

161 (https://biocyc.org/group?id=biocyc14-7907-3764424447). Since Mtb YM rich medium

7

162   used in this study was supplemented with numerous nutrients, many genes that were

163   related to biosynthetic pathways of the supplemented nutrients (e.g. amino acids,

164   pantothenate, purine, flavin and others) were not essential in our study while the past

165   studies categorized these genes as essential. We also identified genes that were only

166   essential in our study but not in other studies. We identified some of these genes as

167   conditionally essential in Mtb YM rich media (described below, Fig. 2, and Table S4). Of

168   note, our study could not detect several genes that were highly expected to be essential

169   and identified as essential in DeJesus study. These genes included short genes such as

170   several ribosomal genes and *folK*. This likely was because DeJesus study used a more

171   saturated transposon library (fourteen replicates vs two replicates).

172   **Comparison of *M. tuberculosis* essential genes between Mtb YM rich**

173   **medium and minimal nutrient medium.** Because gene essentiality can be affected by

174   the external environment (26), in particular by nutrient availability, we next compared

175   gene essentiality between Mtb YM rich medium and Mtb minimal medium. Similar to the

176   transposon library generated on Mtb YM rich medium plates, we generated *M.*

177   *tuberculosis* H37Rv transposon libraries on Mtb minimal medium plates and collected at

178   least 150,000 colonies to create a saturated library. The resultant sequencing data was

179   compared with the Mtb YM rich plate data by using a permutation test-based method to

180   identify genes with statistically significant differences in transposon insertion count (Fig.

181   2A, Table S4). We found that 130 genes were conditionally essential in the Mtb minimal

182   medium compared to the Mtb YM rich medium (https://biocyc.org/group?id=biocyc13-

183   7907-3706551227).

8

184   As anticipated, many of the conditionally essential genes that were identified

185   corresponded to the differences in nutrient composition between Mtb minimal and Mtb

186   YM rich. For example, Mtb YM rich medium contained L-aspartate and pantothenate.

187   Thus, the genes in L-aspartate and pantothenate biosynthesis pathways were

188   dispensable in Mtb YM rich medium (Fig. 2B). However, Mtb YM rich medium did not

189   contain metabolites located downstream of pantothenate. Consequently, these

190   downstream genes (e.g. *coaA*, encoding pantothenate kinase) were essential in both

191   Mtb YM rich and Mtb minimal media (Fig. 2B). This observation is consistent with a

192   previous report that the *M. tuberculosis panCD* double deletion mutant strain can grow

193   in the presence of pantothenate (20), and also suggested that single gene disruptions in

194   the pantothenate biosynthesis pathway (*panB*, *panC*, and *panD*) show a similar

195   pantothenate auxotrophic phenotype as the *panCD* double deletion mutant strain.

196   Another example was the asparagine transporter gene *ansP2* (27) that was

197   conditionally essential in Mtb minimal medium (Fig. 2B) presumably because Mtb

198   minimal medium contained arginine as the sole nitrogen source whereas Mtb YM rich

199   medium contained multiple nitrogen sources (Fig. 1A).

200   Unexpectedly, we also identified 98 genes that were conditionally essential in

201   Mtb YM rich media compared to the Mtb minimal medium

202   (https://biocyc.org/group?id=biocyc13-7907-3710604059). Such genes included *ponA1*

203   and *ponA2,* encoding penicillin binding proteins (PBPs) involved in cell wall

204   peptidoglycan (PG) biogenesis. It was previously shown that *ponA1* and *ponA2* are only

205   essential *in vivo* but not during growth in culture medium (28, 29). Thus, one of nutrients

206   that is uniquely present in Mtb YM rich medium might be also present *in vivo* and may

207   be responsible for the *in vivo* fitness defect of the *ponA1* mutant. LdtB is one of the

208   major L,D-transpeptidases (Ldts) that is also involved in PG biogenesis. *ldtB* was also

209   conditionally essential in Mtb YM rich medium. Interestingly, several genes that were

210   conditionally essential in the absence of *ponA1, ponA2,* or *ldtB* (e.g. *Rv1086, Rv1248c,*

211   *Rv3490, treS, otsA*) were also conditionally essential in Mtb YM rich medium suggesting

212   Mtb YM rich medium negatively affects PG biogenesis in *M. tuberculosis* (30). We also

213   found the *metH*, encoding one of the two methionine synthases, was only essential in

214   Mtb YM rich medium whereas *metE*, encoding the other methionine synthase, was only

215   essential in Mtb minimal medium (Fig.2C). The observed conditional essentialities were

216   consistent with a previous study that *M. tuberculosis metE* expression was inhibited in

217   the presence of vitamin B_{12} by a *metE* B_{12} riboswitch, and the vitamin B_{12}-dependent

218   MetH is predominantly used when vitamin B_{12} is available (31). These findings

219   confirmed that the conditionally essential genes identified by Tn-seq in this study are

220   consistent with previous findings (25).

221   The supplementation of nutrients in the Mtb YM rich medium could potentially

222   subvert the need for enzymes in at least 35 metabolic pathways (Fig. 1A, Fig. 2D, and

223   Table S5). We found fewer gene essentialities in 22 of these pathways in Mtb YM rich

224   medium, suggesting that *M. tuberculosis* can functionally utilize these nutrients. Notably,

225   many genes in these pathways were identified as essential genes by the previous

226   studies (8, 13) and listed as essential genes in public database such as Tuberculist (17)

227   and BioCyc (18). Thus, our results revealed that these genes are essential only in the

228   absence of the corresponding nutrients. In contrast, we also found that essentiality of

229   genes in 13 other metabolic pathways were not altered by nutrient supplementation (Fig.

230   2D). Previous studies have demonstrated that some auxotrophic mutants, such as

231   those for L-arginine, L-lysine, and inositol, require supplementation with the

232   corresponding nutrient at a relatively high concentration to support their growth (32-34).

233   Certain auxotrophic mutants are also known to show growth or survival defects even in

234   the presence of the corresponding nutrient (35, 36). These results were consistent with

235   previous findings that essentiality of some central metabolic pathways could not be

236   bypassed in pathogenic mycobacteria (12) and also provide a more comprehensive

237   understanding of the nutrient utilization capacity of *M. tuberculosis*.

238   **Identification of highly conserved genes in the *M. tuberculosis* complex.**

239   Essential genes that were identified by Tn-seq were further interrogated through

240   comparative genomic analysis in order to see whether essentiality correlated with high

241   levels of sequence conservation within the *M. tuberculosis* complex. A total of 226

242   complete genome sequences of *M. tuberculosis* complex were available from the

243   PATRIC database (37). We excluded non-pathogenic strains (e.g. *M. tuberculosis*

244   H37Ra and *Mycobacterium bovis* BCG) and used 199 genome sequences for

245   comparative genomic analysis (Fig. 3A, Table S6). We identified total of 17,813 genes

246   from 199 strains of the *M. tuberculosis* complex (Fig. 3B and C). Among these genes, 2,

247   206 genes (1,030 core genes and 1,176 soft core genes) were highly conserved in most

248   of the strains ($\geq$ 95 % of strains).

249   Our Tn-seq analysis identified 601 genes as essential in Mtb YM rich medium

250   (Dataset S1, Table S2 and smart table). Among them, we confirmed at least 60 % of the

251   essential genes (356 genes) were included in the list of core/softcore genes (Fig. 3D)

252   (https://biocyc.org/group?id=biocyc14-7907-3764449513). Of note, not all of the target

253   genes for existing anti-tubercular drugs were categorized as core/softcore genes (Table

254   1).

255   **Prediction of *M. tuberculosis* essential genes *in silico.*** Genome-scale

256   metabolic models have been used to computationally simulate a range of cellular

257   functions (38). We utilized a genome-scale model of *M. tuberculosis*, iSM810, to predict

258   gene essentiality *in silico* using flux balance analysis (FBA) (39). The principles for FBA

259   have been described previously (40). Briefly, a metabolic network is represented by a

260   stoichiometric matrix that contains reactions and metabolites. A biomass reaction is

261   defined based on experimentally determined amounts of specific metabolites required

262   for cellular growth. The stoichiometric matrix can be converted into a system of linear

263   equations. FBA then solves this system by optimization, typically of the biomass

264   reaction. Constraints on reaction fluxes can be used to simulate metabolite availability

265   or enzyme presence. For example, a reaction's uptake flux is constrained to zero when

266   the simulated environment does not contain the associated metabolite. Reaction fluxes

267   that simulate metabolite uptake within iSM810 were used to define *in silico* growth

268   media. Uptake was not bounded for metabolites present in the selected growth medium,

269   however; not all media components were represented within the iSM810 transport

270   reactions (Table S7). Genes were assessed for essentiality by systematically closing

271   the flux on each reaction within iSM810 and assessing simulated biomass production on

272   each medium using FBA. Genes were defined essential if biomass production after

273   knockout was <1e-10.

274   iSM810 includes 810 metabolic genes (including 1 orphan gene) and 938

275   metabolic reactions (39). Among the 810 genes in iSM810, our FBA analysis predicted

276   that 159 genes were essential in Mtb YM rich medium and 221 genes were essential in

277   Mtb minimal medium (Fig. 4A, Table S7). We then compared gene essentiality

278   prediction by FBA with essential genes identified by Tn-seq (Fig. 4B). We found that the

279   sensitivity of the FBA based gene essentiality prediction was low as there were a

280   number of genes that were predicted to be essential by FBA but not identified as

281   essential genes by Tn-seq (Fig. 4B) (https://biocyc.org/group?id=biocyc14-7907-

282   3764439908). For instance, we found that genes related to riboflavin biosynthesis were

283   non-essential in Mtb YM rich medium by Tn-seq analysis, but were essential *in silico*.

284   We examined why iSM810 could not accurately predict the essentiality of genes in the

285   riboflavin biosynthesis pathway and found that the model lacked a transport reaction for

286   riboflavin (Table S7). Similarly, we found that the model lacked transport reactions for

287   vitamin B_{12}, *para*-aminobenzoic acid (PABA), H_2O, and myo-inositol. To investigate

288   whether the addition of these transport reactions could improve gene essentiality

289   prediction by iSM810, we added these transport reactions to the model. These changes

290   fixed multiple mis-matches between iSM810 and Tn-seq, by causing the model to

291   determine that genes related to thiamine biosynthesis and riboflavin biosynthesis were

292   non-essential in Mtb YM rich medium (Fig. 4A and B, Table S7). Allowing PABA uptake

293   caused no changes. After adding transport reactions, the growth rate prediction in Mtb

294   YM rich medium increased from 0.055 g/L/day to 0.0876 g/L/day.

295   Unlike the Tn-seq results, the FBA predicted that all genes essential in Mtb YM

296   rich medium were also essential in Mtb minimal and failed to predict any genes such as

297   *metH* that were only essential in Mtb YM rich (Table S7). This is expected given the lack

298   of gene regulation implicit in our FBA modeling.

13

299   Comparing the sets of essential genes experimentally identified by Tn-seq to

300   those computationally predicted *in silico* highlighted the limitations of current *M.*

301   *tuberculosis* metabolic models. For example, we noticed that genes located

302   downstream of purine biosynthesis pathway were identified as essential by Tn-seq in

303   both Mtb YM rich and Mtb minimal media (Fig. 4C). However, there were mismatches

304   between iSM810 and Tn-seq. Among these genes, *purC* is the only gene that was

305   characterized by targeted gene deletion analysis in *M. tuberculosis* (41). The *purC* gene

306   was difficult to delete (41) and the *purC* deletion mutant strain of *M. tuberculosis*

307   showed a notable growth defect compared to the parent strain even in the presence of

308   hypoxanthine (22). This observation could explain why the *purC* gene was identified as

309   essential by Tn-seq in Mtb YM rich. However, the other genes in the purine biosynthesis

310   pathway have not been characterized by targeted gene deletion studies. Thus, future

311   studies are necessary to investigate essentiality of *purK*, *purB*, and *purH* by targeted

312   gene deletion.

313   We also identified a potential avenue for improvement of the genome-scale

314   model of *M. tuberculosis* by comparing Tn-seq identified essential genes and to those

315   predicted *in silico.* We found that many genes in known essential metabolic pathways

316   were not predicted to be essential (e.g. glycolysis, folate metabolism, mycolate

317   biosynthesis, ATP biosynthesis, and several amino acid biosynthesis pathways)

318   (https://biocyc.org/group?id=biocyc14-7907-3764451955). Examination of the biomass

319   function in iSM810 revealed that several known essential metabolites are not connected

320   to the biomass function in iSM810. For example, folate is not connected to the biomass

321   function and this explains why addition of a PABA transport reaction did not affect any

14

322   gene essentiality predictions. None of ATP synthase genes in iSM810 (*atpA*, *atpB*, *atpD*,

323   *atpE*, *atpG*, and *atpH*) was predicted as essential because these were linked to the

324   same reaction using "or" Boolean logic, meaning the presence of one gene was

325   sufficient to allow the reaction to proceed. Thus, further improvement of iSM810 for

326   more accurate prediction of essential genes could be achieved by connecting such

327   essential metabolites to the biomass function.

328   **Conclusions.** In this study, we utilized functional genomics and comparative

329   genomics approaches to identify *M. tuberculosis* essential genes. We identified distinct

330   sets of essential and conditionally essential genes from different approaches and

331   different growth conditions. In addition, comparison of essential genes identified by

332   functional genomics approach to *in silico* predicted essential genes highlighted current

333   gaps in knowledge regarding *M. tuberculosis* metabolism. Our study provides a

334   promising platform to shed new light on essential cellular functions in *M. tuberculosis*

335   that can lead to the discovery of novel targets for antitubercular drugs.

336

337   **Materials and methods**

338   **Media and growth conditions.** *M. tuberculosis* H37Rv was grown aerobically at

339   37 °C in Middlebrook 7H9 medium supplemented with oleate-albumin-dextrose-catalase

340   (OADC; 10%, vol/vol), glycerol (0.2%, vol/vol), and tyloxapol (0.05%, vol/vol) unless

341   otherwise noted. Mtb minimal medium (0.2%, vol/vol glycerol as a sole carbon source)

342   (21) and Mtb YM rich medium (Table S1) agar plates were used to generate *M.*

343   *tuberculosis* H37Rv transposon libraries. Tyloxapol (0.05%, vol/vol) were added to the

344   both Mtb minimal and Mtb YM rich agar plates.

15

345 **Construction of saturated transposon libraries of *M. tuberculosis*.**

346 Transposon mutagenesis were performed as previously described (42).

347 Mycobacteriophage phAE180 (42) was used to transduce a mariner derivative

348 transposon Tn5371 (43) to *M. tuberculosis* H37Rv that was grown until mid-log growth

349 phase (OD_{600} 0.5). Transduced *M. tuberculosis* H37Rv were spread on Mtb minimal

350 plate and Mtb YM rich plate and incubated at 37 °C for 2-3 weeks. To generate a

351 transposon library in saturated size, at least 150,000 colonies were collected from each

352 plate. Each transposon library was aliquoted and stored at -80 °C. Each transposon

353 library was generated in duplicates.

354 **Tn-seq.** Genomic DNA (gDNA) was prepared from each sample as previously

355 described (44). gDNA was then fragmented using an S220 Acoustic DNA Shearing Device

356 (Covaris). After shearing, adapters were added using an Illumina TruSeq Nano DNA library

357 prep kit according to the manufacturer's instructions. Transposon junctions were amplified

358 by using a transposon specific primer Mariner_1R_TnSeq_noMm

359 (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG<u>CCGGGGACTTATCAGCCAACC</u>)

360 and p7 primers (CAAGCAGAAGACGGCATACGAG) with HotStarTaq Master Mix Kit

361 (Qiagen) with the following PCR condition (94 °C for 3 min, 30 cycles of 94 °C for 30

362 seconds, 65 °C for 30 seconds, and 72 °C for 60 seconds, 72 °C for 10 minutes). The

363 transposon junction enriched sample was diluted 1:50 with water and then amplified to add

364 the flow cell adapter and i5 index to the enriched Tn-containing fragments using the

365 following primers:

366 i5 indexing primer [XXXXXXX denotes the position of the 8 bp index sequence):

367 AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTCGTCGGCAGCGTC

16

368   p7 primer: CAAGCAGAAGACGGCATACGAG

369   The amplification reaction was as follows: 5 µl template DNA (from PCR 1), 1 µl nuclease-

370   free water, 2 µl 5x KAPA HiFi buffer (Kapa Biosystems), 0.3 µl 10 mM dNTPs (Kapa

371   Biosystems), 0.5 µl DMSO (Fisher Scientific), 0.2 µl KAPA HiFi Polymerase (Kapa

372   Biosystems, 0.5 µl i5 indexing primer (10 µM), 0.5 µl p7 primer (10 µM). Cycling conditions

373   were: 95°C for 5 minutes, followed by 10 cycles of 98°C for 20 seconds, 63°C for 15

374   seconds, 72°C for 1 minute, followed by a final extension at 72°C for 10 minutes.

375   Amplification products were purified with AMPure XP beads (Beckman Coulter) and the

376   uniquely indexed libraries were quantified using a Quant-IT PicoGreen dsDNA assay

377   (Thermo Fisher Scientific) and the resulting fragment size distribution was assessed

378   using a Bioanalyzer (Agilent Technologies). The resultant Tn-seq library was sequenced

379   using a HiSeq 2500 HO, 125 bp PE run using v4 chemistry (Illumina).

380   **Tn-seq analysis.** Sequence reads were trimmed using CutAdapt (45). We first

381   trimmed sequence reads for transposon sequences

382   (CCGGGGACTTATCAGCCAACCTGT) at the 5` ends. Reads that did not contain

383   transposon sequence at the 5` end were discarded. After the 5` trimming process, all the

384   sequence reads begin with TA. We then trimmed sequence reads for adaptor sequences

385   ligated to 3`end (GATCCCACTAGTGTCGACACCAGTCTC). After the trimming, we

386   discarded the sequence reads that were shorter than 18 bp. The default error rate of 0.1

387   was used for all for all trimming processes.

388   The trimmed sequence reads were mapped (allowing 1 bp mismatch) to the *M.*

389   *tuberculosis* H37Rv genome (GenBank: AL123456.3) using Bowtie 2 (46). The number of

390   reads at each TA site were counted and converted to the .wig format, the input file format

17

391   for TRANSIT (14) using a custom Python script (S1 Appendix). Subsequent statistical

392   analysis for gene essentiality (Bayesian/Gumbel Method, HMM, and resampling method)

393   were performed using TRANSIT (version, 2.0.2) (14).

394   The Bayesian/Gumbel Method determines posterior probability of essentiality of

395   each gene (shown in zbar column in Dataset S1). When the value is 1 or near 1 within the

396   threshold, gene is called essential. When the value is zero or near zero the threshold,

397   gene is called non-essential. When the value is between the two threshold, neither near

398   zero nor 1), gene is called uncertain. When the value is -1, gene is called small because it

399   is considered that gene is too small to determine posterior probability of essentiality. Thus,

400   we analyzed essentialities of small and uncertain genes by HMM. All essential genes

401   identified from uncertain or small genes were listed in Table S2.

402   Total 601 genes (https://biocyc.org/group?id=biocyc14-7907-3764257976) essential

403   for *M. tuberculosis* survival in the Mtb YM rich medium (542 genes by the

404   Bayesian/Gumbel Method and additional 59 genes by the HMM) were used for the

405   comparisons to two past Tn-seq studies (8, 9). List of essential genes identified by past

406   studies were obtained from Dataset S1 in ref (8) and ref (9).

407   **Flux Balance Analysis (FBA).** FBA solutions were obtained using a simulated

408   environment designed to mimic the Mtb YM rich medium designed for this study. This

409   was done by altering uptake boundaries to match the concentration of each metabolite.

410   Most metabolites present in the media were given unlimited boundaries, because these

411   metabolites were not expected to be limiting, and also because they were present at an

412   undefined concentration in the Mtb YM rich medium, due to their source being casamino

413   acids. Metabolites added to the Mtb YM rich medium in known concentrations were

414   bounded in FBA at those concentrations.

415   The iSM810 model contains 938 metabolic reactions and 810 genes (including 1

416   orphan gene) (39). The biomass reaction originally described for iSM810 was chosen to

417   define growth. FBA was performed using the COBRA Toolbox Matlab package (47, 48).

418   The unconstrained uptake fluxes were set to '1'. Gene essentiality was assessed using the

419   COBRA Toolbox single gene deletion function in Matlab. Through single gene deletion,

420   reactions associated with each gene were systematically closed and the model was

421   optimized for biomass production. Any biomass accumulation > 1e-10 (which could occur

422   due to numerical errors) was defined as growth and the gene was classified as non-

423   essential. Biomass accumulation <1e-10 resulted in a gene being called essential. All FBA

424   optimizations were done using the Gurobi Optimizer 7.0 software under the free academic

425   license (49).

426   **Comparative genomic analysis.**

427   Following comparative genomic analysis was carried out as previously reported (50).

428   In brief, the pan and core genomes were defined using Roary software (51). Complete and

429   draft genome sequences of pathogenic strains (non-highlighted strains were used and

430   obtained from PATRIC database (1st October, 2017) summarized in Table S6 were

431   reannotated to generate gff3 files using PROKKA ver. 1.1.12 software (52) and to include

432   annotation of a reference strain H37Rv. Homologous proteins (i.e. protein families) were

433   clustered using the CD-Hit and MCL algorithms. The BLASTp cut-off value was set at 95%.

434   The number of core- and pan-genome protein families were estimated via genome

435   sampling up to the number of input genomes at the default setting in Roary (Dataset S2).

19

Acknowledgments

References

1. World Health Organization. 2017. Global Tuberculosis Report.
2. Zumla A, Nahid P, Cole ST. 2013. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov 12:388-404.
3. van Opijnen T, Camilli A. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol 11:435-42.
4. Zhang YJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, Trauner A, Wallis D, Galaviz S, Huttenhower C, Sacchettini JC, Behar SM, Rubin EJ. 2013. Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. Cell 155:1296-308.
5. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. 2013. Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. PLoS Pathog 9:e1003800.
6. van Opijnen T, Camilli A. 2012. A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res 22:2541-51.
7. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, Rubin EJ. 2012. Global assessment of genomic regions required for growth in *Mycobacterium tuberculosis*. PLoS Pathog 8:e1002946.
8. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7:e1002251.
9. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, Rubin EJ, Schnappinger D, Ehrt S, Fortune SM, Sassetti CM, Ioerger TR. 2017. Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. MBio 8.
10. Carey AF, Rock JM, Krieger IV, Chase MR, Fernandez-Suarez M, Gagneux S, Sacchettini JC, Ioerger TR, Fortune SM. 2018. TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. PLoS Pathog 14:e1006939.
11. Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, Sassetti CM. 2012. Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. Chem Biol 19:218-27.

472   12.   Sassetti CM, Boyd DH, Rubin EJ. 2001. Comprehensive identification of conditionally essential
473   genes in mycobacteria. Proc Natl Acad Sci U S A 98:12712-7.
474   13.   Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high
475   density mutagenesis. Mol Microbiol 48:77-84.
476   14.   DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR. 2015. TRANSIT--A Software Tool for
477   Himar1 TnSeq Analysis. PLoS Comput Biol 11:e1004401.
478   15.   DeJesus MA, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, Ioerger TR. 2013. Bayesian analysis
479   of gene essentiality based on sequencing of transposon insertion libraries. Bioinformatics
480   29:695-703.
481   16.   Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR.
482   2003. A postgenomic method for predicting essential genes at subsaturation levels of
483   mutagenesis: application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 100:7213-8.
484   17.   Lew JM, Kapopoulou A, Jones LM, Cole ST. 2011. TubercuList--10 years after. Tuberculosis
485   (Edinb) 91:1-7.
486   18.   Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M,
487   Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD. 2016. The
488   MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of
489   pathway/genome databases. Nucleic Acids Res 44:D471-80.
490   19.   Chen WH, Minguez P, Lercher MJ, Bork P. 2012. OGEE: an online gene essentiality database.
491   Nucleic Acids Res 40:D901-6.
492   20.   Sambandamurthy VK, Wang X, Chen B, Russell RG, Derrick S, Collins FM, Morris SL, Jacobs WR.
493   2002. A pantothenate auxotroph of *Mycobacterium tuberculosis* is highly attenuated and
494   protects mice against tuberculosis. Nat Med 8:1171-4.
495   21.   Pandey AK, Sassetti CM. 2008. Mycobacterial persistence requires the utilization of host
496   cholesterol. Proc Natl Acad Sci U S A 105:4376-80.
497   22.   Jackson M, Phalen SW, Lagranderie M, Ensergueix D, Chavarot P, Marchal G, McMurray DN,
498   Gicquel B, Guilhot C. 1999. Persistence and protective efficacy of a *Mycobacterium tuberculosis*
499   auxotroph vaccine. Infect Immun 67:2867-73.
500   23.   Thiede JM, Kordus SL, Turman BJ, Buonomo JA, Aldrich CC, Minato Y, Baughn AD. 2016.
501   Targeting intracellular *p*-aminobenzoic acid production potentiates the anti-tubercular action of
502   antifolates. Sci Rep 6:38083.
503   24.   Gopinath K, Venclovas C, Ioerger TR, Sacchettini JC, McKinney JD, Mizrahi V, Warner DF. 2013. A
504   vitamin B_{12} transporter in *Mycobacterium tuberculosis*. Open Biol 3:120175.
505   25.   Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C,
506   Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S,
507   Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp
508   PD. 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic
509   Acids Res 45:D543-D550.
510   26.   D'Elia MA, Pereira MP, Brown ED. 2009. Are essential genes really essential? Trends Microbiol
511   17:433-8.
512   27.   Gouzy A, Larrouy-Maumus G, Bottai D, Levillain F, Dumas A, Wallach JB, Caire-Brandli I, de
513   Chastellier C, Wu TD, Poincloux R, Brosch R, Guerquin-Kern JL, Schnappinger D, Sório de
514   Carvalho LP, Poquet Y, Neyrolles O. 2014. *Mycobacterium tuberculosis* exploits asparagine to
515   assimilate nitrogen and resist acid stress during infection. PLoS Pathog 10:e1003928.
516   28.   Kieser KJ, Boutte CC, Kester JC, Baer CE, Barczak AK, Meniche X, Chao MC, Rego EH, Sassetti CM,
517   Fortune SM, Rubin EJ. 2015. Phosphorylation of the Peptidoglycan Synthase PonA1 Governs the
518   Rate of Polar Elongation in Mycobacteria. PLoS Pathog 11:e1005010.

519 29. Vandal OH, Roberts JA, Odaira T, Schnappinger D, Nathan CF, Ehrt S. 2009. Acid-susceptible
520 mutants of *Mycobacterium tuberculosis* share hypersusceptibility to cell wall and oxidative
521 stress and to the host environment. J Bacteriol 191:625-31.
522 30. Kieser KJ, Baranowski C, Chao MC, Long JE, Sassetti CM, Waldor MK, Sacchettini JC, Ioerger TR,
523 Rubin EJ. 2015. Peptidoglycan synthesis in *Mycobacterium tuberculosis* is organized into
524 networks with varying drug susceptibility. Proc Natl Acad Sci U S A 112:13087-92.
525 31. Warner DF, Savvi S, Mizrahi V, Dawes SS. 2007. A riboswitch regulates expression of the
526 coenzyme B12-independent methionine synthase in *Mycobacterium tuberculosis*: implications
527 for differential methionine synthase function in strains H37Rv and CDC1551. J Bacteriol
528 189:3655-9.
529 32. Gordhan BG, Smith DA, Alderton H, McAdam RA, Bancroft GJ, Mizrahi V. 2002. Construction and
530 phenotypic characterization of an auxotrophic mutant of *Mycobacterium tuberculosis* defective
531 in L-arginine biosynthesis. Infect Immun 70:3080-4.
532 33. Movahedzadeh F, Smith DA, Norman RA, Dinadayala P, Murray-Rust J, Russell DG, Kendall SL,
533 Rison SC, McAlister MS, Bancroft GJ, McDonald NQ, Daffe M, Av-Gay Y, Stoker NG. 2004. The
534 *Mycobacterium tuberculosis ino1* gene is essential for growth and virulence. Mol Microbiol
535 51:1003-14.
536 34. Pavelka MS, Jacobs WR. 1999. Comparison of the construction of unmarked deletion mutations
537 in *Mycobacterium smegmatis*, *Mycobacterium bovis* bacillus Calmette-Guérin, and
538 *Mycobacterium tuberculosis* H37Rv by allelic exchange. J Bacteriol 181:4780-9.
539 35. Murphy HN, Stewart GR, Mischenko VV, Apt AS, Harris R, McAlister MS, Driscoll PC, Young DB,
540 Robertson BD. 2005. The OtsAB pathway is essential for trehalose biosynthesis in
541 *Mycobacterium tuberculosis*. J Biol Chem 280:14524-9.
542 36. Korte J, Alber M, Trujillo CM, Syson K, Koliwer-Brandl H, Deenen R, Köhrer K, DeJesus MA,
543 Hartman T, Jacobs WR, Bornemann S, Ioerger TR, Ehrt S, Kalscheuer R. 2016. Trehalose-6-
544 Phosphate-Mediated Toxicity Determines Essentiality of OtsB2 in *Mycobacterium tuberculosis* In
545 Vitro and in Mice. PLoS Pathog 12:e1006043.
546 37. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard
547 JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE,
548 Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H,
549 Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and
550 Analysis Resource Center. Nucleic Acids Res 45:D535-D542.
551 38. O'Brien EJ, Monk JM, Palsson BO. 2015. Using Genome-scale Models to Predict Biological
552 Capabilities. Cell 161:971-987.
553 39. Ma S, Minch KJ, Rustad TR, Hobbs S, Zhou SL, Sherman DR, Price ND. 2015. Integrated Modeling
554 of Gene Regulatory and Metabolic Networks in *Mycobacterium tuberculosis*. PLoS Comput Biol
555 11:e1004543.
556 40. Orth JD, Thiele I, Palsson B. 2010. What is flux balance analysis? Nat Biotechnol 28:245-8.
557 41. Pelicic V, Jackson M, Reyrat JM, Jacobs WR, Gicquel B, Guilhot C. 1997. Efficient allelic exchange
558 and transposon mutagenesis in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 94:10955-
559 60.
560 42. Kriakov J, Lee S, Jacobs WR. 2003. Identification of a regulated alkaline phosphatase, a cell
561 surface-associated lipoprotein, in *Mycobacterium smegmatis*. J Bacteriol 185:4983-91.
562 43. Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ. 1999. In vivo transposition
563 of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci U S A
564 96:1645-50.
565 44. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR. 2007. Genetic Manipulation of
566 *Mycobacterium tuberculosis*. Curr Protoc Microbiol Chapter 10:Unit 10A.2.

567 45.   Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
568   EMBnetjournal; Vol 17, No 1: Next Generation Sequencing Data Analysis.
569 46.   Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of
570   short DNA sequences to the human genome. Genome Biol 10:R25.
571 47.   Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ. 2007. Quantitative prediction
572   of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc 2:727-38.
573 48.   Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis
574   NE, Rahmanian S, Kang J, Hyduke DR, Palsson B. 2011. Quantitative prediction of cellular
575   metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:1290-307.
576 49.   Gurobi Optimization, Inc. Gurobi Optimizer Reference Manual. Available at:
577   http://www.gurobi.com

578 50.   Yano H, Iwamoto T, Nishiuchi Y, Nakajima C, Starkova DA, Mokrousov I, Narvskaya O, Yoshida S,
579   Arikawa K, Nakanishi N, Osaki K, Nakagawa I, Ato M, Suzuki Y, Maruyama F. 2017. Population
580   Structure and Local Adaptation of MAC Lung Disease Agent *Mycobacterium avium* subsp.
581   *hominissuis*. Genome Biol Evol 9:2403-2417.
582 51.   Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA,
583   Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics
584   31:3691-3.
585 52.   Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.

586

587 **Supporting Information Legends**

588 **Table S1. Composition of Mtb YM rich, Mtb minimal media.**

589 **Table S2 List of short and uncertain genes that were identified as essential by**

590 **TRANSIT HMM.**

591 **Table S3 List of essential genes identified in this study and past studies.**

592 **Table S4 Comparison of essential genes between Mtb YM rich and Mtb minimal**

593 **media identified by Tn-seq and analyzed by TRANSIT re-sampling method.**

594 **Table S5 Comparison of essential metabolic pathways between Mtb YM rich and**

595 **Mtb minimal media.**

596 **Table S6 Complete and draft genome sequences of virulent *M. tuberculosis***

597 **complex used for comparative genomic analysis.**

598 **Table S7 *in silico* growth predictions of a single gene knockout mutant strains of**

599 ***M. tuberculosis.***

600 **Dataset S1 Analysis of Tn-seq gene essentiality analysis by TRANSIT**

601 **Bayesian/Gumbel Method and Hidden Markov Model.**

602 **Data S2. Presence or absence of genes within complete genomes of the**

603 ***Mycobacterium tuberculosis* complex analyzed by Roary.**

604 **Appendix S1. Custom Python script used in this analysis.**

Figure Legends

Figure 1. Identification of essential genes in Mtb YM rich medium. (A) Media composition of Mtb minimal and Mtb YM rich. 7H9 media contains glutamic acid and ammonium sulfate in addition to the nutrients shown in bold. DAP, diaminopimelic acid. FAC, ferric ammonium citrate. PABA, *para*-aminobenzoic acid. *, 0.5 % casamino acids and 0.98 mM tryptophan. (B) *M. tuberculosis* H37Rv growth in Mtb YM rich (red), 7H9 (black), and Mtb minimal (blue). (C) Bayesian/Gumbel results for *M. tuberculosis* H37Rv in Mtb YM rich. (D) HMM results for *M. tuberculosis* H37Rv in Mtb YM rich. (E) Venn diagram comparisons of essential genes by Tn-seq in this study and two previous studies (8, 9). This study (red), Griffin *et al* (blue) (8), and DeJesus *et al* (green) (9).

Figure 2. Identification of conditionally essential genes in Mtb YM rich and Mtb minimal media. (A) Results for comparative analysis between Mtb YM rich and Mtb minimal media. Numbers of conditionally essential genes (adjusted p-value <0.05) were identified by the resampling method in TRANSIT (14). (B, C) Essentialities of genes in pantothenate biosynthesis pathway (B) and methionine biosynthesis pathway (C). Genes in red color (essential in both Mtb YM rich and Mtb minimal), genes in blue color (essential in Mtb minimal but not essential in Mtb YM rich), genes in orange color (essential in Mtb YM rich but not essential in Mtb minimal) genes in black color (non-essential). Information on genes in each pathway was obtained from BioCyc database (18). (D) Ratio of essential genes in each metabolic pathway. All genes are listed in Table S5. NAD, Nicotinamide adenine dinucleotide. PABA, *para*-aminobenzoic acid. purine, 5-aminoimidazole ribonucleotide.

25

Figure 3. Identification of highly conserved essential genes among virulent *M. tuberculosis* complex. (A) Strains used for comparative genomic analysis. (B) Results for pan-genome analysis by Roary. (C) Number of highly conserved genes among virulent *M. tuberculosis* complex. Core genes; conserved in >99 % strains, soft core genes; conserved in 95-99 % strains, shell genes; conserved in 15-95% strains, cloud genes; conserved in 0-15 % strains. (D) Number of genes that are essential for *M. tuberculosis* H37Rv survival in Mtb YM rich and highly conserved among virulent *M. tuberculosis* complex strains.

Figure 4. Comparison of Tn-seq identified essential genes and *in silico* predicted essential genes. (A) Results for *in silico* gene essentiality prediction by iSM810 and iSM810 (modified). (B) Comparison of Tn-seq identified essential genes (blue) and iSM810 predicted essential genes (red). (C) Comparison of Tn-seq identified essential genes and iSM810 predicted essential genes in purine biosynthesis pathway. Genes in red color (essential in both Mtb YM rich and Mtb minimal), genes in blue color (essential in Mtb minimal but not essential in Mtb YM rich), genes in black color (non-essential). Information on genes in each pathway was obtained from BioCyc database (18).

643 **Table 1 Gene essentialities and *in silico* essentiality predictions of anti-tubercular**

644 **drug target genes.**

Gene (Rv number)	Drug	Essentiality in Mtb YM rich media	Core/soft core gene (conserved in % strains)
inhA(*Rv1484*)	isoniazid/ethionamide	essential	yes (96%)
embA (*Rv3794*)	ethambutol	essential	no (79%)
embB (*Rv3795*)	ethambutol	essential	no (83%)
rpoB (*Rv0667*)	rifampin	essential	yes (97%)
atpE (*Rv1305*)	bedaquiline	essential	yes (100%)
gyrA (*Rv0006*)	fluoroquinolones	essential	yes (99%)
gyrB (*Rv0005*)	fluoroquinolones	essential	no (89%)
dfrA (*Rv2763c*)	*para*-aminosalicylic acid	essential	yes (99%)
alr (*Rv3423c*)	D-cycloserine	essential	yes (98%)

645

646

647

648

649

A — Mtb YM rich medium

Mtb minimal medium

glycerol
asparagine
KH$_2$PO$_4$
Na$_2$HPO$_4$
FAC
CaCl$_2$
ZnSO$_4$
MgSO$_4$

acetate
glucose
pyruvate
ribose
amino acids*
biotin
citrate
DAP
hemin
hypoxanthine
lipoic acid

mio-inositol
nicotinamide
oleic acid
PABA
pantothenate
pyridoxin
thiamine
thymine
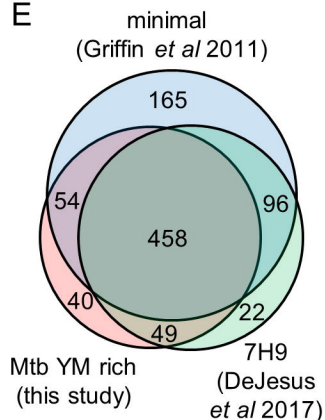trehalose
vitamin B$_{12}$
riboflavin
CuSO$_4$

B

growth (O.D.$_{600}$) vs time (days)
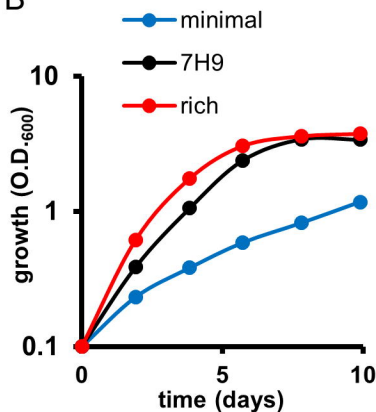
- minimal
- 7H9
- rich

C

	numbers of genes
essential	542
uncertain	115
non-essential	3102
short	231

D

	% of TA sites
essential	13.3
growth defect	2.8
non-essential	78.6
growth advantage	5.3

E

minimal (Griffin *et al* 2011): 165

54

96

458

40

49

22

Mtb YM rich (this study)

7H9 (DeJesus *et al* 2017)

A

	number of genes
conditionally essential in Mtb minimal	130
conditionally essential in Mtb YM rich	98

B

3-methyl-2-oxobutanoate

oxaloacetate

aspB *aspC*

panB

L-aspartate

2-dehydropantoate

panD

pantoate β-alanine

panC

asnA

pantothenate

coaA

L-asparagine

phosphopantothenate

ansP2

C

L-cysteine

O-succinyl-L-homoserine

metB

L-cystathionine

L-homocycteine

metH *metE*

L-methionine
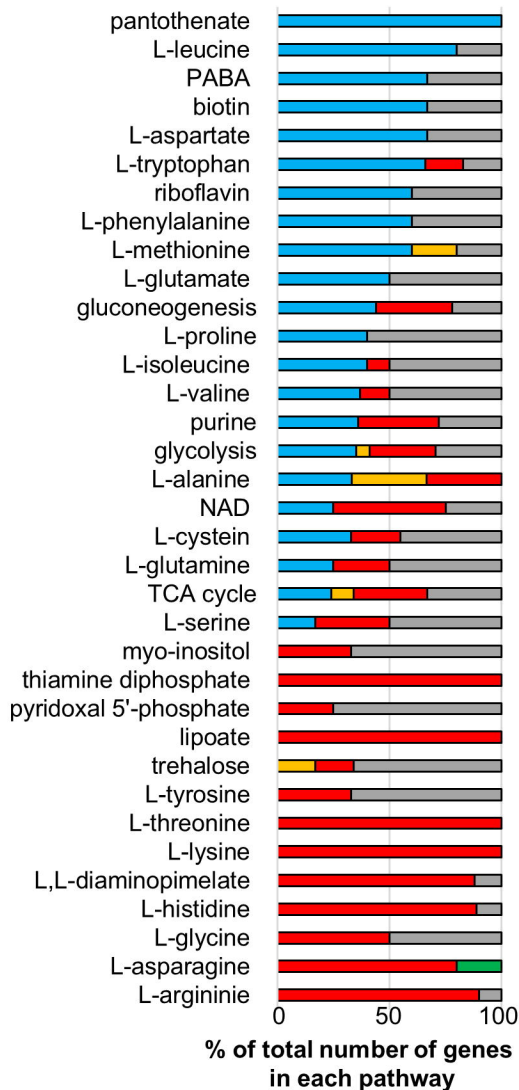
D
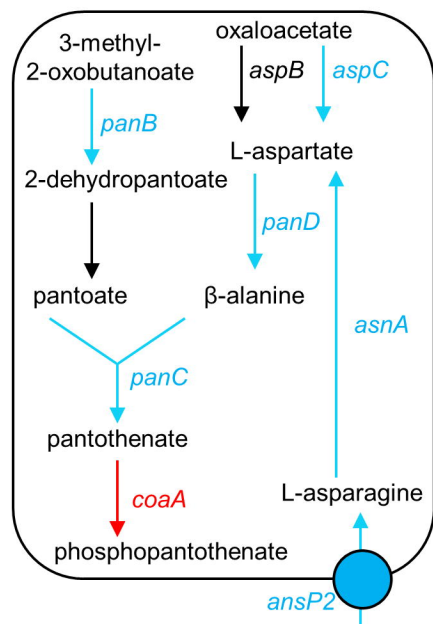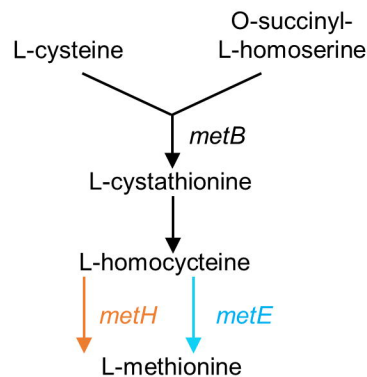
- conditionally essential in Mtb minimal
- conditionally essential in Mtb YM rich
- essential in both conditions
- non essential
- uncertain

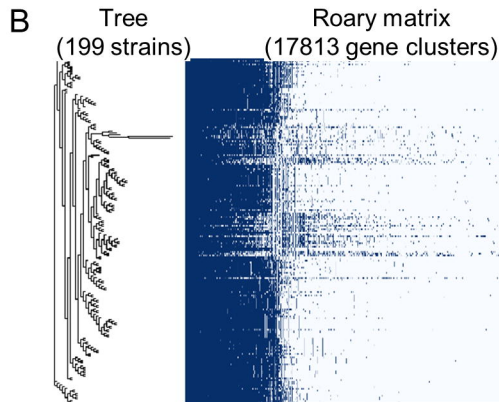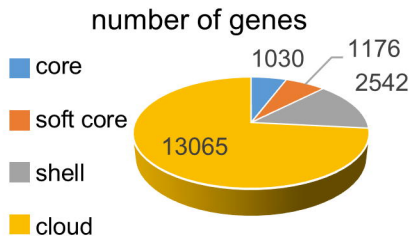**biosynthesis pathway**

pantothenate
L-leucine
PABA
biotin
L-aspartate
L-tryptophan
riboflavin
L-phenylalanine
L-methionine
L-glutamate
gluconeogenesis
L-proline
L-isoleucine
L-valine
purine
glycolysis
L-alanine
NAD
L-cystein
L-glutamine
TCA cycle
L-serine
myo-inositol
thiamine diphosphate
pyridoxal 5'-phosphate
lipoate
trehalose
L-tyrosine
L-threonine
L-lysine
L,L-diaminopimelate
L-histidine
L-glycine
L-asparagine
L-argininie

% of total number of genes in each pathway

A

Species	number of strains
M. africanum	3
M. bovis	4
M. canettii	5
M. caprae	1
M. microti	1
M. tuberculosis	185
Total	199

C

number of genes

- core
- soft core
- shell
- cloud

1030
1176
2542
13065

B

Tree
(199 strains)

Roary matrix
(17813 gene clusters)

D

core or soft core genes

356

601
essential genes
in Mtb YM rich
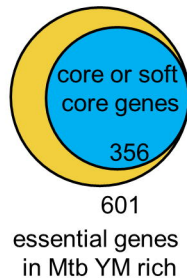
A

	iSM810		iSM810 (modified)	
	Mtb YM rich	Mtb minimal	Mtb YM rich	Mtb minimal
essential	156	221	141	221
non-essential	654	589	669	589
total	810	810	810	810

B

i) Mtb YM rich

136 | 114 | 42
Tn-seq | iSM810
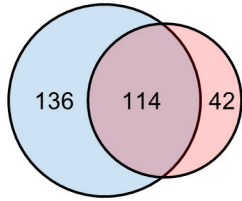
Sensitivity (46%)
Specificity (93%)

140 | 110 | 31
Tn-seq | iSM810 (modified)
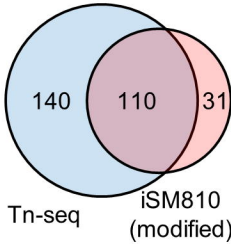
Sensitivity (44%)
Specificity (95%)

ii Mtb minimal

133 | 173 | 48
Tn-seq | iSM810
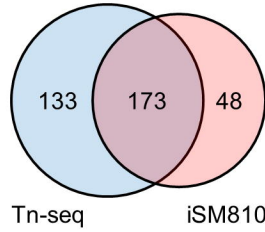
Sensitivity (57%)
Specificity (91%)

C

5-phospho-α-D-ribose 1-diphosphate
→ *purF* →
L-glycine   5-phospho-β-D-ribosylamine
→ *purD* →
N^1-(5-phospho-β-D-ribosyl) glycinamide
→ *purN* →
N^2-formyl-N^1-(5-phospho-β-D-ribosyl)glycinamide
→ *purQ* →   *purL*
2-(formamido)-N^1-(5-phospho-β-D-ribosyl) acetamidine
→ *purM* →
5-amino-1-(5-phospho-β-D-ribosyl)imidazole
→ *purK* →
N^5-carboxyaminoimidazole ribonucleotide
→ *purE* →
5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxylate
→ *purC* →
5'-phosphoribosyl-4-(N-succinocarboxamide)-5-aminoimidazole
→ *purB* →
5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide
→ *purH* →
IMP

	iSM810	iSM810 (modified)	Tn-seq
essential	*purB, purD, purF, purL, purM, purN*	*purB*	*purB, purC, purK, purH*
non-essential	*purC, purK, purH,*	*purC, purD, purF, purH, purK, purL, purM, purN*	*purD, purF, purL, purM, purN*