bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Genome-wide assessment of *Mycobacterium tuberculosis* conditionally essential metabolic pathways

Yusuke Minato[a], Daryl M Gohl [b], Joshua M. Thiede[a], Jeremy M. Chacón[c], William R. Harcombe[c], Fumito Maruyama[d,e,f], and Anthony D. Baughn [a#]

[a]Department of Microbiology and Immunology, University of Minnesota Medical School, Minneapolis, Minnesota, USA.

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

[c]Biotechnology Institute and Department of Ecology, Evolution and Behavior, University of Minnesota, Minnesota, USA.

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

[e]The Japan Science and Technology Agency/Japan International Cooperation Agency, Science and Technology Research Partnership for Sustainable Development (JST/JICA, SATREPS), Tokyo, Japan.

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

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 conditional 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 metabolites. 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*.

Author Summary

Mycobacterium tuberculosis causes 10 million cases of tuberculosis (TB) each year resulting in over one million deaths. TB therapy is challenging because it requires a minimum of six months of treatment with multiple drugs. Protracted treatment times and

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the emergent spread of drug resistant TB necessitate the identification of novel targets for drug discovery to curb this global health threat. Essential functions, defined as those indispensable for growth and/or survival, are potential targets for new antimicrobial drugs. In this study, we aimed to define gene essentialities of *M. tuberculosis* on a genome-wide scale to comprehensively identify potential targets for drug discovery. We utilized a combination of experimental (functional genomics) and *in silico* approaches (comparative genomics and flux balance analysis). Our functional genomics approach identified sets of genes whose essentiality was affected by nutrient availability. Comparative genomics revealed that not all essential genes were fully conserved within the *M. tuberculosis* complex. Comparing sets of essential genes identified by functional genomics to those predicted by flux balance analysis highlighted gaps in current knowledge regarding *M. tuberculosis* metabolic capabilities. Thus, our study identifies numerous potential anti-tubercular drug targets and provides a comprehensive picture of the complexity of *M. tuberculosis* essential cellular functions.

Introduction

Mycobacterium tuberculosis is responsible for approximately 10.4 million new cases of active tuberculosis (TB) infection and 1.4 million deaths annually [1]. While TB chemotherapy has a high success rate in curing drug susceptible TB infections, it is challenging, in part because it requires a minimum of 6 months of treatment with drugs associated with adverse reactions. Thus, finding targets for new TB drugs that are more potent than existing drugs is needed [2].

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Essential genes, defined as genes indispensable for growth and/or survival, are potential targets for new types of antimicrobial drugs. Gene essentiality can be assessed by targeted gene-disruptions, where genes that cannot be disrupted are typically categorized as being essential. However, such traditional genetic approaches are labor-intensive and not easily adaptable to genome-scale screening. Recent advances in Next-Generation Sequencing (NGS)-based approaches have transformed our ability to examine gene functions in a genome-wide manner. Transposon insertion sequencing (Tn-seq) has been widely used to conduct fitness profiling of gene functions in many bacterial species, including *M. tuberculosis* [3-10]. In addition to fitness profiling, a lack of representation of specific transposon insertions within a saturated transposon library has been used to identify essential genes in genome-wide screens [3, 9, 10].

In *M. tuberculosis*, several systematic genome-wide studies, including studies using Tn-seq, have been performed to identify essential genes *in vitro* [7-16] and *in vivo* [4]. The gene essentialities determined by Tn-seq studies are accessible through publicly available databases, such as Tuberculist [17], Biocyc [18], and OGEE [19]. Most of these gene essentiality data were obtained from Tn-seq studies that were carried out using a defined growth medium that was supplemented with a limited number of nutrients [7-10, 13, 15]. Under this growth condition, genes for numerous essential central metabolic pathways can be rendered dispensable through supplementation. Thus, these previous studies have likely miscategorized a large set of genes as essential rather than conditionally essential. For example, pantothenate, an essential precursor in coenzyme A biosynthesis, is not supplemented in most of the commonly used media for *M. tuberculosis*. As a result, *panC*, encoding pantothenate

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

synthetase, the enzyme catalyzing the last step in pantothenate biosynthesis, was defined as an essential gene in these three databases [17-19] despite a previous study showing that the *panCD* double deletion mutant of *M. tuberculosis* H37Rv strain can grow in the presence of supplemental pantothenate [20]. Thus, more precise definition of *M. tuberculosis* essential genes on a genome-wide scale, and annotation of genes which are conditionally essential, would improve the usefulness of these databases.

In this study, we developed a defined-nutrient rich medium (Mtb YM rich) that included a variety of nutrient sources for *M. tuberculosis* to ease identification of conditionally essential genes. We used Tn-seq to identify essential genes in Mtb YM rich and minimal medium (Mtb minimal). As expected, essentialities of many genes involved in metabolite biosynthesis pathways were affected by the supplemental nutrients in Mtb YM rich medium. However, we found that essentialities of certain metabolic pathways were unaffected by the absence or presence of relevant nutrient sources. We also found that both growth conditions had some essential genes that were unique to each environment. In addition, we compared essential genes identified by Tn-seq with highly conserved genes that were identified by a comparative genomics analysis and a modified *in silico* metabolic model. These comparisons indicated that essential genes were highly enriched amongst the conserved core genome and that such gene essentiality measurements can be used to refine metabolic models.

Results and Discussions

Identification of *M. tuberculosis* essential genes in a chemically defined nutrient-rich media. We developed a defined nutrient-rich medium for *M. tuberculosis*

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Mtb YM rich medium) (Fig. 1A, Table S1) by supplementing numerous nutrients into a minimal *M. tuberculosis* growth medium (Mtb minimal medium) [21]. The supplements included several carbon sources, nitrogen sources, co-factors, amino acids, nucleotide bases and other nutrients. Four of the supplements, lipoic acid, nicotinamide, hemin, and ribose, inhibited *M. tuberculosis* H37Rv growth at higher concentrations. Thus, we utilized concentrations of these supplements that did not impair growth. The Mtb YM rich medium supported the *M. tuberculosis* H37Rv growth similar to the commonly used 7H9 medium that contains essential salts and relatively few nutrients (Fig. 1B).

We next generated a library of *M. tuberculosis* H37Rv transposon insertion mutants on Mtb YM rich medium plates. As the *M. tuberculosis* H37Rv genome contains 74,602 TA dinucleotides, we collected at least 150,000 colonies to approach saturation of *himar1* transposon insertion sites in the library. Genomic DNA was isolated from the library. Next the DNA was sheared, end repaired, and sequencing adapters were added, and the transposon adjacent regions were enriched by PCR before massively parallel sequencing. The resultant sequencing data was analyzed using TRANSIT, a recently developed software package for analyzing Tn-seq data [14]. TRANSIT contains two statistical methods, Bayesian/Gumbel Method and Hidden Markov Model (HMM), to identify essential genes or essential genomic regions, respectively, in a single growth condition. By using the Bayesian/Gumbel Method, we found that 542 genes were essential for *M. tuberculosis* growth in the Mtb YM rich medium (Fig. 1C, Table S2). The Bayesian/Gumbel method assesses gene essentiality based on consecutive sequence of TA sites lacking insertion within a gene. However, one limitation of this method is the inability to assess the essentiality of short genes. As a result, 231 short genes were not

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

classified by this analysis (Fig. 1C). To overcome this issue, we also identified essential genomic regions using an HMM [14]. The HMM is based on the read count at a given site and the distribution over the surrounding sites. This analysis identified 13.3 % of TA sites as essential (Fig. 1D, Table S3). These essential regions included 15 short genes whose essentiality was not able to be assessed by the Bayesian/Gumbel Method. In total, we identified 557 genes (542 genes by the Bayesian/Gumbel Method and 15 short essential genes by the HMM) as essential genes for *M. tuberculosis* survival in the Mtb YM rich medium. A list of these genes is available through the BioCyc smart table format (https://biocyc.org/group?id=biocyc17-7907-3718043660) [22]. These essential genes included known targets for existing anti-tubercular drugs further validating that the Tn-seq assay successfully identified *M. tuberculosis* essential genes (Table 1).

Comparison of *M. tuberculosis* essential genes between Mtb YM rich medium and minimal nutrient medium. Because gene essentiality can be affected by the external environment [23], in particular by nutrient availability, we next compared gene essentiality between Mtb YM rich medium and Mtb minimal medium. Similar to the transposon library generated on Mtb YM rich medium plates, we generated *M. tuberculosis* H37Rv transposon libraries on Mtb minimal medium plates and collected at least 150,000 colonies to create a saturated library. The resultant sequencing data was compared with the Mtb YM rich plate data by using a permutation test-based method to identify genes with statistically significant differences in transposon insertion count (Fig. 2A, Table S4). We found that 130 genes were conditionally essential in the Mtb minimal medium compared to the Mtb YM rich medium (https://biocyc.org/group?id=biocyc13-7907-3706551227).

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As anticipated, many of the conditionally essential genes that were identified corresponded to the differences in nutrient composition between Mtb minimal and Mtb YM rich. For example, Mtb YM rich medium contained L-aspartate and pantothenate. Thus, the genes in L-aspartate and pantothenate biosynthesis pathways were dispensable in Mtb YM rich medium (Fig. 2B). However, Mtb YM rich medium did not contain metabolites located downstream of pantothenate. Consequently, these downstream genes (e.g. *coaA*, encoding pantothenate kinase) were essential in both Mtb YM rich and Mtb minimal media (Fig. 2B). This observation is consistent with a previous report that the *M. tuberculosis panCD* double deletion mutant strain can grow in the presence of pantothenate [20], and also suggested that single gene disruptions in the pantothenate biosynthesis pathway (*panB*, *panC*, and *panD*) show a similar pantothenate auxotrophic phenotype as the *panCD* double deletion mutant strain. Another example was the asparagine transporter gene, *ansP2* that was conditionally essential in Mtb minimal medium (Fig. 2B) presumably because Mtb minimal medium contained arginine as the sole nitrogen source whereas Mtb YM rich medium contained multiple nitrogen sources (Fig. 1A).

Unexpectedly, we also identified 98 genes that were conditionally essential in Mtb YM rich media compared to the Mtb minimal medium (https://biocyc.org/group?id=biocyc13-7907-3710604059). Such genes included *ponA1* and *ponA2,* encoding penicillin binding proteins (PBPs) involved in cell wall peptidoglycan (PG) biogenesis. It was previously shown that *ponA1* and *ponA2* are only essential *in vivo* but not during growth in culture medium [24, 25]. Thus, one of nutrients that is uniquely present in Mtb YM rich medium might be also present *in vivo* and may

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

be responsible for the *in vivo* fitness defect of the *ponA1* mutant. LdtB is one of the major L,D-transpeptidases (Ldts) that is also involved in PG biogenesis. *ldtB* was also conditionally essential in Mtb YM rich media. Interestingly, several genes that were conditionally essential in the absence of *ponA11, ponA2,* or *ldtB* (e.g. *Rv1086*, *Rv1248c*, *Rv3490, treS, otsA*) were also conditionally essential in Mtb YM rich media suggesting Mtb YM rich medium negatively affects PG biogenesis in *M. tuberculosis* [26]. We also found the *metH*, encoding one of the two methionine synthases, was only essential in Mtb YM rich medium whereas *metE*, encoding the other methionine synthase, was only essential in Mtb minimal medium (Fig.2C). The observed conditional essentialities were consistent with a previous study that *M. tuberculosis metE* expression was inhibited in the presence of vitamin B_{12} by a *metE* B_{12} riboswitch, and the vitamin B_{12}-dependent MetH is predominantly used when vitamin B_{12} is available [27]. These findings confirmed that the conditionally essential genes identified by Tn-seq in this study are consistent with previous findings [22].

The supplementation of nutrients in the Mtb YM rich medium could potentially subvert the need for enzymes in at least 35 metabolic pathways (Fig. 1A, Fig. 2D, and Table S5). In these pathways, we found fewer gene essentialities in 22 pathways in Mtb YM rich medium, suggesting that *M. tuberculosis* can functionally utilize these nutrients. Notably, many genes in these pathways were identified as essential genes by the previous studies [8, 13] and listed as essential genes in public database such as Tuberculist [17] and BioCyc [18]. Thus, our results revealed that these genes are essential only in the absence of the corresponding nutrients. In contrast, we also found that essentiality of genes in 13 other metabolic pathways were not altered by nutrient

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

supplementation (Fig. 2D). Previous studies have demonstrated that some auxotrophic mutants, such as those for L-arginine, L-lysine, and inositol, require supplementation with the corresponding nutrient at a relatively high concentration to support their growth [28-30]. Certain auxotrophic mutants are also known to show growth or survival defects even in the presence of the corresponding nutrient supplementation [31, 32]. These results were consistent with previous findings that essentiality of some central metabolic pathways could not be bypassed in pathogenic mycobacteria [12] and also provide a more comprehensive understanding of the nutrient utilization capacity of *M. tuberculosis*.

Identification of highly conserved genes in the *M. tuberculosis* complex. Essential genes that were identified by Tn-seq were further interrogated through comparative genomic analysis in order to see whether essentiality correlated with high levels of sequence conservation within the *M. tuberculosis* complex. A total of 226 complete genome sequences of *M. tuberculosis* complex were available from the PATRIC database [33]. We excluded non-pathogenic strains (e.g. *M. tuberculosis H37Ra* and *Mycobacterium bovis BCG*) and used 199 genome sequences for comparative genomic analysis (Fig. 3A, Table S6). We identified total of 17,813 genes from 199 strains of the *M. tuberculosis* complex (Fig. 3B and C). Among these genes, 2,363 genes (1,283 core genes and 1,028 soft core genes) were highly conserved in most of the strains ($\geq$ 95 % of strains).

Our Tn-seq analysis identified 557 genes as essential in Mtb YM rich media (Table S2, S3, smart table). Among them, approximately 60 % of the essential genes (349 genes) were included in the list of core/softcore genes (https://biocyc.org/group?id=biocyc17-7907-3723500017). Of note, not all of the target

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes for existing anti-tubercular drugs were categorized as core/softcore genes (Table 1).

Prediction of *M. tuberculosis* essential genes *in silico* Genome-scale metabolic models have been used to computationally simulate a range of cellular functions [34]. We utilized a genome-scale model of *M. tuberculosis*, iSM810, to predict gene essentiality *in silico* using flux balance analysis (FBA) [35]. The principles for FBA have been described previously [36]. Briefly, a metabolic network is represented by a stoichiometric matrix that contains reactions and metabolites. A biomass reaction is defined based on experimentally determined amounts of specific metabolites required for cellular growth. The stoichiometric matrix can be converted into a system of linear equations. FBA then solves this system by optimization, typically optimization of the biomass reaction. Constraints on reaction fluxes can be used to simulate metabolite availability or enzyme presence. For example, a reaction's uptake flux is constrained to zero when the simulated environment does not contain the associated metabolite. Reaction fluxes that simulate metabolite uptake within iSM810 were used to define *in silico* growth media. Uptake was not bounded for metabolites present in the selected growth medium, however; not all media components were represented within the iSM810 transport reactions (Data S1). Genes were assessed for essentiality by systematically closing the flux on each reaction within iSM810 and assessing simulated biomass production on each medium using FBA. Genes were defined essential if biomass production after knockout was zero.

iSM810 includes 810 metabolic genes and 938 metabolic reactions [35]. Among the 810 genes in iSM810, our FBA analysis predicted that 159 genes were essential in

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mtb YM rich medium and 221 genes were essential in Mtb minimal medium (Fig. 4A, Table S7). We then compared gene essentiality prediction by FBA with essential genes identified by Tn-seq (Fig. 4B). We found that the sensitivity of the FBA based gene essentiality prediction was low as there were a number of genes that were predicted to be essential by FBA but not identified as essential genes by Tn-seq (https://biocyc.org/group?id=biocyc17-7907-3715275234). For instance, we found that genes related to riboflavin biosynthesis were non-essential in Mtb YM rich medium by Tn-seq analysis, but were essential *in silico*. We examined why iSM810 could not accurately predict the essentiality of genes in the riboflavin biosynthesis pathway and found that the model lacked a transport reaction for riboflavin (Data S1). Similarly, we found that the model lacked transport reactions for vitamin B_{12}, *para*-aminobenzoic acid (PABA), H_2O, myo-inositol, and hypoxanthine. To investigate whether the addition of these transport reactions could improve gene essentiality prediction by iSM810, we added these transport reactions to the model (Data S1). These changes fixed multiple mis-matches between iSM810 and Tn-seq, by causing the model to determine that genes related to thiamine biosynthesis, riboflavin biosynthesis, and purine biosynthesis were non-essential in Mtb YM rich medium (Fig. 4 A and B, Table S7). Allowing PABA uptake caused no changes. After adding transport reactions, the growth rate prediction in Mtb YM rich medium increased from 0.055 g/L/day to 0.0876 g/L/day.

Unlike the Tn-seq results, the FBA predicted that all genes essential in Mtb YM rich medium were also essential in Mtb minimal and failed to predict any genes that were only essential in Mtb YM rich (Fig. S1). This is expected given the lack of gene

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

regulation implicit in our FBA modeling, and thus genes such as *metH* that was predicted to be non-essential in both Mtb minimal and Mtb YM rich media.

Comparing the sets of essential genes experimentally identified by Tn-seq to those computationally predicted *in silico* highlighted the limitations of current *M. tuberculosis* metabolic models. For example, we noticed that genes located downstream of purine biosynthesis pathway were identified as essential by Tn-seq in both Mtb YM rich and Mtb minimal media (Fig. 4C). However, even after the addition of a hypoxanthine transport reaction, there were still mismatches between iSM810 and Tn-seq. Among these genes, *purC* is the only gene that was characterized by targeted gene deletion analysis in *M. tuberculosis* [37]. The *purC* gene was difficult to delete [37] and the *purC* deletion mutant strain of *M. tuberculosis* showed a notable growth defect compared to the parent strain even in the presence of hypoxanthine [38]. This observation could explain why the *purC* gene was identified as essential by Tn-seq in Mtb YM rich. However, the other genes in purine biosynthesis pathway have not been characterized by targeted gene deletion studies. Thus, future studies are necessary to investigate essentiality of *purK*, *purB*, and *purH* by targeted gene deletion studies.

We also identified a potential avenue for the improvement of genome-scale model of *M. tuberculosis* by comparing Tn-seq identified essential genes and to those predicted *in silico.* We found that many genes in known essential metabolic pathways were not predicted to be essential (e.g. glycolysis, folate metabolism, mycolate biosynthesis, ATP biosynthesis, and several amino acid biosynthesis pathways) (https://biocyc.org/group?id=biocyc17-7907-3715282404). Examination of the biomass function in iSM810 and revealed that several known essential metabolites are not

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

connected to the biomass function in iSM810. For example, folate is not connected to the biomass function and this explains why addition of a PABA transport reaction did not affect any gene essentiality predictions. None of ATP synthase genes in iSM810 (*atpA*, *atpB*, *atpD*, *atpE*, *atpG*, and *atpH*) was predicted as essential because these were linked to the same reaction using "or" Boolean logic, meaning the presence of one gene was sufficient to allow the reaction to proceed. Thus, further improvement of iSM810 for more accurate prediction of essential genes could be achieved by connecting such essential metabolites to the biomass function.

Conclusions. In this study, we utilized functional genomics and comparative genomics approaches to identify *M. tuberculosis* essential genes. We identified distinct sets of essential and conditionally essential genes from different approaches and different growth conditions. In addition, comparison of essential genes identified by functional genomics approach to *in silico* predicted essential genes highlighted current gaps in knowledge regarding *M. tuberculosis* metabolism. Our study provides a promising platform to shed new light on essential cellular functions in *M. tuberculosis* that can lead to the discovery of novel targets for antitubercular drugs.

Materials and methods

Media and growth conditions. *M. tuberculosis* H37Rv was grown aerobically at 37 °C in Middlebrook 7H9 medium supplemented with oleate-albumin-dextrose-catalase (OADC; 10%, vol/vol), glycerol (0.2%, vol/vol), and tyloxapol (0.05%, vol/vol) unless otherwise noted. Mtb minimal medium (0.2%, vol/vol glycerol as a sole carbon source) [21] and Mtb YM rich medium (Table S1) agar plates were used to generate *M.*

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tuberculosis H37Rv transposon libraries. Tyloxapol (0.05%, vol/vol) were added to the both Mtb minimal and Mtb YM rich agar plates.

Construction of saturated transposon libraries of *M. tuberculosis*. Transposon mutagenesis were performed as previously described [39]. Mycobacteriophage phAE180 [39] was used to transduce a mariner derivative transposon Tn5371 [40] to *M. tuberculosis* H37Rv that was grown until mid-log growth phase (OD_{600} 0.5). Transduced *M. tuberculosis* H37Rv were spread on Mtb minimal plate and Mtb YM rich plate and incubated at 37 °C for 2-3 weeks. To generate a transposon library in saturated size, at least 150,000 colonies were collected from each plate. Each transposon library was aliquoted and stored at -80 °C. Each transposon library was generated in duplicates.

Tn-seq. Genomic DNA (gDNA) was prepared from each sample as previously described [41]. gDNA was then fragmented using an S220 Acoustic DNA Shearing Device (Covaris). After shearing, adapters were added using an Illumina TruSeq Nano DNA library prep kit according to the manufacturer's instructions. Transposon junctions were amplified by using a transposon specific primer Mariner_1R_TnSeq_noMm (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG<u>CCGGGGACTTATCAGCCAACC</u>) and p7 primers (CAAGCAGAAGACGGCATACGAG) with HotStarTaq Master Mix Kit (Qiagen) with the following PCR condition (94 °C for 3 min, 30 cycles of 94 °C for 30 seconds, 65 °C for 30 seconds, and 72 °C for 60 seconds, 72 °C for 10 minutes). The transposon junction enriched sample was diluted 1:50 with water and then amplified to add the flow cell adapter and i5 index to the enriched Tn-containing fragments using the following primers:

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

p7 primer: CAAGCAGAAGACGGCATACGAG

The amplification reaction was as follows: 5 µl template DNA (from PCR 1), 1 µl nuclease-free water, 2 µl 5x KAPA HiFi buffer (Kapa Biosystems), 0.3 µl 10 mM dNTPs (Kapa Biosystems), 0.5 µl DMSO (Fisher Scientific), 0.2 µl KAPA HiFi Polymerase (Kapa Biosystems, 0.5 µl i5 indexing primer (10 µM), 0.5 µl p7 primer (10 µM). Cycling conditions were: 95°C for 5 minutes, followed by 10 cycles of 98°C for 20 seconds, 63°C for 15 seconds, 72°C for 1 minute, followed by a final extension at 72°C for 10 minutes. Amplification products were purified with AMPure XP beads (Beckman Coulter) and the uniquely indexed libraries were quantified using a Quant-IT PicoGreen dsDNA assay (Thermo Fisher Scientific) and the resulting fragment size distribution was assessed using a Bioanalyzer (Agilent Technologies). The resultant Tn-seq library was sequenced using a HiSeq 2500 HO, 125 bp PE run using v4 chemistry (Illumina).

Tn-seq analysis. Sequence reads were trimmed using CutAdapt [42]. We first trimmed sequence reads for transposon sequences (CCGGGGACTTATCAGCCAACCTGT) at the 5` ends. Reads did not contain transposon sequence at the 5` end were discarded. After the 5` trimming process, all the sequence reads begin with TA. We then trimmed sequence reads for adaptor sequences ligated to 3`end (GATCCCACTAGTGTCGACACCAGTCTC). After the trimming, we discarded the sequence reads that were shorter than 18 bp. The default error rate of 0.1 was used for all for all trimming processes.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The trimmed sequence reads were mapped (allowing 1 bp mismatch) to the *M. tuberculosis* H37Rv genome (GenBank: AL123456.3) using Bowtie 2 [43]. The number of reads at each TA site were counted and converted to the .wig format, the input file format for TRANSIT [14] using a custom Python script (S1 Appendix). Subsequent statistical analysis for gene essentiality were performed using TRANSIT [14].

Flux Balance Analysis (FBA). FBA solutions were obtained using a simulated environment designed to mimic the Mtb YM rich media designed for this study. This was done by altering uptake boundaries to match the concentration of each metabolite. Most metabolites present in the media were given unlimited boundaries, because these metabolites were not expected to be limiting, and also because they were present at an undefined concentration in the Mtb YM rich medium, due to their source being casamino acids. Metabolites added to the Mtb YM rich medium in known concentrations were bounded in FBA at those concentrations.

The iSM810 model contains 938 metabolic reactions and 810 genes [35]. The biomass reaction originally described for iSM810 was chosen to define growth. FBA was performed using the COBRA Toolbox Matlab package [44, 45]. The unconstrained uptake fluxes were set to ‘1’. Gene essentiality was assessed using the COBRA Toolbox single gene deletion function in Matlab. Through single gene deletion, reactions associated with each gene were systematically closed and the model was optimized for biomass production. Any biomass accumulation > 1e-10 (which could occur due to numerical errors) was defined as growth and the gene was classified as non-essential. Biomass accumulation <1e-10 resulted in a gene being called essential. All FBA optimizations were done using the Gurobi Optimizer 7.0 software under the free academic license [46].

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Comparative genomic analysis.

Following comparative genomic analysis was carried out as previously [47]. In brief, the pan and core genomes were defined using Roary software [48]. Complete and draft genome sequences of pathogenic strains (non-highlighted strains were used and obtained from PATRIC database (1st October, 2017) summarized in Table S6 were reannotated to generate gff3 files using PROKKA ver. 1.1.12 software [49] and to include annotation of a reference strain H37Rv. Homologous proteins (i.e. protein families) were clustered using the CD-Hit and MCL algorithms. The BLASTp cut-off value was set at 95%. The number of core- and pan-genome protein families were estimated via genome sampling up to the number of input genomes at the default setting in Roary.

Acknowledgments

We thank Nicholas D. Peterson and Dr. Igor Libourel for their assistance in initial conception and design of the study. This study was supported by funds from the Minnesota Partnership for Biotechnology and Medical Genomics (ML2012, chapter 5, article 1, section 5, subdivision 5e to A.D.B), the American Lung Association (to A.D.B), the National Institutes of Health (GM121498, to W.R.H. and AI123146 to A.D.B.) the Japan Agency for Medical Research (AMED; project number 17fk0108116h0401to FM), Kakenhi (18K19674 and 16H05501 to F.M).

References

1. World Health Organization. Global Tuberculosis Report. 2017.
2. Zumla A, Nahid P, Cole ST. Advances in the development of new tuberculosis drugs and treatment regimens. Nat Rev Drug Discov. 2013;12(5):388-404. doi: 10.1038/nrd4001. PubMed PMID: 23629506.
3. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. 2013;11(7):435-42. doi: 10.1038/nrmicro3033. PubMed PMID: 23712350; PubMed Central PMCID: PMCPMC3842022.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4. Zhang YJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, et al. Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. Cell. 2013;155(6):1296-308. doi: 10.1016/j.cell.2013.10.045. PubMed PMID: 24315099; PubMed Central PMCID: PMCPMC3902092.
5. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. Gene fitness landscapes of Vibrio cholerae at important stages of its life cycle. PLoS Pathog. 2013;9(12):e1003800. doi: 10.1371/journal.ppat.1003800. PubMed PMID: 24385900; PubMed Central PMCID: PMCPMC3873450.
6. van Opijnen T, Camilli A. A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res. 2012;22(12):2541-51. doi: 10.1101/gr.137430.112. PubMed PMID: 22826510; PubMed Central PMCID: PMCPMC3514683.
7. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, et al. Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. PLoS Pathog. 2012;8(9):e1002946. doi: 10.1371/journal.ppat.1002946. PubMed PMID: 23028335; PubMed Central PMCID: PMCPMC3460630.
8. Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 2011;7(9):e1002251. doi: 10.1371/journal.ppat.1002251. PubMed PMID: 21980284; PubMed Central PMCID: PMCPMC3182942.
9. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, et al. Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. MBio. 2017;8(1). Epub 2017/01/17. doi: 10.1128/mBio.02133-16. PubMed PMID: 28096490; PubMed Central PMCID: PMCPMC5241402.
10. Carey AF, Rock JM, Krieger IV, Chase MR, Fernandez-Suarez M, Gagneux S, et al. TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities. PLoS Pathog. 2018;14(3):e1006939. Epub 2018/03/05. doi: 10.1371/journal.ppat.1006939. PubMed PMID: 29505613; PubMed Central PMCID: PMCPMC5854444.
11. Griffin JE, Pandey AK, Gilmore SA, Mizrahi V, McKinney JD, Bertozzi CR, et al. Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. Chem Biol. 2012;19(2):218-27. doi: 10.1016/j.chembiol.2011.12.016. PubMed PMID: 22365605; PubMed Central PMCID: PMCPMC3292763.
12. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A. 2001;98(22):12712-7. doi: 10.1073/pnas.231275498. PubMed PMID: 11606763; PubMed Central PMCID: PMCPMC60119.
13. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol. 2003;48(1):77-84. PubMed PMID: 12657046.
14. DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR. TRANSIT--A Software Tool for Himar1 TnSeq Analysis. PLoS Comput Biol. 2015;11(10):e1004401. Epub 2015/10/08. doi: 10.1371/journal.pcbi.1004401. PubMed PMID: 26447887; PubMed Central PMCID: PMCPMC4598096.
15. DeJesus MA, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, Ioerger TR. Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. Bioinformatics. 2013;29(6):695-703. Epub 2013/01/29. doi: 10.1093/bioinformatics/btt043. PubMed PMID: 23361328; PubMed Central PMCID: PMCPMC3597147.
16. Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, et al. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2003;100(12):7213-8. Epub 2003/05/29. doi: 10.1073/pnas.1231432100. PubMed PMID: 12775759; PubMed Central PMCID: PMCPMC165855.
17. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList--10 years after. Tuberculosis (Edinb). 2011;91(1):1-7. Epub 2010/10/25. doi: 10.1016/j.tube.2010.09.008. PubMed PMID: 20980199.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

18. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2016;44(D1):D471-80. Epub 2015/11/02. doi: 10.1093/nar/gkv1164. PubMed PMID: 26527732; PubMed Central PMCID: PMCPMC4702838.

19. Chen WH, Minguez P, Lercher MJ, Bork P. OGEE: an online gene essentiality database. Nucleic Acids Res. 2012;40(Database issue):D901-6. Epub 2011/11/10. doi: 10.1093/nar/gkr986. PubMed PMID: 22075992; PubMed Central PMCID: PMCPMC3245054.

20. Sambandamurthy VK, Wang X, Chen B, Russell RG, Derrick S, Collins FM, et al. A pantothenate auxotroph of Mycobacterium tuberculosis is highly attenuated and protects mice against tuberculosis. Nat Med. 2002;8(10):1171-4. doi: 10.1038/nm765. PubMed PMID: 12219086.

21. Pandey AK, Sassetti CM. Mycobacterial persistence requires the utilization of host cholesterol. Proc Natl Acad Sci U S A. 2008;105(11):4376-80. Epub 2008/03/11. doi: 10.1073/pnas.0711159105. PubMed PMID: 18334639; PubMed Central PMCID: PMCPMC2393810.

22. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, et al. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic Acids Res. 2017;45(D1):D543-D50. Epub 2016/11/28. doi: 10.1093/nar/gkw1003. PubMed PMID: 27899573; PubMed Central PMCID: PMCPMC5210515.

23. D'Elia MA, Pereira MP, Brown ED. Are essential genes really essential? Trends Microbiol. 2009;17(10):433-8. Epub 2009/09/16. doi: 10.1016/j.tim.2009.08.005. PubMed PMID: 19765999.

24. Kieser KJ, Boutte CC, Kester JC, Baer CE, Barczak AK, Meniche X, et al. Phosphorylation of the Peptidoglycan Synthase PonA1 Governs the Rate of Polar Elongation in Mycobacteria. PLoS Pathog. 2015;11(6):e1005010. Epub 2015/06/26. doi: 10.1371/journal.ppat.1005010. PubMed PMID: 26114871; PubMed Central PMCID: PMCPMC4483258.

25. Vandal OH, Roberts JA, Odaira T, Schnappinger D, Nathan CF, Ehrt S. Acid-susceptible mutants of Mycobacterium tuberculosis share hypersusceptibility to cell wall and oxidative stress and to the host environment. J Bacteriol. 2009;191(2):625-31. Epub 2008/11/14. doi: 10.1128/JB.00932-08. PubMed PMID: 19011036; PubMed Central PMCID: PMCPMC2620805.

26. Kieser KJ, Baranowski C, Chao MC, Long JE, Sassetti CM, Waldor MK, et al. Peptidoglycan synthesis in Mycobacterium tuberculosis is organized into networks with varying drug susceptibility. Proc Natl Acad Sci U S A. 2015;112(42):13087-92. Epub 2015/10/05. doi: 10.1073/pnas.1514135112. PubMed PMID: 26438867; PubMed Central PMCID: PMCPMC4620856.

27. Warner DF, Savvi S, Mizrahi V, Dawes SS. A riboswitch regulates expression of the coenzyme B12-independent methionine synthase in Mycobacterium tuberculosis: implications for differential methionine synthase function in strains H37Rv and CDC1551. J Bacteriol. 2007;189(9):3655-9. doi: 10.1128/JB.00040-07. PubMed PMID: 17307844; PubMed Central PMCID: PMCPMC1855906.

28. Gordhan BG, Smith DA, Alderton H, McAdam RA, Bancroft GJ, Mizrahi V. Construction and phenotypic characterization of an auxotrophic mutant of Mycobacterium tuberculosis defective in L-arginine biosynthesis. Infect Immun. 2002;70(6):3080-4. PubMed PMID: 12011001; PubMed Central PMCID: PMCPMC127984.

29. Movahedzadeh F, Smith DA, Norman RA, Dinadayala P, Murray-Rust J, Russell DG, et al. The Mycobacterium tuberculosis ino1 gene is essential for growth and virulence. Mol Microbiol. 2004;51(4):1003-14. PubMed PMID: 14763976.

30. Pavelka MS, Jacobs WR. Comparison of the construction of unmarked deletion mutations in Mycobacterium smegmatis, Mycobacterium bovis bacillus Calmette-Guérin, and Mycobacterium tuberculosis H37Rv by allelic exchange. J Bacteriol. 1999;181(16):4780-9. PubMed PMID: 10438745; PubMed Central PMCID: PMCPMC93962.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

31. Murphy HN, Stewart GR, Mischenko VV, Apt AS, Harris R, McAlister MS, et al. The OtsAB pathway is essential for trehalose biosynthesis in Mycobacterium tuberculosis. J Biol Chem. 2005;280(15):14524-9. Epub 2005/02/09. doi: 10.1074/jbc.M414232200. PubMed PMID: 15703182.

32. Korte J, Alber M, Trujillo CM, Syson K, Koliwer-Brandl H, Deenen R, et al. Trehalose-6-Phosphate-Mediated Toxicity Determines Essentiality of OtsB2 in Mycobacterium tuberculosis In Vitro and in Mice. PLoS Pathog. 2016;12(12):e1006043. Epub 2016/12/09. doi: 10.1371/journal.ppat.1006043. PubMed PMID: 27936238; PubMed Central PMCID: PMCPMC5148154.

33. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017;45(D1):D535-D42. Epub 2016/11/29. doi: 10.1093/nar/gkw1017. PubMed PMID: 27899627; PubMed Central PMCID: PMCPMC5210524.

34. O'Brien EJ, Monk JM, Palsson BO. Using Genome-scale Models to Predict Biological Capabilities. Cell. 2015;161(5):971-87. doi: 10.1016/j.cell.2015.05.019. PubMed PMID: 26000478; PubMed Central PMCID: PMCPMC4451052.

35. Ma S, Minch KJ, Rustad TR, Hobbs S, Zhou SL, Sherman DR, et al. Integrated Modeling of Gene Regulatory and Metabolic Networks in Mycobacterium tuberculosis. PLoS Comput Biol. 2015;11(11):e1004543. Epub 2015/11/30. doi: 10.1371/journal.pcbi.1004543. PubMed PMID: 26618656; PubMed Central PMCID: PMCPMC4664399.

36. Orth JD, Thiele I, Palsson B. What is flux balance analysis? Nat Biotechnol. 2010;28(3):245-8. doi: 10.1038/nbt.1614. PubMed PMID: 20212490; PubMed Central PMCID: PMCPMC3108565.

37. Pelicic V, Jackson M, Reyrat JM, Jacobs WR, Gicquel B, Guilhot C. Efficient allelic exchange and transposon mutagenesis in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 1997;94(20):10955-60. PubMed PMID: 9380741; PubMed Central PMCID: PMCPMC23543.

38. Jackson M, Phalen SW, Lagranderie M, Ensergueix D, Chavarot P, Marchal G, et al. Persistence and protective efficacy of a Mycobacterium tuberculosis auxotroph vaccine. Infect Immun. 1999;67(6):2867-73. PubMed PMID: 10338493; PubMed Central PMCID: PMCPMC96594.

39. Kriakov J, Lee S, Jacobs WR. Identification of a regulated alkaline phosphatase, a cell surface-associated lipoprotein, in Mycobacterium smegmatis. J Bacteriol. 2003;185(16):4983-91. PubMed PMID: 12897018; PubMed Central PMCID: PMCPMC166462.

40. Rubin EJ, Akerley BJ, Novik VN, Lampe DJ, Husson RN, Mekalanos JJ. In vivo transposition of mariner-based elements in enteric bacteria and mycobacteria. Proc Natl Acad Sci U S A. 1999;96(4):1645-50. PubMed PMID: 9990078; PubMed Central PMCID: PMCPMC15546.

41. Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WR. Genetic Manipulation of Mycobacterium tuberculosis. Curr Protoc Microbiol. 2007;Chapter 10:Unit 10A.2. doi: 10.1002/9780471729259.mc10a02s6. PubMed PMID: 18770603.

42. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal; Vol 17, No 1: Next Generation Sequencing Data Analysis. 2011.

43. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25. doi: 10.1186/gb-2009-10-3-r25. PubMed PMID: 19261174; PubMed Central PMCID: PMCPMC2690996.

44. Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc. 2007;2(3):727-38. doi: 10.1038/nprot.2007.99. PubMed PMID: 17406635.

45. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6(9):1290-307. Epub 2011/08/04. doi: 10.1038/nprot.2011.308. PubMed PMID: 21886097; PubMed Central PMCID: PMCPMC3319681.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

46. Gurobi Optimization, Inc. Gurobi Optimizer Reference Manual. Available at: http://www.gurobi.com

47. Yano H, Iwamoto T, Nishiuchi Y, Nakajima C, Starkova DA, Mokrousov I, et al. Population Structure and Local Adaptation of MAC Lung Disease Agent Mycobacterium avium subsp. hominissuis. Genome Biol Evol. 2017;9(9):2403-17. doi: 10.1093/gbe/evx183. PubMed PMID: 28957464; PubMed Central PMCID: PMCPMC5622343.

48. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3. Epub 2015/07/20. doi: 10.1093/bioinformatics/btv421. PubMed PMID: 26198102; PubMed Central PMCID: PMCPMC4817141.

49. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9. Epub 2014/03/18. doi: 10.1093/bioinformatics/btu153. PubMed PMID: 24642063.

50. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, et al. inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science. 1994;263(5144):227-30. PubMed PMID: 8284673.

Supporting Information Legends

Table S1. Compositions of Mtb YM rich and Mtb minimal media.

Table S2 Results of Tn-seq gene essentiality analysis by TRANSIT Bayesian/Gumbel method.

Table S3 Results of Tn-seq gene essentiality analysis by TRANSIT HMM.

Table S4 Comparison of essential genes between Mtb YM rich and Mtb minimal media identified by Tn-seq and analyzed by TRANSIT re-sampling method.

Table S5 Comparison of essential metabolic pathways between Mtb YM rich and Mtb minimal media.

Table S6 Complete and draft genome sequences of virulent *M. tuberculosis* complex used for comparative genomics analysis.

Table S7 *in silico* growth predictions of a single gene knockout mutant strains of *M. tuberculosis*.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S1. Comparison of conditional essential genes identified by Tn-seq and predicted *in silico*. (A) Comparison of essential genes in Mtb YM rich (red) and Mtb minimal (blue) predicted by iSM810 predicted. (B) Comparison of essential genes in Mtb YM rich (red) and Mtb minimal (blue) identified by Tn-seq.

Data S1 List of transport reactions in iSM810 model.

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Figure 1. Identification of essential genes in Mtb YM rich medium. (A) Media compositions of Mtb minimal and Mtb YM rich. 7H9 media contains glutamic acid in addition to the nutrients shown in bold. DAP, diaminopimelic acid. FAC, ferric ammonium citrate. PABA, *para*-aminobenzoic acid. *, 0.5 % casamic 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.

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.

Figure 3. Identification of highly conserved essential genes among virulent *M. tuberculosis* complex. (A) Strains used for comparative genomics analysis. (B) Results for pan-genome analysis by Roary. (C) Number of highly conserved genes

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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].

bioRxiv preprint doi: https://doi.org/10.1101/534289; this version posted January 29, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1 Gene essentialities and *in silico* essentiality predictions of anti-tubercular drug target genes.

Gene (Rv number)	Drug	Essentiality in 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%)

A

Mtb minimal media

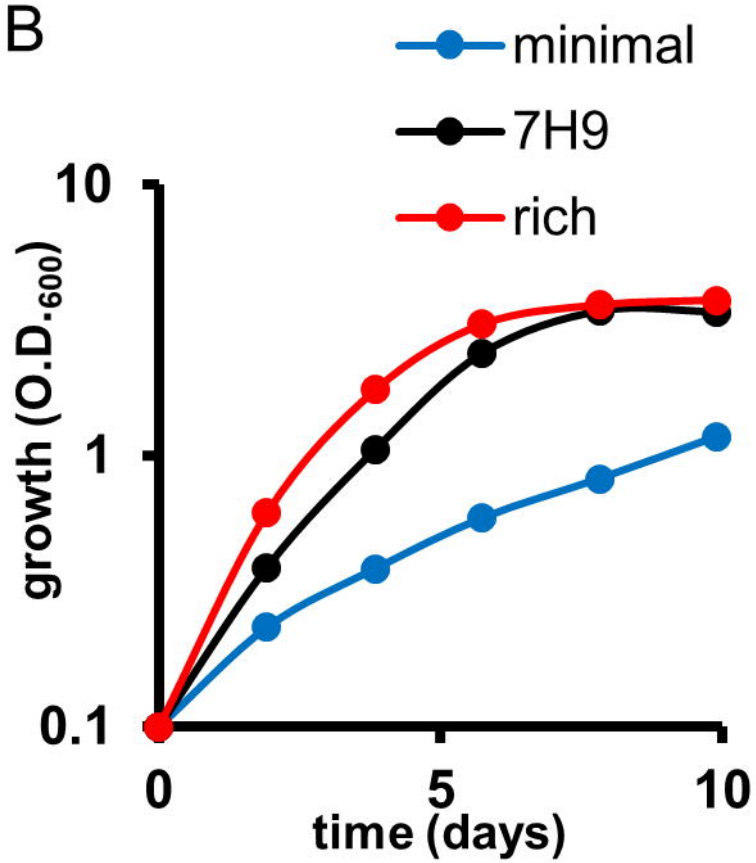
glycerol
asparagine
KH_2PO_4
Na_2HPO_4
FAC
$CaCl_2$
$ZnSO_4$
$MgSO_4$

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
$CuSO_4$

Mtb rich media

B



C

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

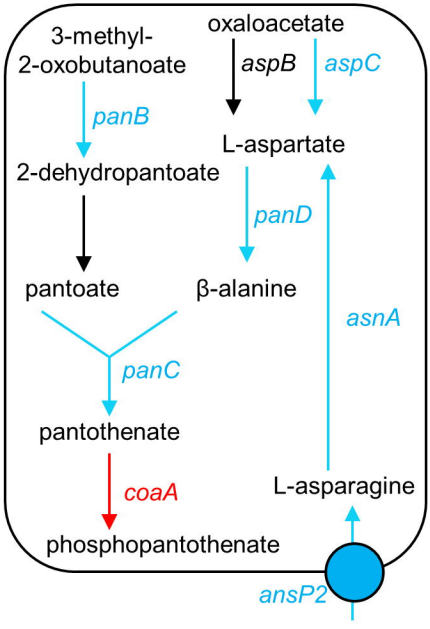
D

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

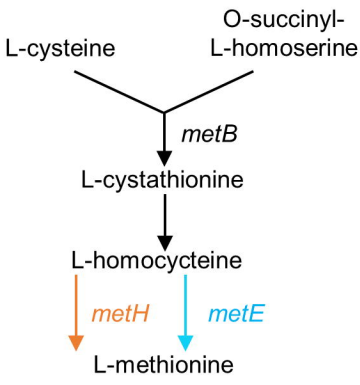
A

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

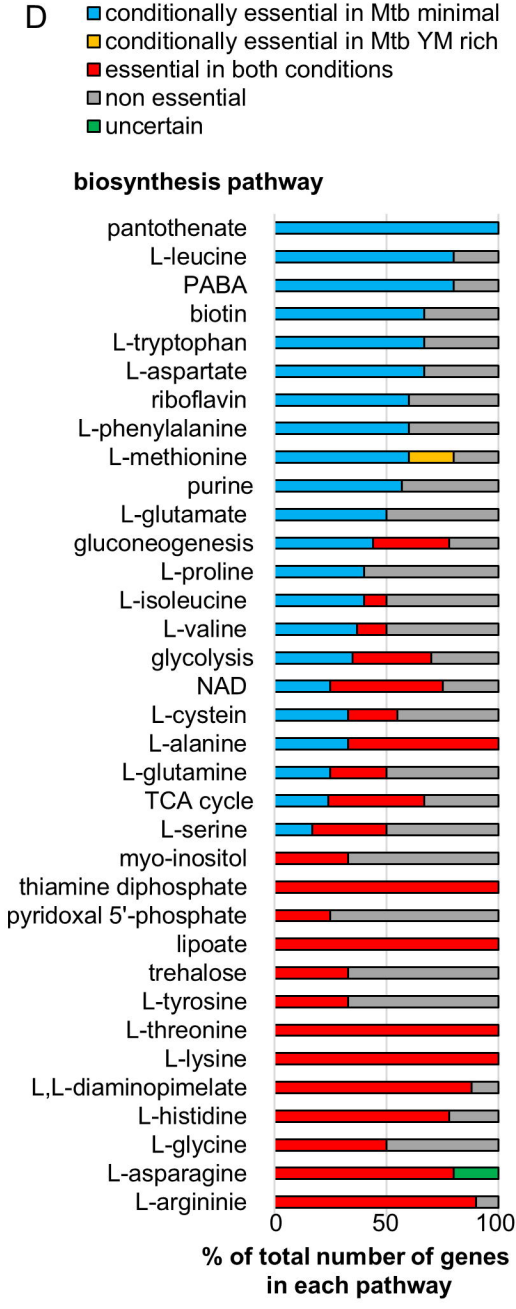
B



C



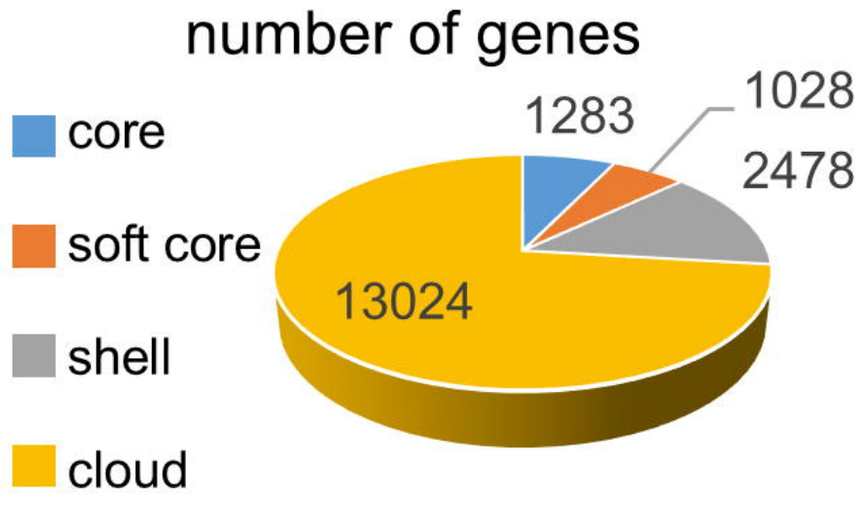
D



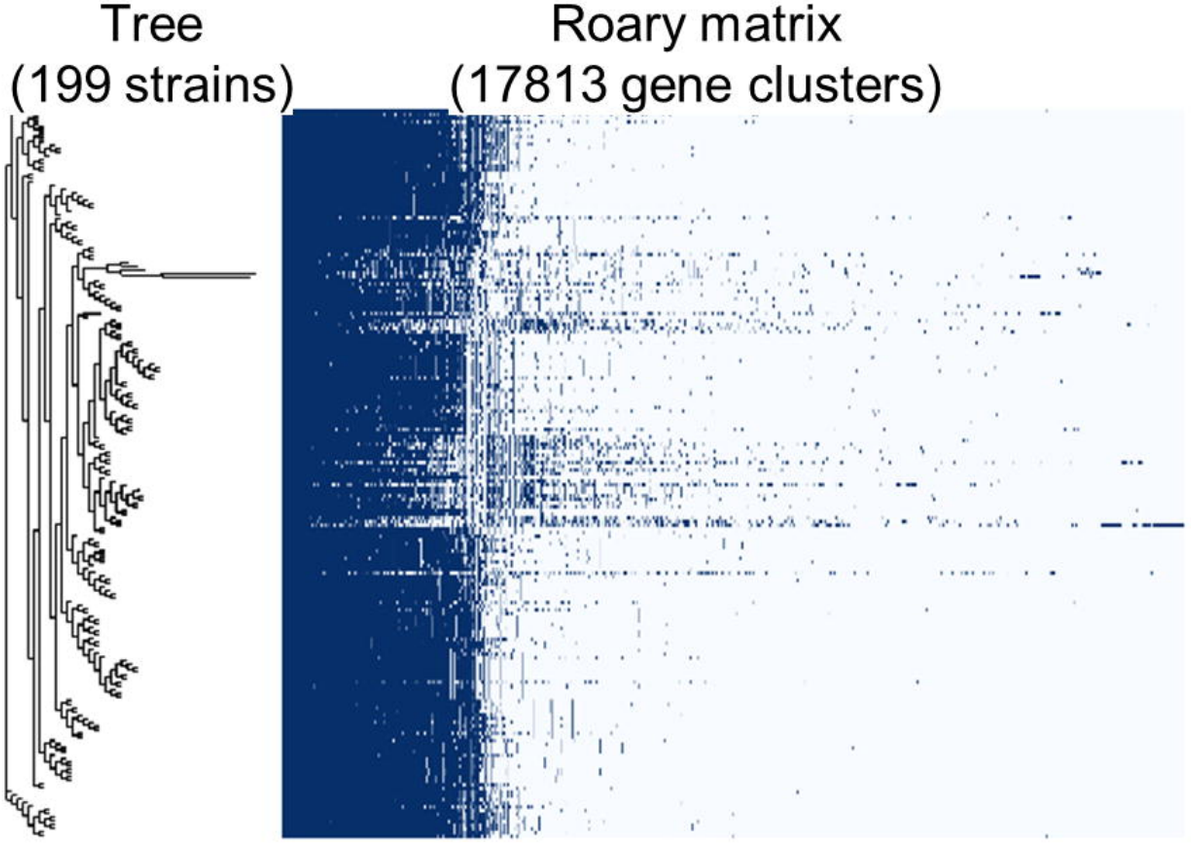
A

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

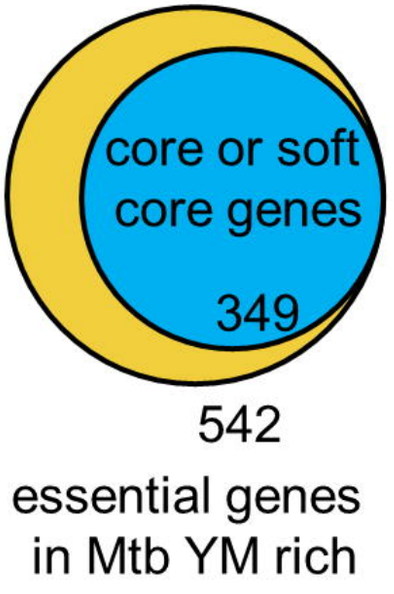
C



B



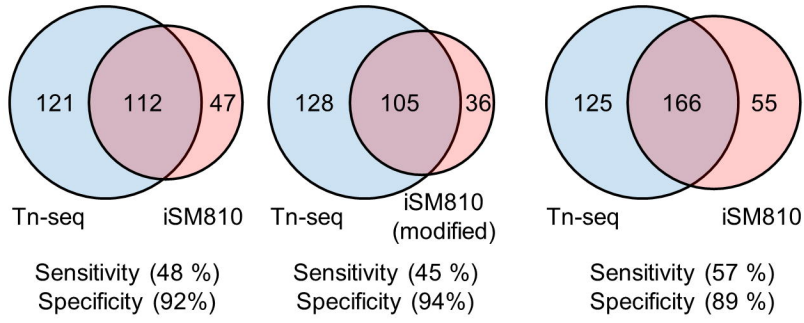
D



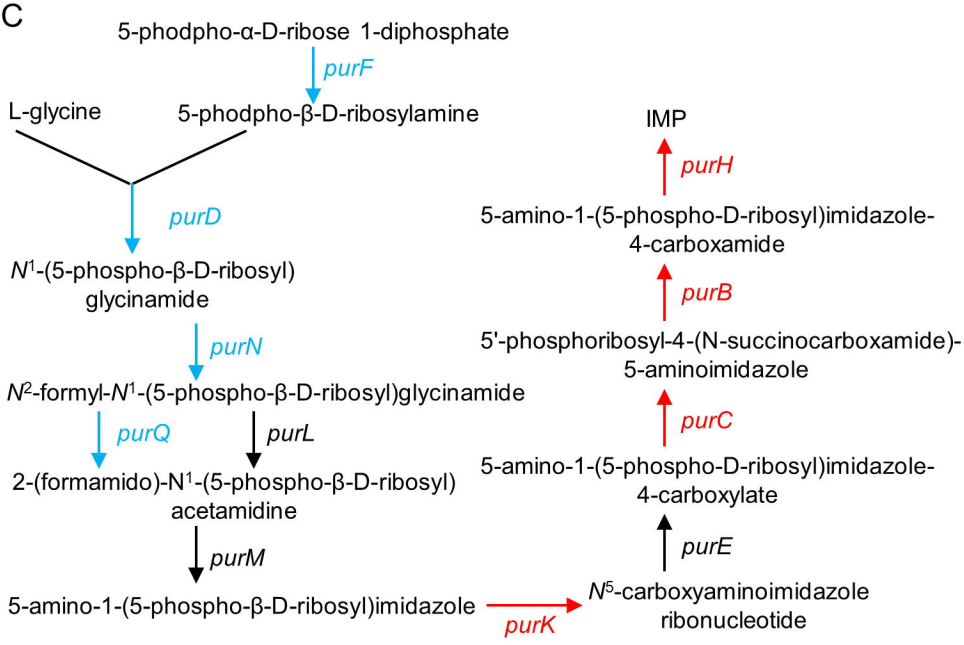
A

	iSM810		iSM810 (modified)	
	Mtb YM rich	Mtb minimal	Mtb YM rich	Mtb Minimal
essential	159	221	141	221
non-essential	651	589	669	589
total	810	810	810	810

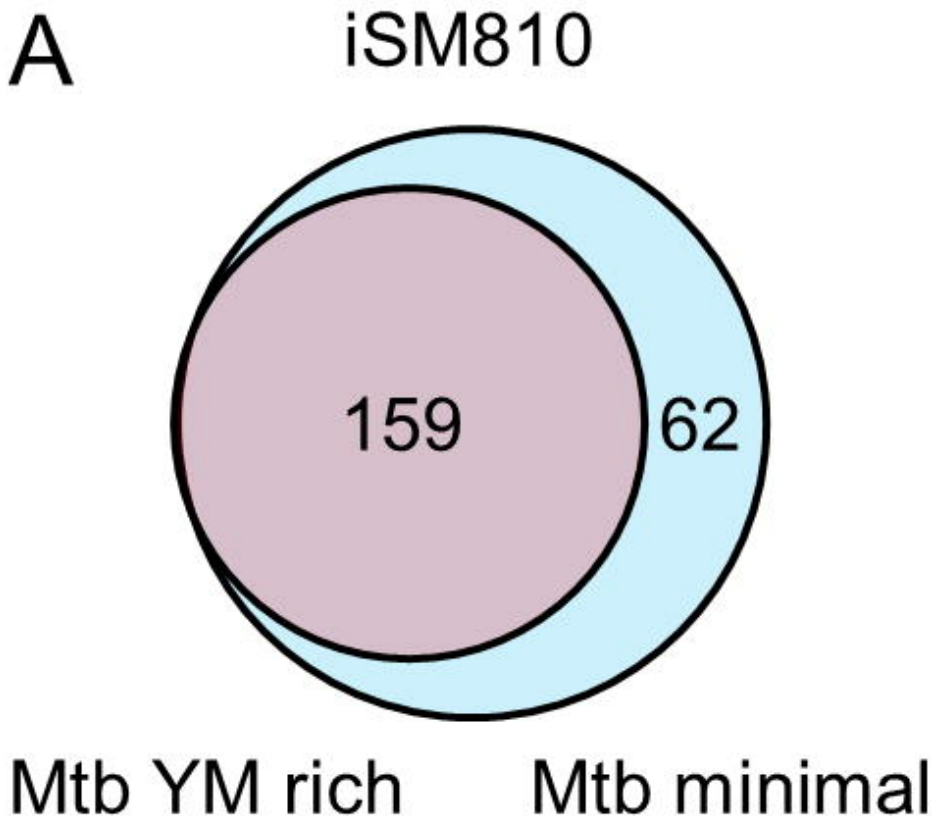
B

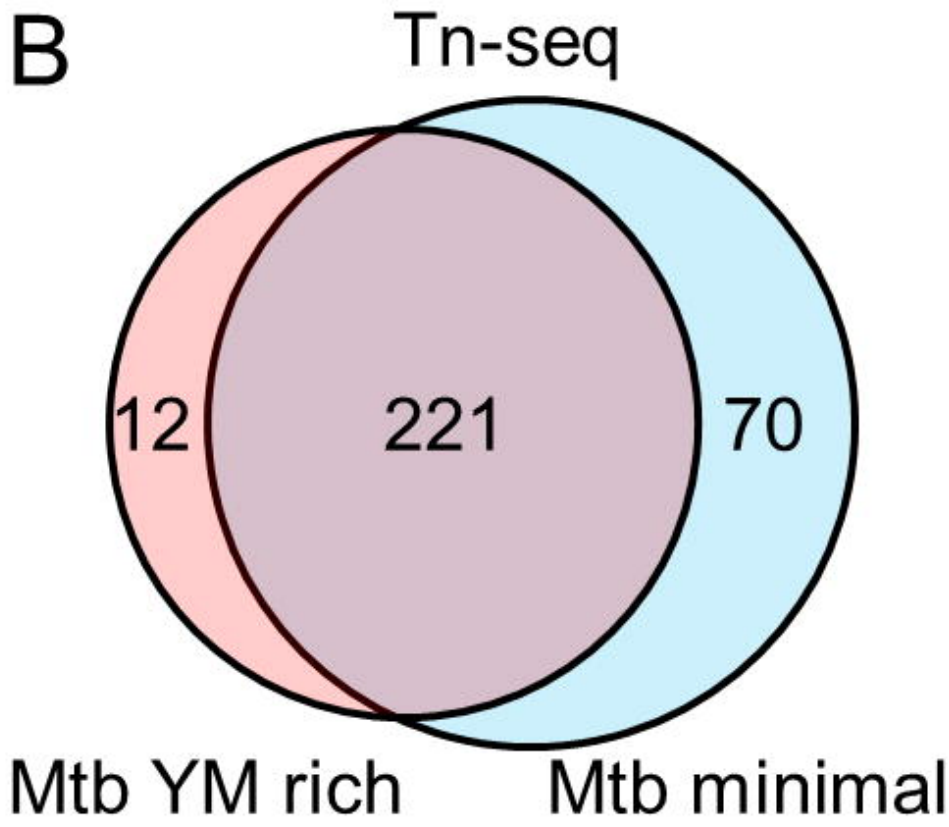


C



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


A
iSM810
159
62
Mtb YM rich
Mtb minimal


B
Tn-seq
12
221
70
Mtb YM rich
Mtb minimal