

Mycobacterial Metabolic Model Development for Drug Target Identification

Bridget P. Bannerman^{1,3}, Alex Oarga², and Jorge Júlvez²

¹University of Cambridge, Cambridge, UK bpc28@cam.ac.uk.

²Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain.

³Science Resources Foundation, UK

ABSTRACT

Antibiotic resistance is increasing at an alarming rate, and three related mycobacteria are the source of widespread infections in humans. According to the World Health Organization, *Mycobacterium leprae*, which causes leprosy, is still endemic in tropical countries; *Mycobacterium tuberculosis* is the second largest infectious disease killer worldwide after COVID-19; *Mycobacteroides abscessus*, a group of non-tuberculous mycobacteria, causes lung infections and other health-care-associated infections in humans. Due to the rise in resistance to common antibacterial drugs, it is critical to develop alternatives to traditional treatment procedures. Furthermore, an understanding of the biochemical mechanisms underlying pathogenic evolution is important for the treatment and management of these diseased conditions.

In this study, metabolic models have been developed for two bacterial pathogens, *M. leprae*, and *M. abscessus*, and a new computational tool has been used to identify potential drug targets, which are referred to as bottleneck reactions. The genes, reactions, and pathways in each of these organisms have been highlighted; the potential drug targets can be further explored as broad-spectrum antibacterials and the unique drug targets to each pathogen are significant for precision medicine initiatives.

The models and associating datasets are available in GigaScience and the following repositories:

- *M. abscessus*
 - Biomodels: <https://www.ebi.ac.uk/biomodels/MODEL2203300002>

- <https://www.patmedb.org/Bacteria/Mabscessus>
-
- *M. leprae*
 - Biomodels: <https://www.ebi.ac.uk/biomodels/MODEL2203300001>
 - <https://www.patmedb.org/Bacteria/Mleprae>

KEYWORDS

M. abscessus, *M. leprae*, *M. tuberculosis*, drug targets, bottleneck reactions

DATA DESCRIPTION

Context

M. leprae and *M. tuberculosis* are two related pathogenic mycobacteria responsible for leprosy and tuberculosis in humans. Another related mycobacterium, *M. abscessus*, causes opportunistic infections in healthcare-related settings.

The previous analyses of the metabolic models of *M. tuberculosis* have supported studies demonstrating the evolutionary drivers of antibiotic resistance and the identification of novel drug targets against mycobacteria [Kavvas *et al.*, 2018, Bannerman *et al.*, 2019].

Here we demonstrate the newly built genome-scale metabolic models of *M. leprae* and *M. abscessus*, including the curation, simulation, and model optimization strategies [Ebrahim *et al.*, 2013, Oarga *et al.*, 2020, Bannerman *et al.*, 2021]. To ensure the development of standardised metabolic models for the global systems biology community, we have implemented the recently released community standards and used the MEMOTE quality control software to evaluate our models [Carey *et al.*, 2020, Lieven *et al.*, 2020].

METHODS

GEM reconstruction, curation, and simulation

Automated draft reconstructions of *M. leprae* and *M. abscessus* were downloaded from BioModels and evaluated them against other organism-specific databases, such as BioCyc and KEGG [Caspi, 2014; Kanehisa, 2021]. COBRAPy, a Python toolbox for the construction,

manipulation, and analysis of constraint-based models [Ebrahim *et al.*, 2013] and GNU Linear Programming Kit (GLPK), a software package which solves efficiently large-scale linear programming problems (<https://www.gnu.org/software/glpk/>) were used to manipulate and simulate the models.

To create the new genome-scale metabolic models (GEM) of *M. abscessus* and *M. leprae*, additional reactions, gene-to-reaction associations, and pathways (that were not in the automated model) were integrated from Kegg and BioCyc [Caspi, 2014; Kanehisa, 2021]. The annotation of genes and metabolites were improved by comparing and transferring annotations from the related *M. tuberculosis* models in iEKVIII model and BioCyc database [Caspi, 2014; Kavvas *et al.*, 2018]. Further improvements to the models were made by comparing them with the compound formula and charge from the Metanetx database and by mapping the genes and reactions of the GEMS to the BiGG, ChEBI, KEGG, and MetaCyc, databases [Hastings *et al.*, 2016; Norsigian *et al.*, 2020; Moretti *et al.*, 2021]. Figure 1 describes the full process of the mycobacterial metabolic models (*Mab22* and *Mlep22*) of the pathogens, *M. abscessus* and *M. leprae*. The revised model reconstructions of *M. leprae* (*Mlep22*) and *M. abscessus* (*Mab22*) can be instantiated without error on the COBRA software (version 0.16.0) [Ebrahim *et al.*, 2013].

BIOMASS reactions for *M. leprae* and *M. abscessus*

We generated biomass reactions for *M. abscessus* and *M. leprae* using the methodology in the pathway tools software [Karp *et al.*, 2019] and the corresponding BioCyc databases [Karp *et al.*, 2019]. The software tool, *findCPcli*, which implements the computational method for identifying bottleneck reactions as drug targets are available at <https://github.com/findCP/findCPcli>.

DATA VALIDATION AND QUALITY CONTROL

Model optimisation

The dead-end metabolite reactions that were previously present in the automated model were eliminated to enhance the model's quality. The model was then iteratively evaluated, considering model-specific reactions for the *M. abscessus* and *M. leprae* organisms and comparisons to the BioCyc, Kegg, MetaNetX 4.2, BiGG, and ChEBI databases [Caspi *et al.*, 2014; Hastings *et al.*, 2016; Norsigian *et al.*, 2020; Moretti *et al.*, 2021].

To improve the quality of the model, the reactions with dead-end metabolites, previously found in the automated models were removed. MEMOTE, a standardised genome-scale metabolic model testing programme was used to undertake quality control checks during the model's iterations and optimization [Leiven *et al.*, 2020]. In the process, SBO annotations and gene annotations from the KEGG database and 728 new formulae from the MetaNetX database were added [Moretti *et al.*, 2021]. As a result, the Memote score increased from 49% to 63% on the *M. leprae* model and from 48% to 66% on the *M. abscessus* model.

The new GEMS for *M. abscessus* and *M. leprae* are encoded in the Systems Biology Markup Language (SBML) [Keating *et al.*, 2020] and designated as follows:

iMlep22 for the *M. leprae* model: i for in silico, Mlep for *M. leprae*, and published in 2022.

iMlep22 consists of 5625 reactions, 4016 metabolites, 871 genes.

iMab22 for the *M. abscessus* model: i for in silico, Mab for *M. abscessus*, and published in 2022. *iMab22* consists of 8580 reactions, 6273 metabolites, and 1837 genes.

Standardisation and curation have been done according to the community standards for the development of metabolic models as described in Carey *et al.*, 2020 and Lieven *et al.*, 2020 to produce gold standard metabolic network reconstructions of *M. abscessus* and *M. leprae*. The development of *M. abscessus* and *M. leprae* (*iMab22* and *iMlep22*) metabolic models is illustrated in Figure 1, and the overall capability of the models has been summarised in Figure 2.

Comparative analysis

M. abscessus, an opportunistic pathogen, has a large genome size of 5.1 MB (Ripoll *et al.* 2009) with more biochemical and bottleneck reactions in the *iMab22* model compared to the obligate pathogens *M. tuberculosis* with a genome size of 4.4MB and *M. leprae*, with an even smaller genome size of 3.3 bp [Singh & Cole, 2011]. The fewer metabolic and bottleneck reactions in the *iMlep22* model can be attributed to the reductive evolution that occurred in *M. leprae* and the subsequent loss of genes. An illustration of the distribution of unique enzymes and bottleneck reactions in each of the models (*iMab22* and *iMlep22*) in comparison with each other and with *M. tuberculosis* is demonstrated in Figure 3. Alternative enzymes are not included in this analysis because they catalyse the same biochemical reactions and do not fit into the category of bottleneck reactions, which are defined as unique reactions responsible for the survival and growth of the organisms in the metabolic network [Bannerman *et al.*, 2021; Yeh *et al.*, 2004; Oarga *et al.*, 2020].

Re-use potential

The standardised genomic scale metabolic models for *M. leprae* (*iMlep22*) and *M. abscessus* (*iMab22*) have been developed using the systems biology community standards for quality control and evaluation of the models [Carey *et al.*, 2020, Lieven *et al.*, 2020] and available for reuse by the global scientific community. The models can be retrieved from:

M. abscessus

<https://www.ebi.ac.uk/biomodels/MODEL2203300002> and

<https://www.patmedb.org/Bacteria/Mabscessus>.

M. leprae

<https://www.ebi.ac.uk/biomodels/MODEL2203300001>

<https://www.patmedb.org/Bacteria/Mleprae>

The *findCPcli* tool used for analysing the models can be retrieved from

biotools (<https://bio.tools/findcpcli>) and

SciCrunch (https://scicrunch.org/resources/about/registry/SCR_023391).

References

Kavvas E, Seif Y, Yurkovich J, Norsigian C, Poudel S, Greenwald WW, et al. Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst Biol*. 2018;**12**:25.

Bannerman BP, Vedithi, S, Julvez J, Munir A, Skwark M, Oliver S and Blundell, TL. Comparative pathway analysis of mycobacterial species provides insight into the identification of new drug targets, 2019. doi: <https://doi.org/10.1101/535856>

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COstraints-based reconstruction and analysis for Python. *BMC Syst Biol* 7: 74. 10.1186/1752-0509-7-74

Oarga A, Bannerman BP, Júlvez J. Growth Dependent Computation of Chokepoints in Metabolic Networks. 2020 In book: Lecture Notes in Computer Science. DOI: 10.1007/978-3-030-60327-4_6.

Bannerman, B. P, Jorge Júlvez, Alexandru Oarga, Tom L Blundell, Pablo Moreno, and R Andres Floto. 'Integrated Human/SARS-CoV-2 Metabolic Models Present Novel Treatment Strategies against COVID-19'. *Life Science Alliance* 4, no. 10 (1 October 2021): e202000954. <https://doi.org/10.26508/lsa.202000954>.

Carey MA, Dräger A, Beber ME, Papin JA, Yurkovich JT (2020) Community standards to facilitate development and address challenges in metabolic modeling. *Mol Syst Biol* 16: e9235. 10.15252/msb.20199235

Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, et al. (2020) MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* 38: 272–276. 10.1038/s41587-020-0446-y

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 42: D459–D471. 10.1093/nar/gkt1103

Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M (2021) KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res* 49: D545–D551. 10.1093/nar/gkaa970

Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* 44: D1214–D1219. 10.1093/nar/gkv1031

Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z (2020) BiGG models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res* 48: D402–D406. 10.1093/nar/gkz1054

Moretti S, Tran VDT, Mehl F, Ibberson M, Pagni M (2021) MetaNetX/MNXref: Unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res* 49: D570–D574. 10.1093/nar/gkaa992

Karp, P.D. and Midford, P.E. and Paley, S.M. and Krummenacker, M. and Billington, R. and Kothari, A. and Ong, W.K. and Subhraveti, P. and Keseler, I.M. and Caspi, R.,

Pathway Tools version 23.0: Integrated Software for Pathway/Genome Informatics and Systems Biology, arXiv 1-111 (2019).

Karp, P.D., et al., The BioCyc collection of microbial genomes and metabolic pathways *Briefings in Bioinformatics* (2019).

Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, Bergmann FT, Finney A, Gillespie CS, Helikar T, et al. (2020) SBML level 3: An extensible format for the exchange and reuse of biological models. *Mol Syst Biol* 16: e9110. 10.15252/msb.2019911

Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, et al. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. *PLoS one*. 2009 Jun 19;4(6):e5660–e5660.

Singh P, Cole S. *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. *Future Microbiol*. 2011;6(1):57–71.

Figure 1: Mycobacterial Metabolic Model Development for Drug Target Identification: *M. abscessus* and *M. leprae*

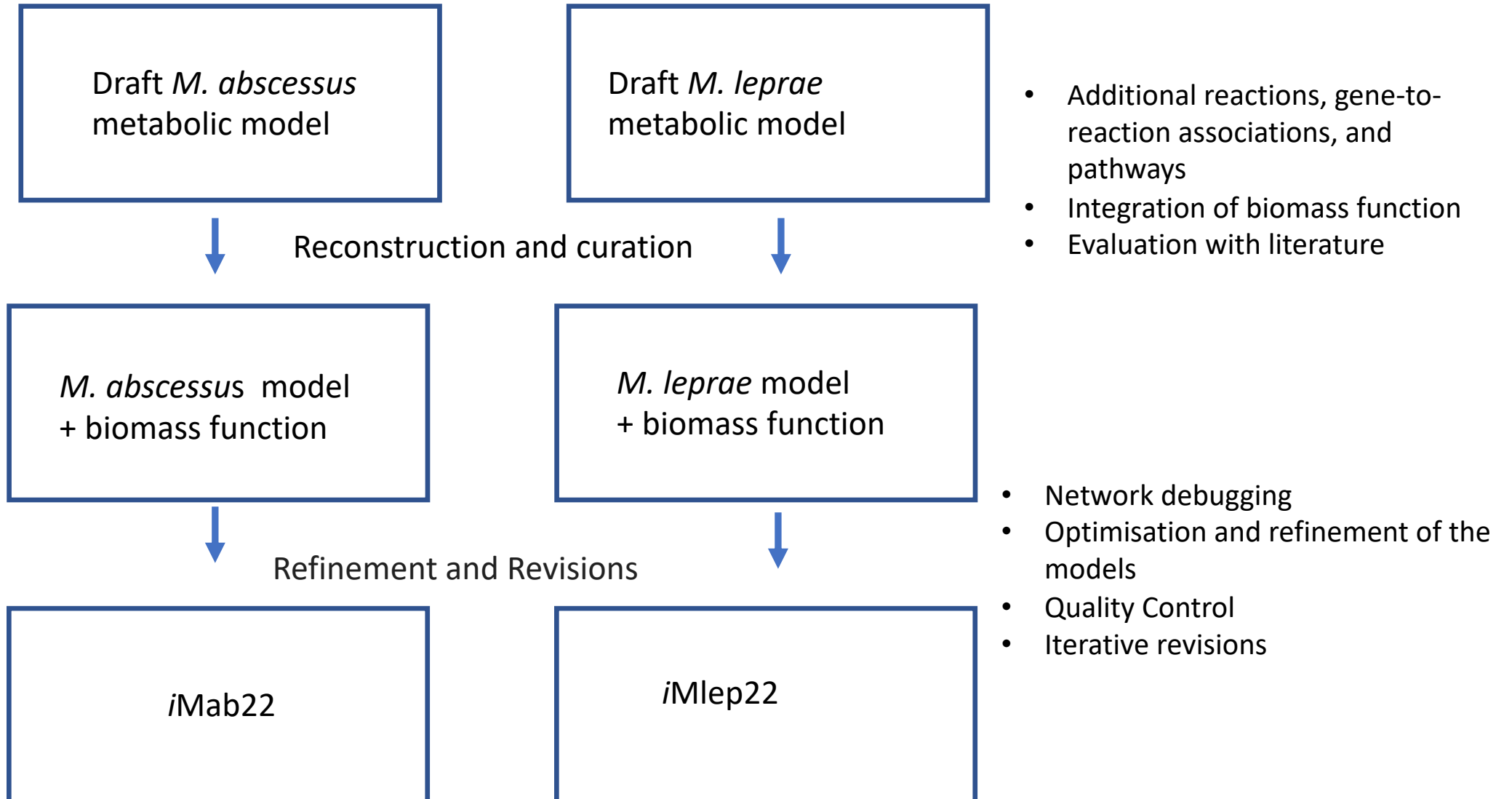


Figure 2: Overall capability of the models : *M. abscessus* (*iMab22*) and *M. leprae* and (*iMlep22*)

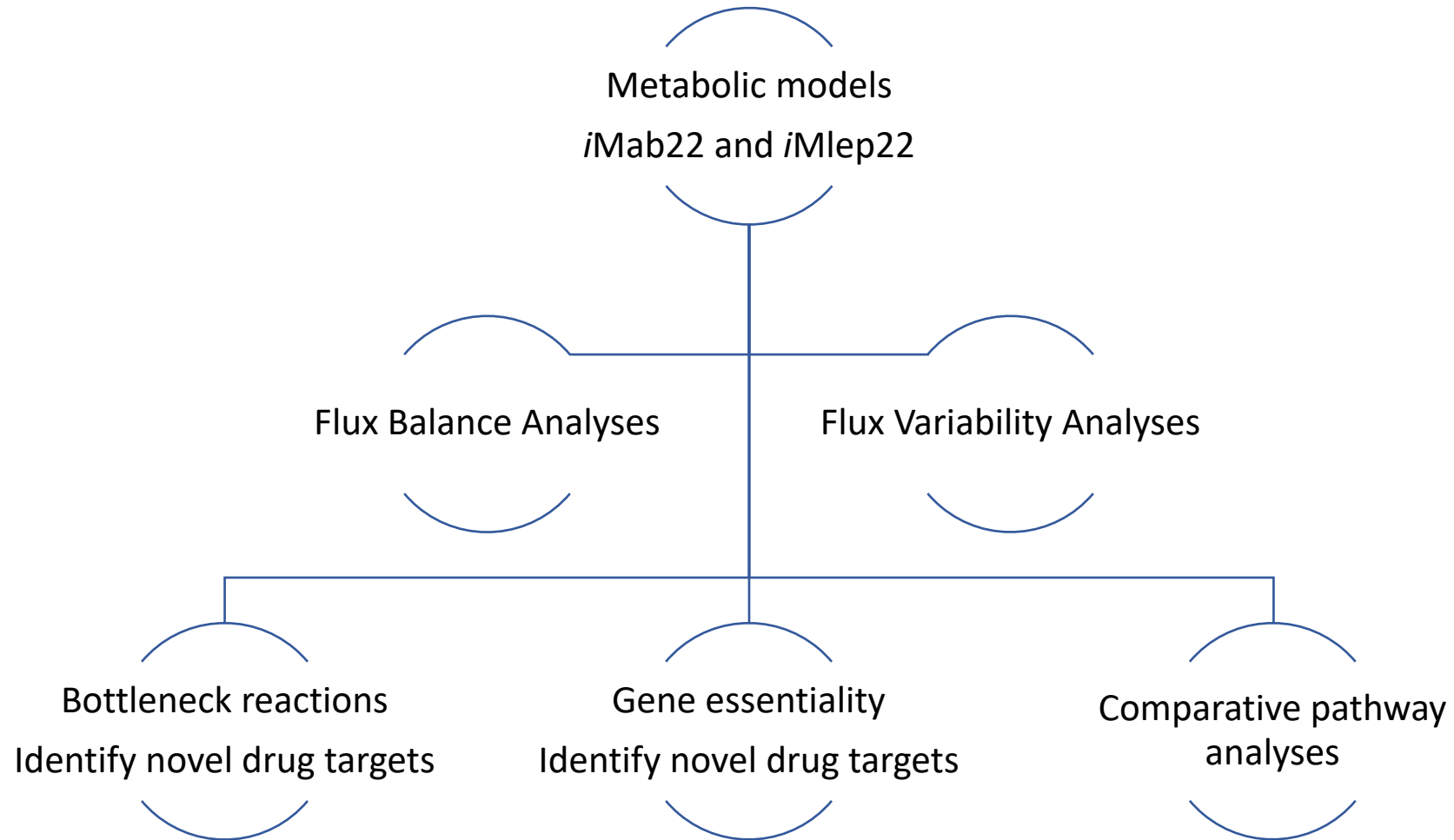


Figure 3: Unique/bottleneck reactions in *M. abscessus*, *M. leprae* in relation to *M. tuberculosis*

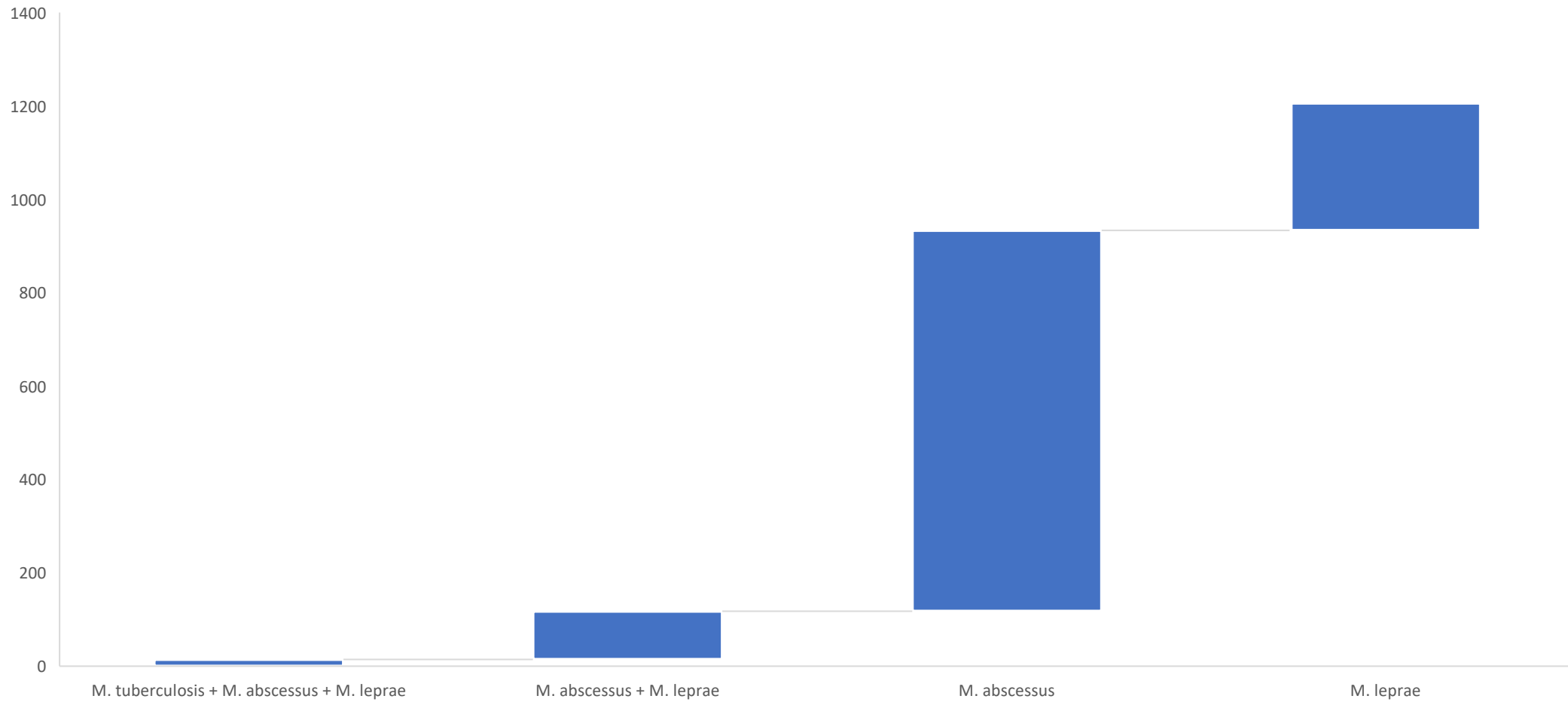


Figure 3b:

1. The number of unique/bottle neck reactions in *M. abscessus* and *M. leprae*
2. The number of bottleneck reactions common to *M. abscessus* and *M. leprae*
3. The number of bottleneck common to *M. abscessus*, *M. leprae* and *M. tuberculosis*