Comparative analysis of metabolic networks of pathogens

to be submitted as part of Major Project Report in partial fulfilment of the requirement for the degree of

Master of Technology

in

Bioinformatics

Submitted by

Kumar Gaurav
(2K14/Bio/09)

Delhi Technological University, Delhi, India

under the supervision of

Dr. Yasha Hasija

Assistant Professor

Department of Biotechnology

Delhi Technological University, Delhi, India

(Formerly Delhi College of Engineering)
Shahbad daulatpur, Bawana road, New Delhi - 110042

CERTIFICATE

This is to certify that the dissertation entitled **"Comparative analysis of metabolic networks of pathogens"** in the partial fulfillment of the requirements for the award of the degree of Masters of Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by him under my guidance. The information and data enclosed in this thesis is original and has not been submitted elsewhere for honoring of any other degree.

Dr. Yasha Hasija
(Project Mentor)
Department of Bio-Technology
Delhi Technological University
Delhi-110042

Dr. D. Kumar
Head of Department
Department of Bio-Technology
Delhi Technological University
Delhi -110042

DECLARATION

I declare that my major project entitled **"Comparative analysis of metabolic networks of pathogens"**, submitted to Department of Biotechnology, Delhi Technological University as a result of the work carried out by me at "Genome Informatics Laboratory", Department of Biotechnology, as Major project.

Date: Kumar Gaurav

ACKNOWLEDGEMENT

I would like to express my sincere thanks to Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Delhi, India for granting permission to carry out my Postgraduate Project.

I would like to take this opportunity to express my gratitude to the Dr. D. Kumar, Professor, HOD, Department of Biotechnology for providing me the opportunity to carry out my project successfully by providing us all the necessary resources and guidance.

I would like to take this opportunity to express my thanks to my project guide Dr. Yasha Hasija, Assistant Professor, Department of Biotechnology for continuously guiding and supporting at each level of the project.

I would like to take this opportunity to express my thanks to all my dear friends for their timely support and motivation.

Kumar Gaurav
2K14/Bio/09

Index

List of Figures and Tables

1. Abstract

Metabolism is an important cellular process and its malfunction is a significant reason behind different human diseases. Metabolic Network is the system of connected chemical reactions. It is the interconnection between the different metabolites and physical processes determining physiological and biochemical properties in the particular metabolic pathway. Connections between biochemical responses through substrate and product metabolites produce complex metabolic networks that may be studied with the help of network theory, stoichiometric analysis, and information on protein structure or function and metabolite properties. Metabolic networks are complex and highly interconnected and hence it needs system level computational approach to identify the genotype - phenotype relationship. Many diseases are caused by failures of metabolic enzymes. These enzymes exist in the perspective of networks well characterized by the static topology of enzyme-metabolite interactions and by the reaction fluxes that are possible at steady state. Flux balance analysis (FBA) is based on the linear programming algorithm. It has developed as a powerful method for the *in silico* analyses of metabolic networks. Here I present the Flux balance analysis of the genomic scale metabolic models of the four microorganisms i.e. *Microbacterium tuberculosis, Staphylococcus aureus, Helicobacter pylori*, and *Salmonella typhimurium*. Using this approach Gene Essentiality and Synthetic Lethality of the metabolic pathway of the given metabolic models has been performed with the help of Fast-SL algorithm. Synthetic lethal genes are the pairs of non-essential genes whose simultaneous deletion limits growth. These attributes helps to find out some of the important factors for the comparative analysis like essential genes and reactions. The study of synthetic lethality plays a crucial role in explaining functional links between genes and gene function predictions. Metabolite connectivity provides a major insight into basic structure of the metabolite networks

2. Introduction

As a complex interactive nonlinear system, much attention has been gained by the study related to the metabolic networks in recent years. Efforts have also been made for the analysis of the structure of whole metabolic networks (Adam M Feist and Bernhard Ø Palsson, 2008). The use of flux balance analysis in order to predict the flux distribution in an entire metabolic network under certain physiological conditions is to analyze and compare the metabolic networks related to different organisms. There is already the availability of genome scale reconstructed models of the different organisms like *H.pylori, M. tuberculosis* etc. These metabolic networks are obtained from different metabolic models databases (Joshua J. Hamilton, Jennifer L. Reed, 2012). Flux Balance Analysis (FBA) is a widely used constraint based approach for studying biochemical metabolic networks, mainly the genomic scale metabolic network that have been reconstructed in the past decade. It uses the concept of linear programming. Linear programming obtains the maximum potential of the objective function that we are looking at, and therefore, when using flux balance analysis, a single solution is obtained for the optimization problem (Karthik Raman and Nagasuma Chandra, 2009). FBA calculates the flow of metabolites taking place in the metabolic network, hence making it possible to predict the growth speed of an organism or the rate of production of a biochemically important metabolite ((Jeremy S Edwards and Bernhard O Palsson, 2002). Essential genes comprise of genes whose specific deletion is lethal under a specific environmental condition. Identification of essential genes in the metabolic network of microorganisms helps to identify potential drug targets and in understanding of minimal requirements for a synthetic cell. However, experimental assessment of essentiality of the coding genes of metabolic network is resource intensive and not viable for all bacterial organisms, in particular if they are infective (Kitiporn Plaimas *et al*, 2009; Costas D. Maranas *et al*, 2015). A constraint-based analysis of reconstructed metabolic networks has proved to be quite useful in various applications such as metabolic engineering prediction of outcomes of gene deletions, drug-target identification and in the elucidation of cellular regulatory networks (Karthik Raman *et al*, 2005). By analogy, Synthetic Lethals (SLs) refer to the pair of non-essential genes whose simultaneous deletion is proved to be lethal. Synthetic gene lethality can arise for different types of reasons. For example, two protein products can be interchanged with respect to an essential function that act in similar pathway or function in two distinct pathways with redundant or complementary essential functions. The study of synthetic lethality plays an important role in

explaining functional links between genes and the prediction of the gene functions (Karthik Raman *et al*, 2015; Patrick F Suthers et al, 2009; Costas D. Maranas *et al*, 2015). Occurrence of synthetic lethality can be seen between genes and small molecules, and it can be used to explain the mechanism of action of drugs. The true potential of synthetic lethality has been widely studied in yeast (Sebastian M.B. Nijman, 2011). Biological systems are prone to mutation, to the variation in environmental conditions and to random variations in the abundance of component molecules of the biological systems. Many physiological and developmental systems are robust to such disturbances and hence immune to any such fluctuations (Joanna Masel and Mark L. Siegal, 2009). Biological systems that have been experimentally proved of being robust to the significant changes in their environmental conditions need mathematical models for robustness analysis of the metabolic network. These mathematical models should themselves be robust. The necessary condition for the model robustness is that the dynamics of the model should be insensitive to small variations in the parameters of the model. The model dynamics may be very much sensitive to simultaneous parameter variations (Kim J *et al*, 2006). Phenotype phase plane analysis can study the optimal utilization of the metabolic network as a function of the constraints. At present metabolic flux maps are typically calculated for the single growth condition, hence gives a limited view of the metabolic phenotype – genotype relation. But phenotypic phase plane analysis maps all the growth conditions characterized by two environmental variables presented in the single plane (Edwards JS *et al*, 2002). Fast-SL is an algorithm which computes combinations of reactions, which when deleted, leads to the elimination of the growth of the organism due to the modification in the metabolic pathway. It achieves this by a combination of limiting the search space and exhaustively iterating through the remaining combinations (Karthik Raman *et al*, 2015). The functional *in silico* model of *Mycobacterium tuberculosis*, *iNJ661*, contains 661 genes, 1025 reactions and 826 metabolites. Genome scale models can be used as hypothesis generating tools and also for analysis and discovery which will expectantly support the rational drug development process (Neema Jamshidi1 and Bernhard Ø Palsson, 2007). The functional *in silico* model of bacterium *Staphylococcus aureus*, *iSB619,* consists of 619 genes by which 640 metabolic reactions are catalysed. The reaction list is the most complete till date for this pathogen (Scott A Becker and Bernhard Ø Palsson, 2005; Matthias Heinemann, Anne Kummel, Reto Ruinatscha, Sven Panke, 2005). The reconstructed network, *iIT341* represents a detailed review of the current literature

about *Helicobacter pylori* as it integrates biochemical and genomic data in a comprehensive framework. In total, it contains 341 metabolic genes, 476 intracellular reactions, 78 exchange reactions, and 485 metabolites (Bernhard Ø. Palsson *et. al*, 2005; Christophe H *et al*, 2002). Salmonella enterica subspecies I serovar Typhimurium is a human pathogen which causes different diseases and its increasing antibiotic resistance poses many public health problems. The functional *in silico* model of *Salmonella typhimurium*, *STM_v1_0*, consists of 1271 genes, 1802 metabolites and 2545 reactions (Thiele I *et al*, 2011; McClelland M *et al*, 2001). Methods of COnstraint Based Reconstruction and Analysis (COBRA) have been successfully applied in the field of microbial metabolic engineering and are being extended to model transcriptional and signaling networks and in the field of public health. The COBRA approach focuses on using physicochemical and biological constraints to analyse the set of feasible phenotypic states of a reconstructed biological network under a given condition (Jan Schellenberger *et. al*, 2011). SBMLToolbox, a toolbox that facilitates importing and exporting models presented in the Systems Biology Markup Language (SBML) format in the MATLAB environment. It also provides functionality that helps an experienced user of either SBML or MATLAB to combine the computing power of MATLAB with the exchangeability and portability of an SBML model. SBMLToolbox supports all levels and versions of SBML (Sarah M. Keating *et al*, 2006). In this work, we describe the similarities and dissimilarities of metabolic networks of *Salmonella typhimurium* (STM_v1_0), *Helicobacter pylori* (iIT341), *Mycobacterium tuberculosis* (iNJ661) and *Staphylococcus aureus* (iSB619) by pairwise comparison with the help of flux balance analysis and gene deletion method.

3. Literature Review

Mycobacterium tuberculosis continues to be a major pathogen in the third world, killing almost 2 million people a year by the most recent estimates. The functional *in silico* bacterium, *iNJ661*, contains 661 genes, 1025 reactions and 826 metabolites can produce many of the complex compounds characteristic to tuberculosis, such as mycolic acids and mycocerosates. Genome-scale models can be used for analysis, discovery, and as hypothesis generating tools, which will hopefully assist the rational drug development process. These models need to be able to assimilate data from large datasets and analyze them (Neema Jamshidi1 and Bernhard Ø Palsson, 2007). *Staphylococcus aureus* is a pathogenic bacterium that has evolved resistance to many antibiotics, representing a significant health care concern. The functional *in silico* bacterium, *iSB619*, consists of 619 genes that catalyze 640 metabolic reactions. The reaction list is the most complete to date for this pathogen (Scott A Becker and Bernhard Ø Palsson, 2005). *Helicobacter pylori* is a human gastric pathogen infecting almost half of the world population. This reconstruction network, *iIT341* GSM/GPR, represents a detailed review of the current literature about *H. pylori* as it integrates biochemical and genomic data in a comprehensive framework. In total, it accounts for 341 metabolic genes, 476 intracellular reactions, 78 exchange reactions, and 485 metabolites (Bernhard Ø. Palsson *et. al*, 2005). Metabolic reconstructions are common denominators in systems biology and represent biochemical, genetic, and genomic (BiGG) knowledge-bases for target organisms by capturing currently available information in a consistent, structured manner Salmonella enterica subspecies I serovar Typhimurium is a human pathogen, causes various diseases and its increasing antibiotic resistance poses a public health problem. The functional *in silico* bacterium, *STM_v1_0*, consists of 1271 genes, 1802 metabolites and 2545 reactions (Bernhard Ø. Palsson *et. al*, 2011). Flux balance analysis (FBA) is a mathematical approach for analyzing the flow of metabolites through a metabolic network. FBA calculates the flow of metabolites through this metabolic network, thus making it possible to predict the growth rate of an organism or the rate of production of a biotechnologically important metabolite. FBA does not require kinetic parameters and can be computed very quickly even for large networks. This makes it well suited to studies that characterize many different perturbations such as different substrates or genetic manipulations (Thiele *et. al*, 2010). Biological systems that have been experimentally verified to be robust to significant changes in

their environments require mathematical models that are themselves robust. A necessary condition for model robustness is that the model dynamics should not be sensitive to small variations in the model's parameters. The concept of robustness has been proposed as a key indicator of validity for models of many types of biological systems (J. Kim *et. al*, 2006). Synthetic lethal sets are sets of reactions or genes where only the simultaneous removal of all reactions or genes in the set inhibits growth of an organism. Fast-SL enables an efficient enumeration of higher order synthetic lethals in metabolic networks, which may help uncover previously unknown genetic interactions and combinatorial drug targets (Karthik Raman *et. al*, 2015). Microbial cells operate under governing constraints that limit their range of possible functions. With the availability of annotated genome sequences, it has become possible to reconstruct genome-scale biochemical reaction networks for microorganisms. A substantial and growing toolbox of computational analysis methods has been developed to study the characteristics and capabilities of microorganisms using a constraint-based reconstruction and analysis (COBRA) approach. This approach provides a biochemically and genetically consistent framework for the generation of hypotheses and the testing of functions of microbial cells (Bernhard Ø. Palsson *et. al*, 2004). COnstraint-Based Reconstruction and Analysis (COBRA) methods have been successfully employed in the field of microbial metabolic engineering and are being extended to modeling transcriptional and signaling networks and the field of public health. COBRA methods have been used to guide metabolic pathway engineering, to model pathogens and host-pathogen interactions and to assess the impact of disease states on human metabolism. The COBRA approach focuses on employing physicochemical, data-driven, and biological constraints to enumerate the set of feasible phenotypic states of a reconstructed biological network in a given condition (Jan Schellenberger *et. al*, 2011). FBA is a constraint-based modeling approach in which the stoichiometry of the underlying biochemical network constrains the solution. FBA assumes that metabolic networks will reach a steady state constrained by the stoichiometry. The stoichiometric constraints lead to an underdetermined system, however, a bounded solution space of all feasible fluxes can be identified (Kenneth J Kauffman *et. al*, 2003). SEED project, initiated in 2003 as an integration of genomic data and analysis tools, now contains > 5,000 complete genomes, a constantly updated set of curated annotations embodied in a large and growing collection of encoded subsystems, a derived set of protein families, and hundreds of genome-scale metabolic models. RAST annotation server, and

provide access to a growing collection of metabolic models that support flux balance analysis. The SEED servers offer open access to regularly updated data, the ability to annotate prokaryotic genomes, the ability to create metabolic reconstructions and detailed models of metabolism, and access to hundreds of existing metabolic models (Ramy K. Aziz *et al*, 2012). BiGG Models is a completely redesigned Biochemical, Genetic and Genomic knowledge base. BiGG Models contains more than 75 high quality, manually curated genome-scale metabolic models. On the website, users can browse, search and visualize models. BiGG Models connects genome-scale models to genome annotations and external databases. Reaction and metabolite identifiers have been standardized across models to conform to community standards and enable rapid comparison across models. Furthermore, BiGG Models provides a comprehensive application programming interface for accessing BiGG Models with modeling and analysis tools (Zachary A. King *et al*, 2015). Comparative analysis of the well-curated networks is now possible. Pairs of metabolites often appear together in several network reactions, linking them topologically. This co-occurrence of pairs of metabolites in metabolic reactions is termed as metabolite coupling. These metabolite pairs can be directly computed from the stoichiometric matrix, S. Metabolite coupling is derived from the matrix ŜŜ^T (product of binary form of S and transpose of binary form of S), whose off-diagonal elements indicate the number of reactions in which any two metabolites participate together, where Ŝ is the binary form of S. Metabolites with high individual metabolite connectivity also tended to be those with the highest metabolite coupling, as the most connected metabolites couple more often (Scott A Becker *et al*, 2006). MATLAB (The MathWorks, Inc.) is a general-purpose technical computing language and development environment that is widely used in scientific and engineering applications. The Bioinformatics Toolbox for MATLAB is a library of functions that adds bioinformatics capabilities to the MATLAB environment. SBMLToolbox, a toolbox that facilitates importing and exporting models represented in the Systems Biology Markup Language (SBML) in and out of the MATLAB environment and provides functionality that enables an experienced user of either SBML or MATLAB to combine the computing power of MATLAB with the portability and exchangeability of an SBML model. SBMLToolbox supports all levels and versions of SBML (Sarah M. Keating *et al*, 2006).

4. Materials and methods

The comparative study of metabolic network for finding the similarity and dissimilarities is conducted for the following four pathogenic bacteria.

1. *Mycobacterium tuberculosis* (iNJ661)
2. *Staphylococcus aureus* (iSB619)
3. *Salmonella typhimurium* (STM_v1_0)
4. *Helicobacter pylori* (iIT341)

The genome scale metabolic networks of the four pathogenic organisms were collected from different sources like BiGG database, BioModel database and TheModelSeed database. BiGG database is a knowledgebase of biochemically, genetically and genomically structured and reconstructed genome scale metabolic network models. BiGG integrates several published genome scale metabolic networks into one source with typical nomenclature which allows components to be compared across different organisms. BioModels Database is a repository of computational models of biological processes. Models defined from literature are manually curated and supplemented with cross-references. The SEED servers offer open access to frequently updated data, the facility to interpret prokaryotic genomes, the ability to create metabolic reconstructions and comprehensive models of metabolism. It also provides access to hundreds of current metabolic models. The MATLAB files of the four given organisms were downloaded from the database from the BiGG Database.

4.1 Flux Balance Analysis

The constraints in FBA have the form $\alpha \leq vi \leq \beta$ where α and β are the lower and upper limits. Thermodynamic constraints according to the reversibility and irreversibility of a reaction can be applied by setting the lower limit (α) for the equivalent flux to 0 if the reaction is irreversible and -1000 if the reaction is reversible. The upper limit is set to 1000 for reversible and irreversible reactions. The COBRA toolbox was used to run the FBA. COBRA toolbox is a collection of MATLAB scripts for constraint based modeling that are run from within the MATLAB environment. The SBML files were read in COBRA toolbox. A particular metabolic flux distribution within the feasible set was found by using linear programming (LP). The Gurobi

Solver package was used to solve linear programming in cobra toolbox. The fluxes of all reaction were obtained and the growth rate of the organism was calculated by setting the biomass reaction as the objective function. The biomass is usually represented as a stoichiometrically balanced reaction, describing the formation of the biomass from various cellular constituents, as well as various co-factors, which are required to drive the process. There are separate commands in the COBRA toolbox for each of the functions performed for the metabolic model. The detailed codes for performing FBA using the COBRA Toolbox in the MATLAB environment are as follows:-

Command to load the MATLAB file: **load ('filename.mat');**

Command to find out the objective function for the growth of the organism:

objective = checkObjective(model);

Command to change reaction bounds:

model = changeRxnBounds(model, 'rxnNameList', value, 'boundType');

Command to change the Objective Function:

model = changeObjective(model, 'rxnNameList', [objectiveCoeff]) ;

Command to calculate the growth rate:

FBAsolution = optimizeCbModel (model, [osenseStr]);

where *osenseStr* is either 'max' or 'min' to maximize or minimize the value of objective respectively.

*FBAsolution.f***;**

4.2 Simulating Gene Knockouts

The GPR associations are included in the MATLAB files. Gene-protein-reaction associations are embodied in *rxnGeneMat* matrix in the MATLAB files, which is a matrix with as many rows as there are reactions in the model and as many columns as there are genes in the model. The ith row and jth column comprises 1 if the jth gene in genes is associated with the ith reaction in rxns entry of the model and 0 otherwise. The Cobra toolbox has a function called *singleGeneDeletion*

that performs single gene deletion. The function ―*singleGeneDeletion* performs the single gene deletion analysis using FBA. The below given command performs the single gene deletion:

[grRatio, grRateWT, grRateKO, hasEffects, delRxns] = singleGeneDeletion(model);

Where grRatio = grRateWT/grRateKO

grRateWT = growth rate of the wild type

grRateKO = growth rate of the knock outs

hasEffect= contains value 1 for each gene whose deletion affects the growth rate.

delRxns = contains a list of all the reactions whose value is set to 0 for each gene deletion.

The command for performing the single reaction deletion:

[grRatio, grRateWT, grRateKO, hasEffects, delRxns] = singleRxnDeletion(model);

The command for performing the double gene deletion:

[grRatio, grRateWT, grRateKO] = doubleGeneDeletion(model);

4.3 Performing Synthetic Lethality Analysis

It was performed using the Fast – SL algorithm that is implemented using the MATLAB.

It was performed using the Fast – SL algorithm that is implemented using the MATLAB.

Code to perform the synthetic lethal reaction analysis - **fastSL (model, cutoff, order, eliList, atpm)**

Code to perform the synthetic lethal gene analysis - **fastSLgenes (model, cutoff, order, flag)**

where ***model*** is metabolic network file, ***cutoff*** is cutoff percentage value for lethality and have default value of 0.01, ***order*** is order of synthetic lethals and have default value of 2 and maximum value 3, ***eliList*** is the list of the reactions to be excluded from the analysis, ***flag*** is set to 1 for rigorous search and have default value 0, ***atpm*** is ATP Maintenance Reaction Id in model.rxns if other than ATP Maintenance Reaction.

4.4 Robustness analysis

Robustness analysis uses the FBA to analyze the metabolic network properties. In this step, the flux was varied through one reaction and the optimal objective value was calculated as a function of this value. This tells about the sensitivity of the objective value to a particular reaction. We have checked the effect of varying glucose and oxygen uptake on the growth rate. COBRAToolbox has the built in function ***robustnessAnalysis*** to perform this method.

4.5 Phenotypic phase plane analysis

While performing the Robustness analysis, we varied one parameter and made one parameter to be constant. However, in phenotypic phase plane analysis, we varied two parameters glucose and oxygen uptake rate simultaneously and the results were plotted as phenotypic phase plane. It revealed the interaction between two reactions. COBRA Toolbox has the built in function ***phenotypePhasePlane*** to perform this step.

4.6 Metabolite Connectivity

Metabolite connectivity is the number of reactions in which the metabolites occur in. In this step the Stoichiometric matrix **S** of the metabolic model was converted into the binary matrix **Sbin**. It replaces all non-zero elements in the matrix S with '1' in the binary matrix. Then all '1s' in each row of the **Sbin** was added up to determine the number of reactions the metabolites occur in.

4.7 Reaction Essentiality and Metabolite Connectivity

Correlation between reaction essentiality and metabolite connectivity can be calculated in the metabolic network. In this step, all the reactions which were associated with the metabolites were deleted and checked whether it could still produce the biomass or not. Reactions associated with the metabolites could be found out by scanning through the corresponding rows in the Stoichiometric **S** matrix. After then all the associated genes were knocked out and FBA was used to predict the possibility of the growth. Lethality fraction was calculated based on the above result. The semi log graph was plotted with this lethality fraction vs. metabolite connectivity.

5. Results and discussion

5.1 Calculation of Growth rate Using FBA

The growth rates of the four organisms were calculated with help of Flux Balance Analysis. The maximization of biomass was set as the objective function to calculate the growth rate. The growth rate is calculated in the unit millimole/hour/gramDryWeight (mmol gDW^{-1} hr^{-1}). The growth rates of four organisms are shown in Table 1.

TABLE 1 Growth rate of all the four organisms using FBA.

Organism	Number of Reactions	Number of Metabolites	Number of Genes	Growth Rate
Mycobacterium tuberculosis	1025	826	661	0.0522
Staphylococcus aureus	743	655	619	0.0687
Salmonella typhimurium	1076	982	801	0.4779
Helicobacter pylori	554	485	339	0.6928

The table showing growth rates, are the growth rates calculated for the wild type strain of each of the model organism. The objective function for each of the organism was different. Apart from this, growth rates under the alternative substrates like glucose, succinate, pyruvate etc. were also calculated. The calculation of growth rates was done under both aerobic and anaerobic conditions. The growth rates calculated under the alternative substrates were different compared to the wild type ones. For most of the substrate, growth rate under the anaerobic condition was found to be 0 for each of the organism studied. Alternative substrates under which growth rates were calculated are the important substrates that are involved in the major metabolic pathways like glycosis, TCA cycle, oxidative phosphorylation, amino acid metabolism etc. The growth

rates calculated under alternative substrates are shown in the given below tables for each of the organism separately. The maximum substrate uptake rate was set to -20 mmol gDW^{-1} hr^{-1} for every substrate.

Table 2 Growth rates of *Mycobacterium tuberculosis* under the alternative substrates

Substrate	**Aerobic**	**Anaerobic**
Glucose	0.3060	0
Succinate	0.1393	-2.0512e-26
Aspartate	0.0680	3.9109e-20
Phosphate	0.1440	-1.7094e-27
Glycerol	0.3060	6.9372e-26
Pyruvate	0.3060	-9.6864e-27

Table 3 Growth rates of *Staphylococcus aureus* under the alternative substrates

Substrate	**Aerobic**	**Anaerobic**
Glucose	0.0141	2.2559e-26
Succinate	Absent	Absent
Aspartate	0.0123	0
Phosphate	0.0103	0
Glycerol	-9.6865e-18	-4.1655e-18
Pyruvate	Absent	Absent

Table 4 Growth rates of *Helicobacter pylori* under the alternative substrates

Substrate	**Aerobic**	**Anaerobic**
Glucose	4.1510	0
Succinate	5.3325	0
Aspartate	5.2489	0
Phosphate	4.1510	0
Glycerol	Absent	Absent
Pyruvate	4.9698	1.5179e-29

Table 5 Growth rates of *Salmonella typhimurium* under the alternative substrates

Substrate	**Aerobic**	**Anaerobic**
Glucose	1.9317	0.4500
Succinate	1.5507	0.0525
Aspartate	1.4971	0.2959
Phosphate	0.4779	0.0525
Glycerol	1.1227e-12	-9.3396e-13
Pyruvate	1.1584e-12	-4.6855e-13

5.2 Retrieving Essential Reactions and Essential Genes from Gene Knockouts

If a single gene is associated with multiple reactions, the deletion of that gene will result in the removal of all associated reactions. On the other hand, a reaction that can be catalyzed by multiple non-interacting gene products will not be removed in a single gene deletion. The essential reactions were extracted by reaction knockout method. The essential genes were extracted by gene knockout method. By the aid of the COBRA toolbox, we can retrieve the

essential reactions in the metabolic network. After gene knockout, a mutant strain of the organism is generated and it also provides the gene-reaction ratio (grRatio) of the mutant strain and wild type of the organisms. The grRatio is the ratio between the grRatio of wild type strain and grRatio of the mutated strain. If the grRatio is less than 0.05, those genes and reactions are the essential genes and essential reactions of the organisms. Numbers of essential reactions and essential genes present in metabolic network of each organism have been shown in Table 6.

TABLE 6 Essential Genes and Essential Reactions

Organism	***S. aureus***	***M. tuberculosis***	***S. typhimurium***	***H. pylori***
Number of Essential Reactions	244	314	345	256
Number of Essential Genes	168	188	201	186

The essential reactions thus obtained for each of the organism can be compared pairwise with essential reactions of each of the organism. The pairwise comparison was performed with the help of Compare plugin in the Notepad++. Common essential reactions thus obtained by the pairwise comparison are shown in the below given table 7

Table 7 Common Essential Reactions.

Pairwise Combination of Organisms	***M. tuberculosis***	***M. tuberculosis***	***M. tuberculosis***	***S. aureus***	***S.aureus***	***S. typhimurim***
	S. typhimurium	***S. aureus***	***H. pylori***	***S. typhimurim***	***H.pylori***	***H. pylori***
No. of Common Essential Reactions	48	47	96	60	93	58

Gene knockout performed on each of the organism can be visualized using the MATLAB. Double gene deletion was performed for all the genes present in each of the organism. The lethality of double gene deletion can be visualized where the scale of lethality varies from 0 to 1. Scale of 1 refers to highly lethal deletion for the particular two genes and scale of 0 refers the lethal effect to be normal as shown in figure 1 to figure 4.

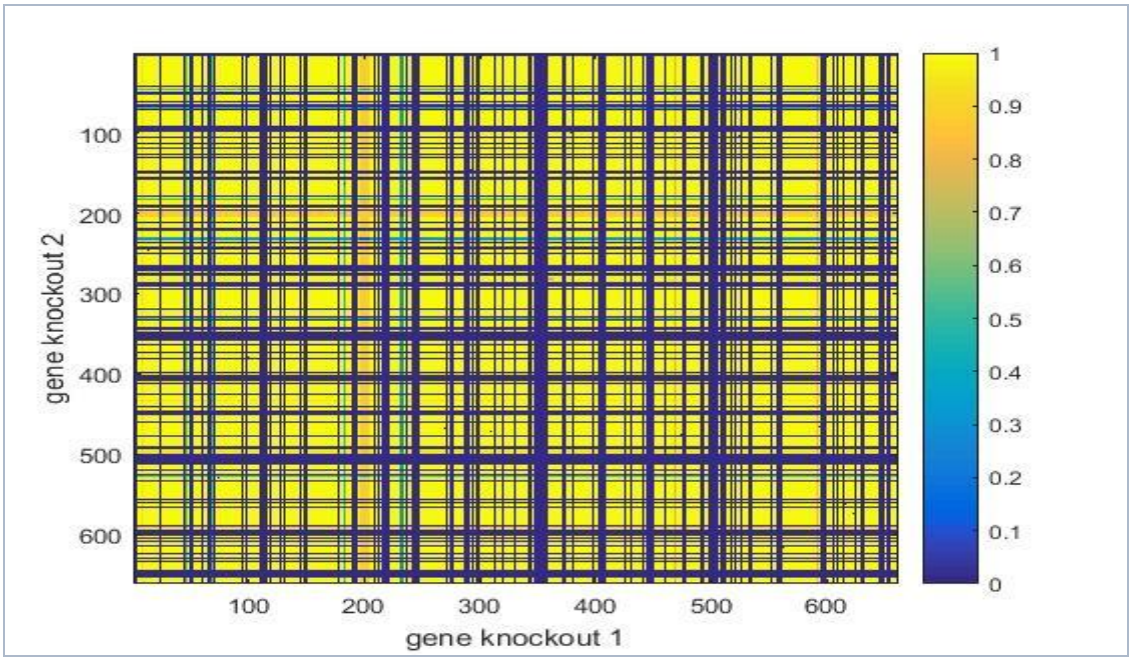

Figure 1 Gene knockout of *Mycobacterium tuberculosis* metabolic network using double gene deletion.

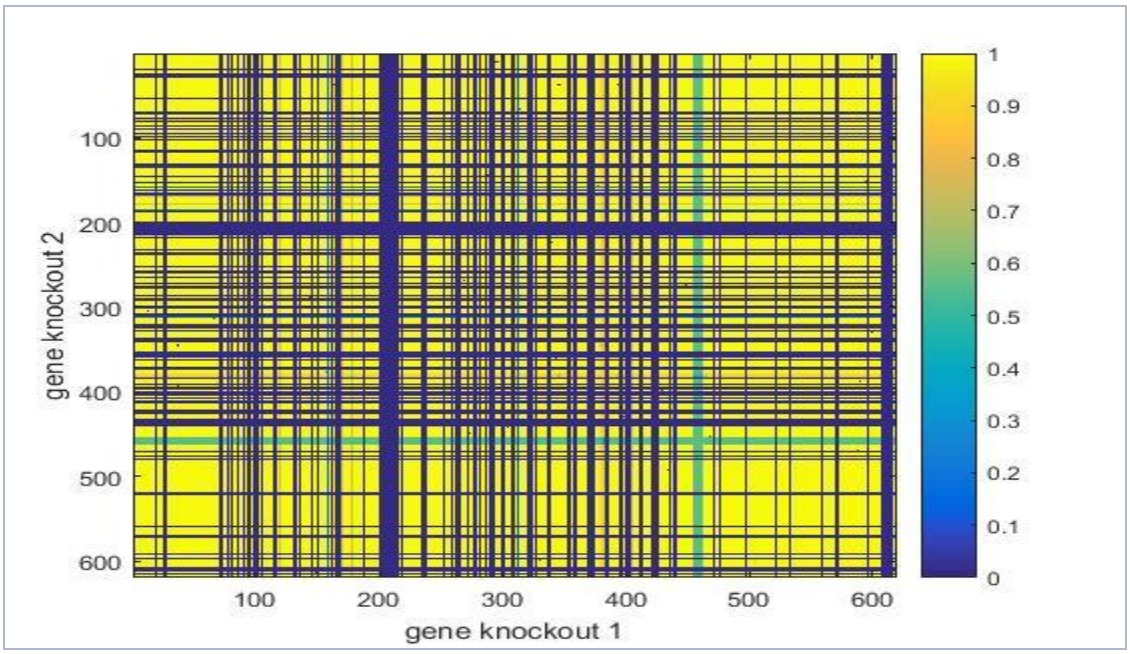

Figure 2 Gene knockout of *Staphylococcus aureus* metabolic network using double gene deletion.

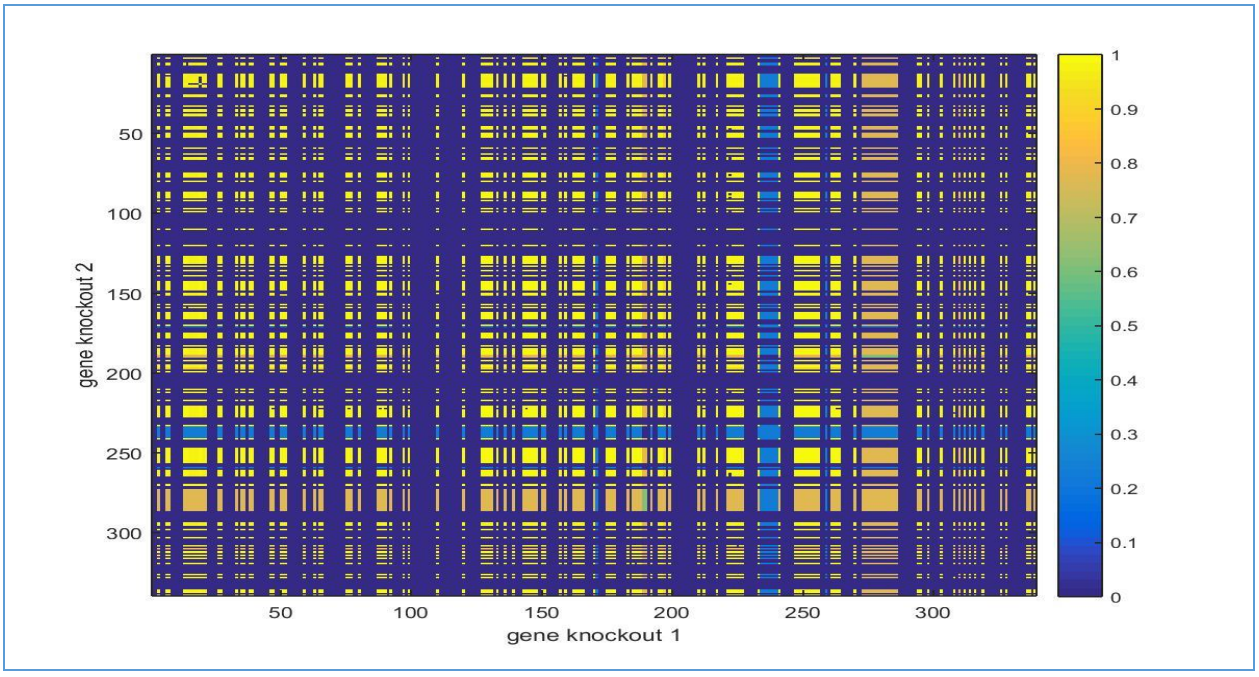

Figure 3 Gene knockout of *Helicobacter pylori* metabolic network using double gene deletion method.

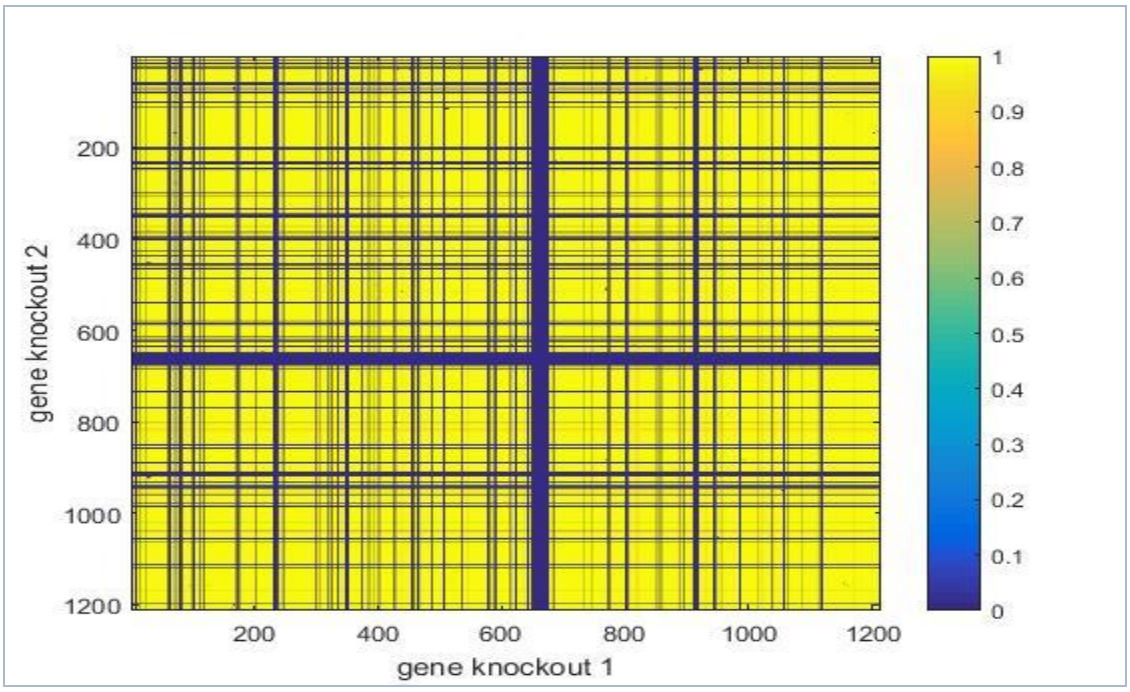

Figure 4 Gene knockout of *Salmonella typhimurium* metabolic network using double gene deletion method.

5.3 Synthetic Lethality Analysis

It was performed using the algorithm Fast–SL algorithm which is implemented in the MATLAB using the COBRAToolbox. The synthetic gene lethality and the synthetic reaction lethality were performed separately. For synthetic gene lethality, both single and double gene lethality was performed. Similarly for synthetic reaction lethality, both single and double reaction lethality was performed. In single gene lethality there was deletion of single gene and similarly in double gene lethality pair of genes were deleted. Similar procedure were followed for reaction lethality also. The result for gene lethality has been shown in the Table 8.

Table 8 Synthetic Gene Lethality

Organisms	**Number of Genes for Single Lethality**	**Number of Genes for Double Lethality**
Mycobacterium tuberculosis	188	188
Staphylococcus aureus	168	168
Salmonella typhimurium	202	204
Helicobacter pylori	168	168

The result for reaction lethality has been shown in the Table 9.

Table 9 Synthetic Reaction Lethality

Organism	**Number of Reactions for Single Lethality**	**Number of Reactions for Double Lethality**
Mycobacterium tuberculosis	314	314
Staphylococcus aureus	244	244
Salmonella typhimurium	346	346
Helicobacter pylori	135	244

5.4 Robustness Analysis

Robustness analysis was performed to study the effect of the glucose and oxygen uptake on the growth rates. This analysis was performed for the metabolic network of each of the organism. The sensitivity of the growth of the metabolic network was checked against the varying glucose and oxygen uptake rate. To determine the effect of varying glucose uptake on growth, oxygen uptake rate was fixed at -17 mmol gDW^{-1} hr^{-1} which is a realistic uptake of oxygen. To determine the effect of varying oxygen uptake on growth, glucose uptake rate was fixed at 10

mmol gDW^{-1} hr^{-1} which is a realistic uptake of glucose. Results for the robustness analysis for each of the organism have been shown in the below given figures.

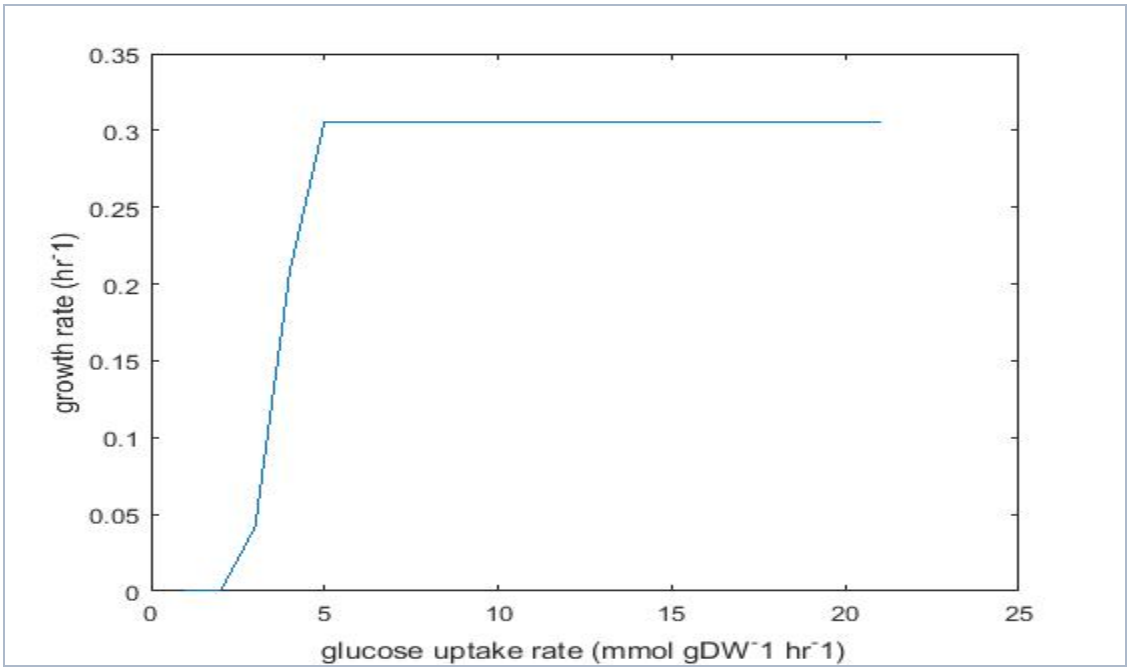

Figure 5 Robustness analysis of *Mycobacterium tuberculosis* metabolic network for maximum growth rate while varying glucose uptake rate with oxygen uptake fixed at 17 mmol gDW^{-1} hr^{-1}.

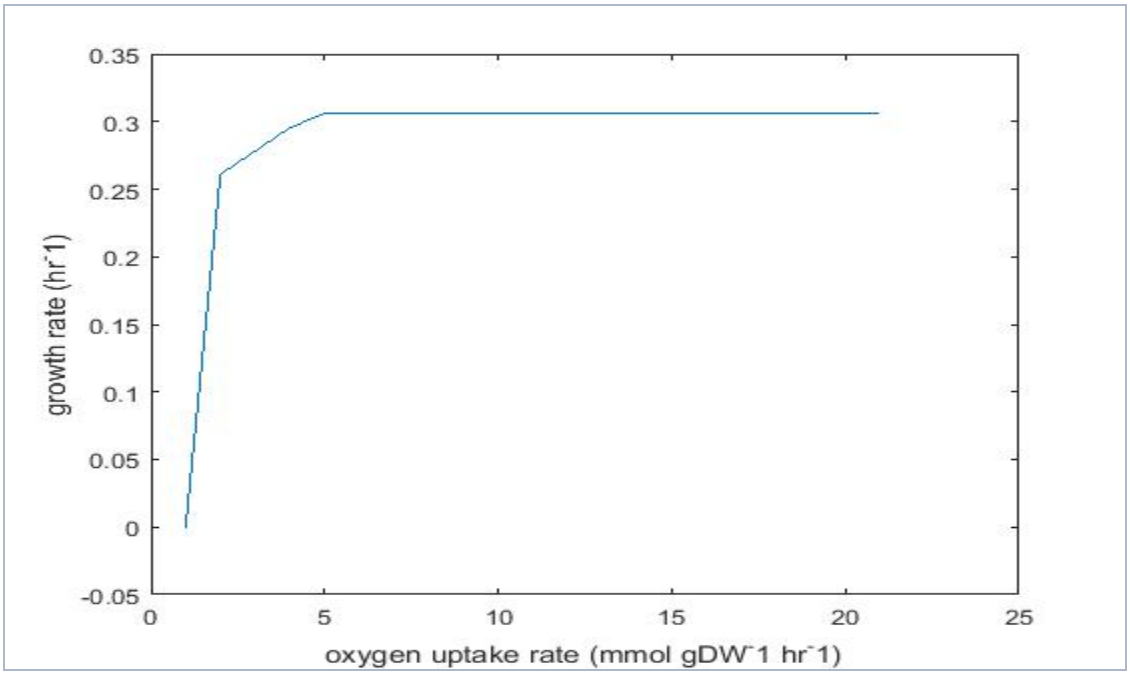

Figure 6 Robustness analysis of *Mycobacterium tuberculosis* for maximum growth rate while varying oxygen uptake rate with glucose uptake fixed at 10 mmol gDW^{-1} hr^{-1}.

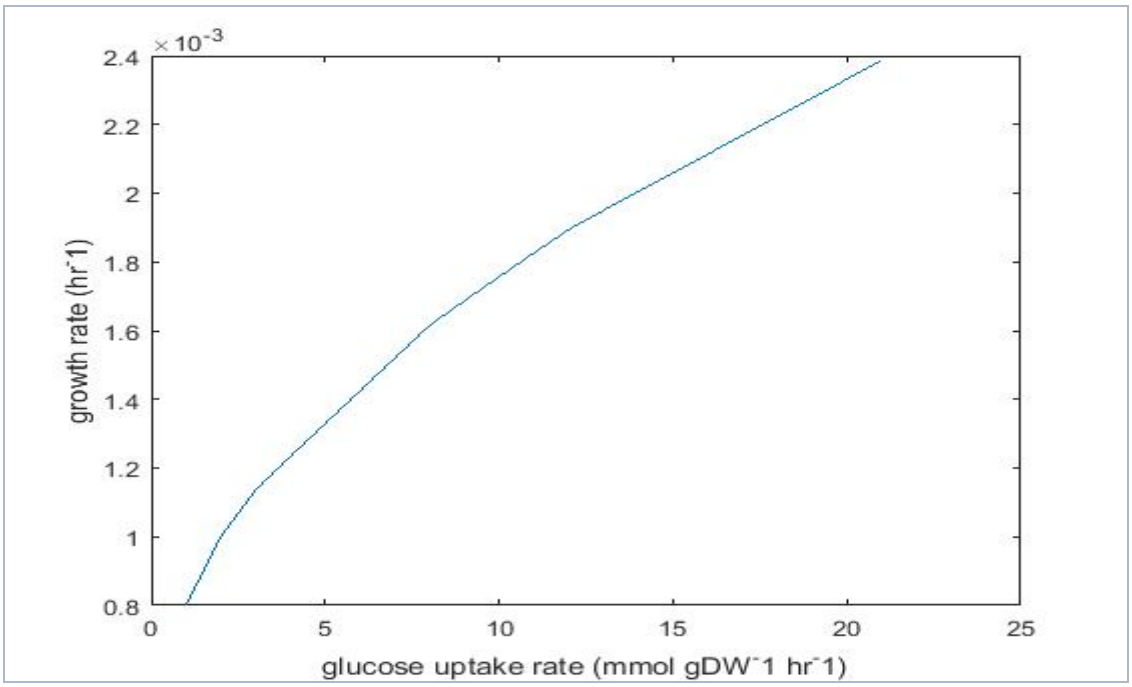

Figure 7 Robustness analysis of Staphylococcus aureus metabolic network for maximum growth rate while varying glucose uptake rate with oxygen uptake fixed at 17 mmol gDW^{-1} hr^{-1}

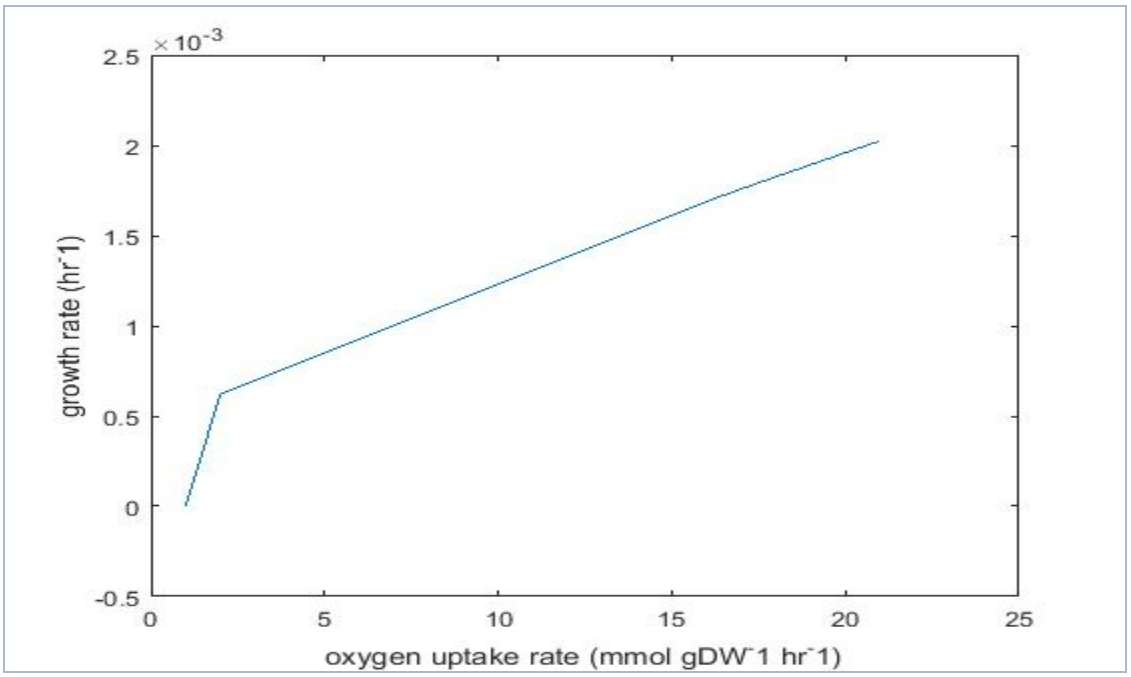

Figure 8 Robustness analysis of *Staphylococcus aureus* metabolic network for maximum growth rate while varying oxygen uptake rate with glucose uptake fixed at 10 mmol gDW^{-1} hr^{-1}.

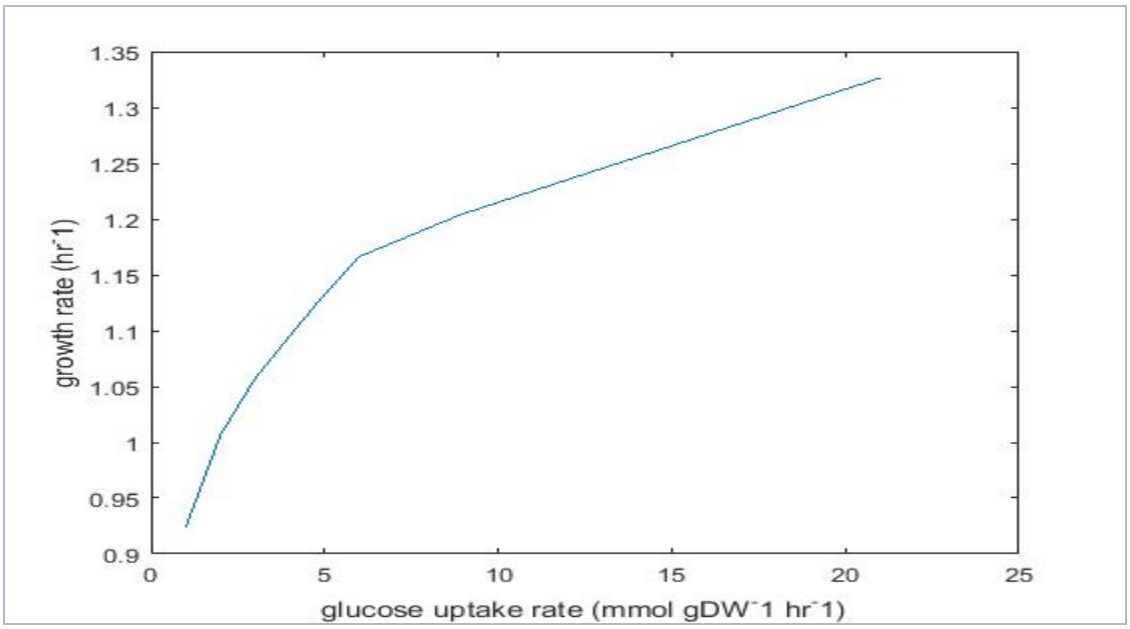

Figure 9 Robustness analysis of *Helicobacter pylori* metabolic network for maximum growth rate while varying glucose uptake rate with oxygen uptake fixed at 17 mmol gDW^{-1} hr^{-1}.

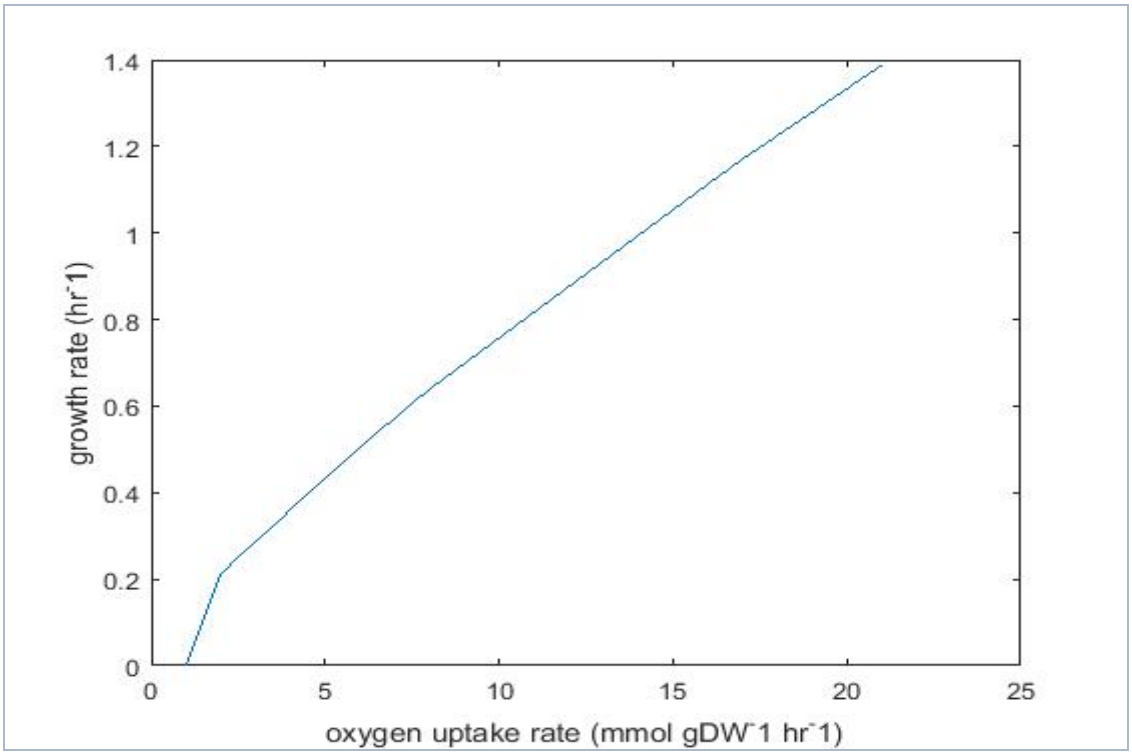

Figure 10 Robustness analysis of *Helicobacter pylori* metabolic network for maximum growth rate while varying oxygen uptake rate with glucose uptake fixed at 10 mmol gDW^{-1} hr^{-1}.

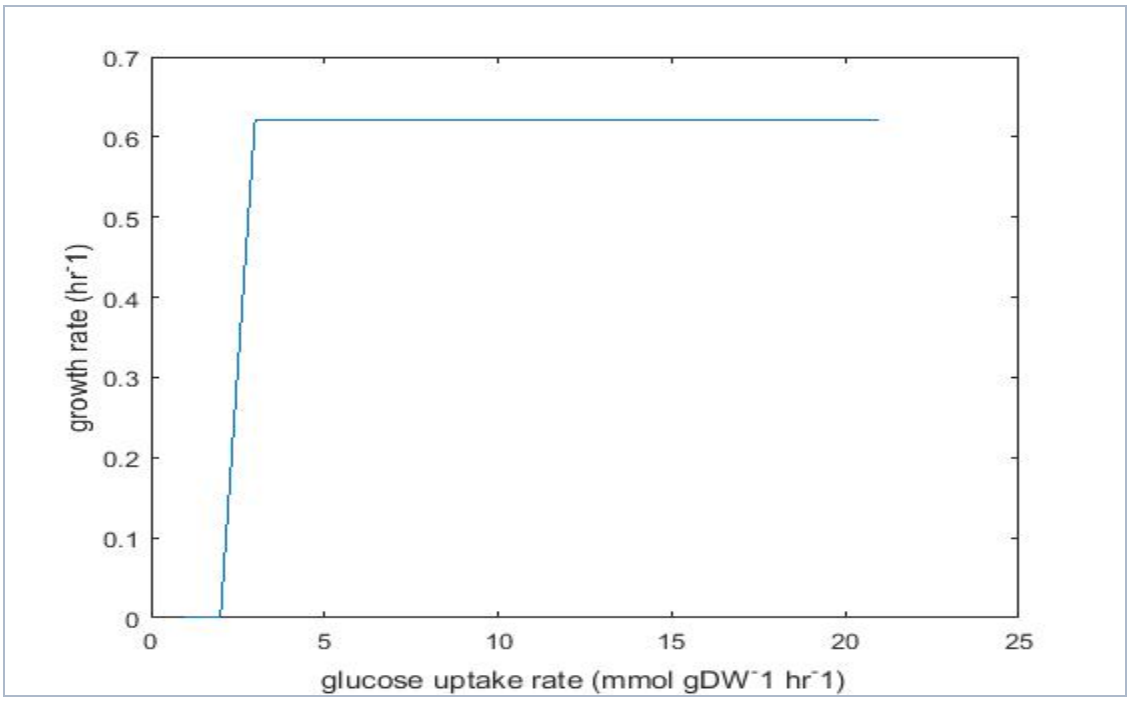

Figure 11 Robustness analysis of *Salmonella typhimurium* metabolic network for maximum growth rate while varying glucose uptake rate with oxygen uptake fixed at 17 mmol gDW^{-1} hr^{-1}.

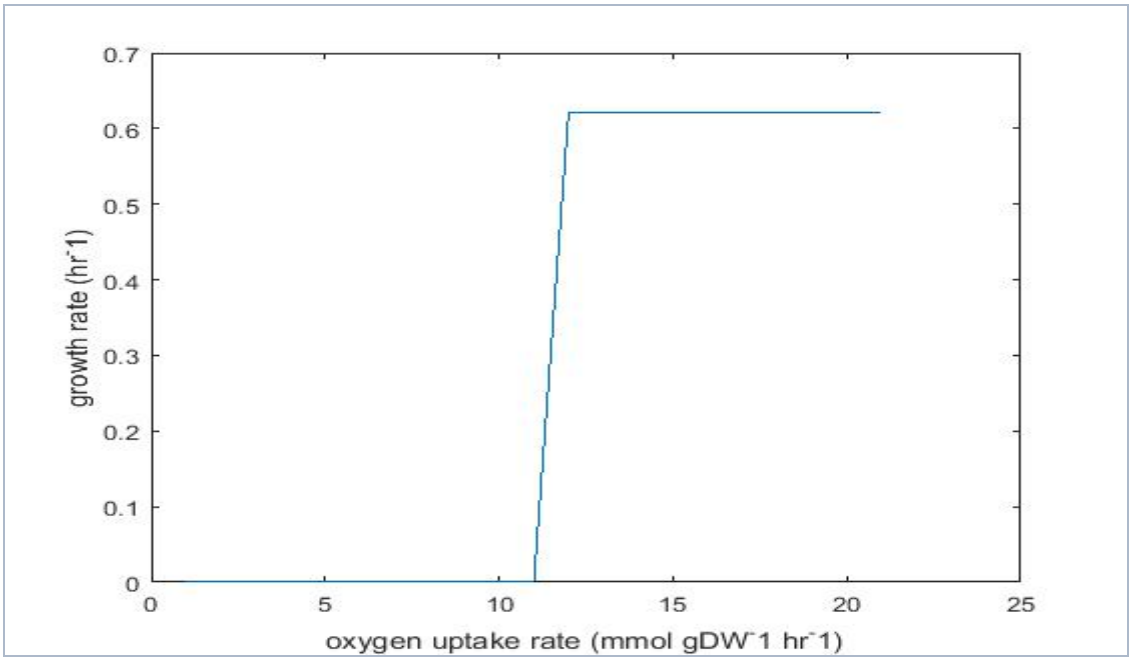

Figure 12 Robustness analysis of *Salmonella typhimurium* metabolic network for maximum growth rate while varying oxygen uptake rate with glucose uptake fixed at 10 mmol gDW^{-1} hr^{-1}.

5.5 Phenotypic Phase Plane Analysis

Phenotypic phase plane analysis was performed by varying two substrates simultaneously. In this study phenotypic phase plane analysis is done by varying the glucose and oxygen uptake rate simultaneously. It assesses the effect of the simultaneous variation of the uptake rate of the glucose and oxygen on the growth rate. The effect was plotted on the 2-D graph as well as on the 3-D graph. The plot is divided into different phases with different color-coding. The 3-D plot was created using the MATLAB function surfl. The results for the phenotypic phase plane analysis for each of the different organism are given below in the figure 13 to figure 16.

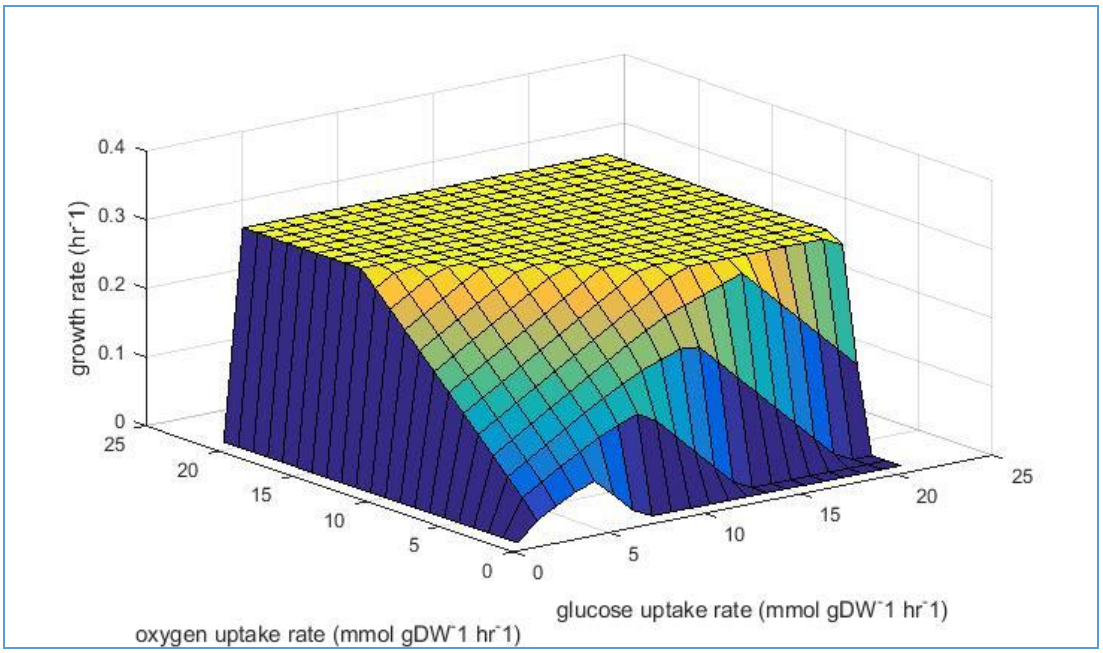

Figure 13 Glucose and oxygen phenotype phase plane for *Mycobacterium tuberculosis* genotype.

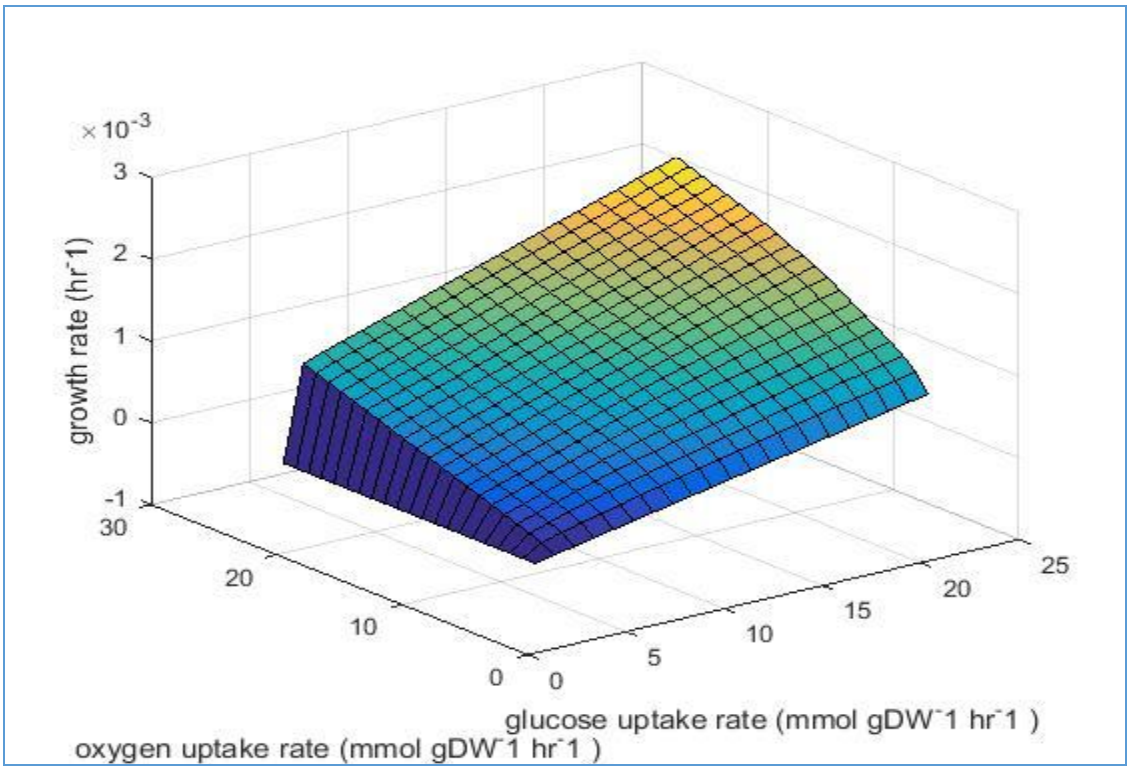

Figure 14 Glucose and oxygen phenotype phase plane for *Staphylococcus aureus* genotype.

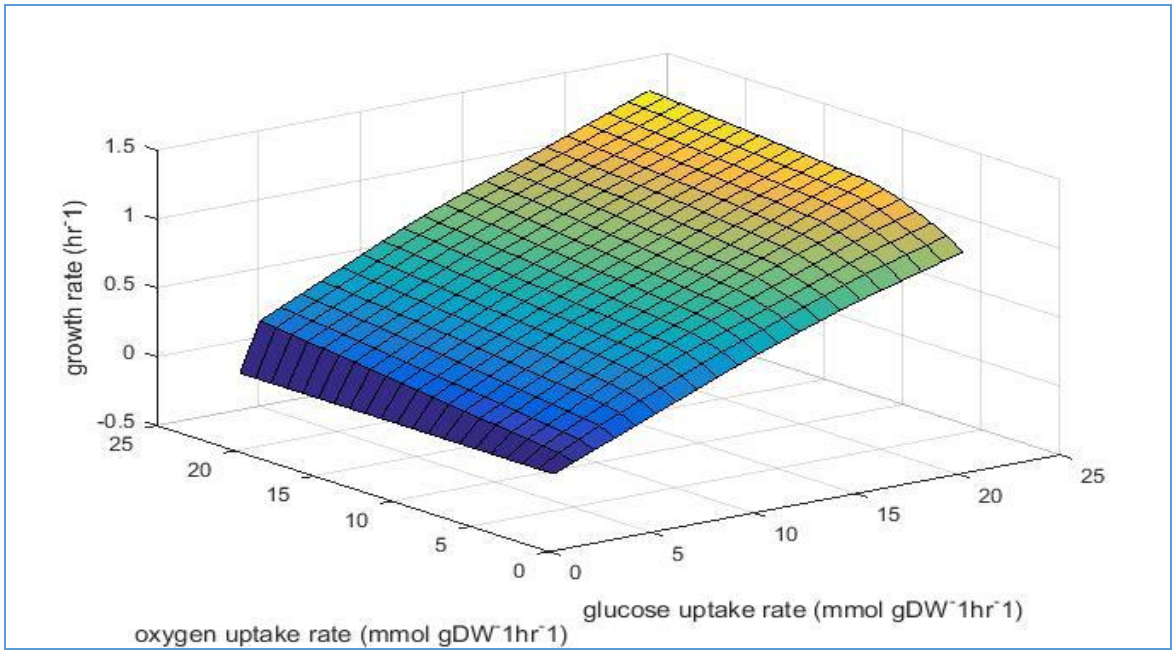

Figure 15 Glucose and oxygen phenotype phase plane for *Helicobacter pylori* genotype.

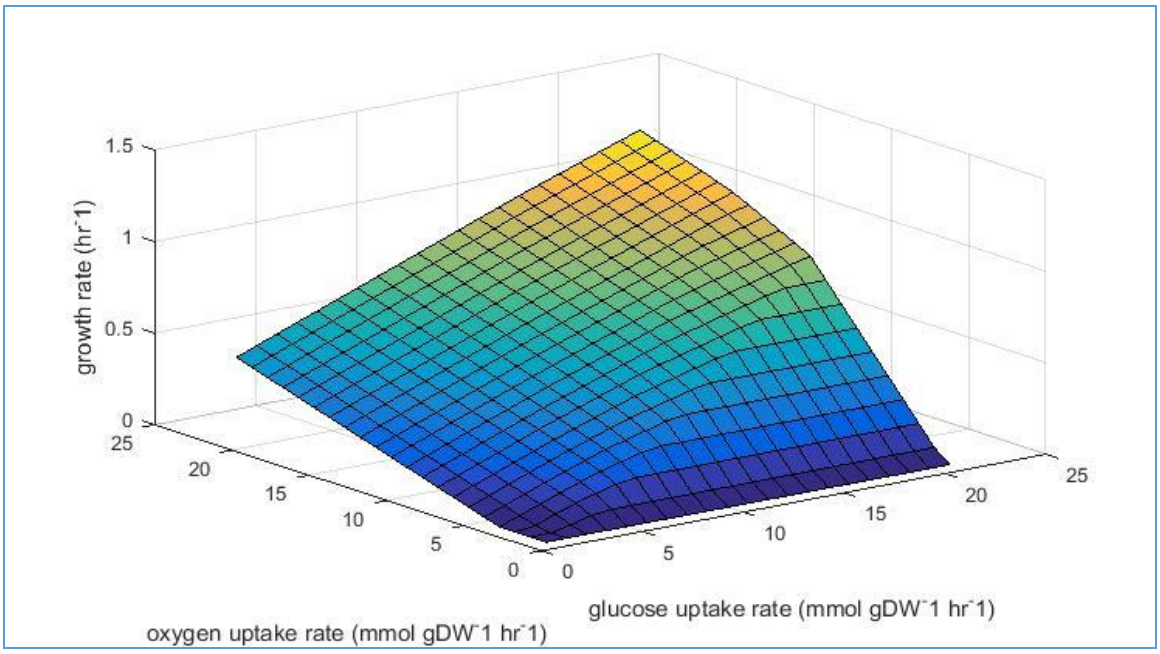

Figure 16 Glucose and oxygen phenotype phase plane for *Salmonella typhimurium* genotype.

5.6 Metabolite Connectivity

For each of the four networks, the number of reactions in which each possible metabolite is occurs was determined. It is the measure of individual connectivity of each of the metabolite in the metabolic network. This metabolite connectivity was plotted on the log-log scale for each of the four metabolic networks. The approximate linear appearance of the curve relates to the power law distribution of the metabolic networks. The plots of four different metabolic networks show that there are very few metabolites, which are highly connected. Most of the metabolites occur only in a few reactions. The highly connected metabolites are the 'global factors' similar to the hub protein in the protein-protein interaction network. The least connected metabolites are the 'local factors' which occur in linear pathway. The power law distribution shows that the networks are scale-free. Metabolite connectivity in the metabolic networks of each of the organisms is shown in below given figures.

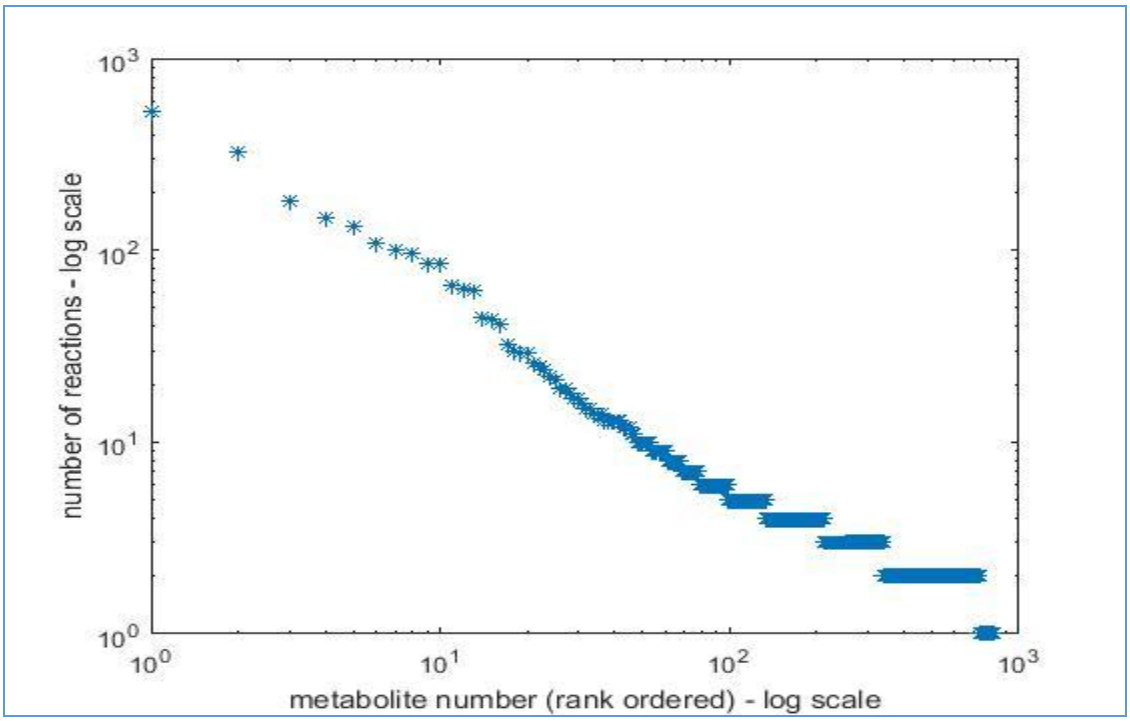

Figure 17 Metabolite connectivity of *Mycobacterium tuberculosis* ((loglog plot).

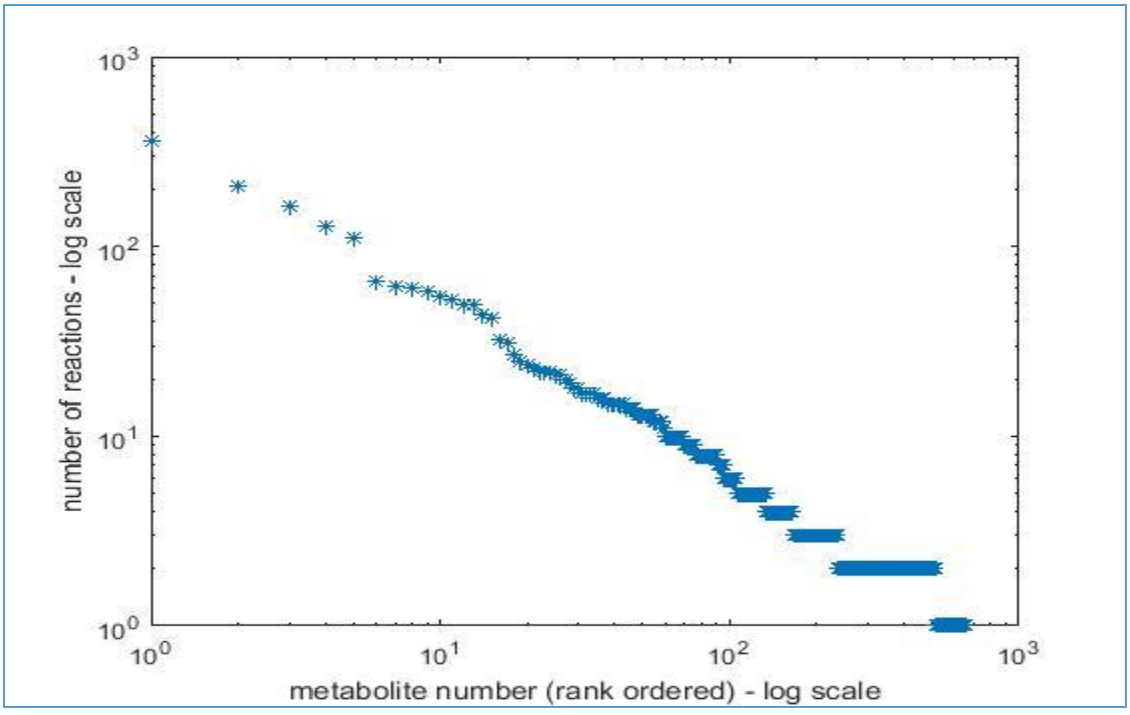

Figure 18. Metabolite connectivity of *Staphylococcus aureus* (loglog plot).

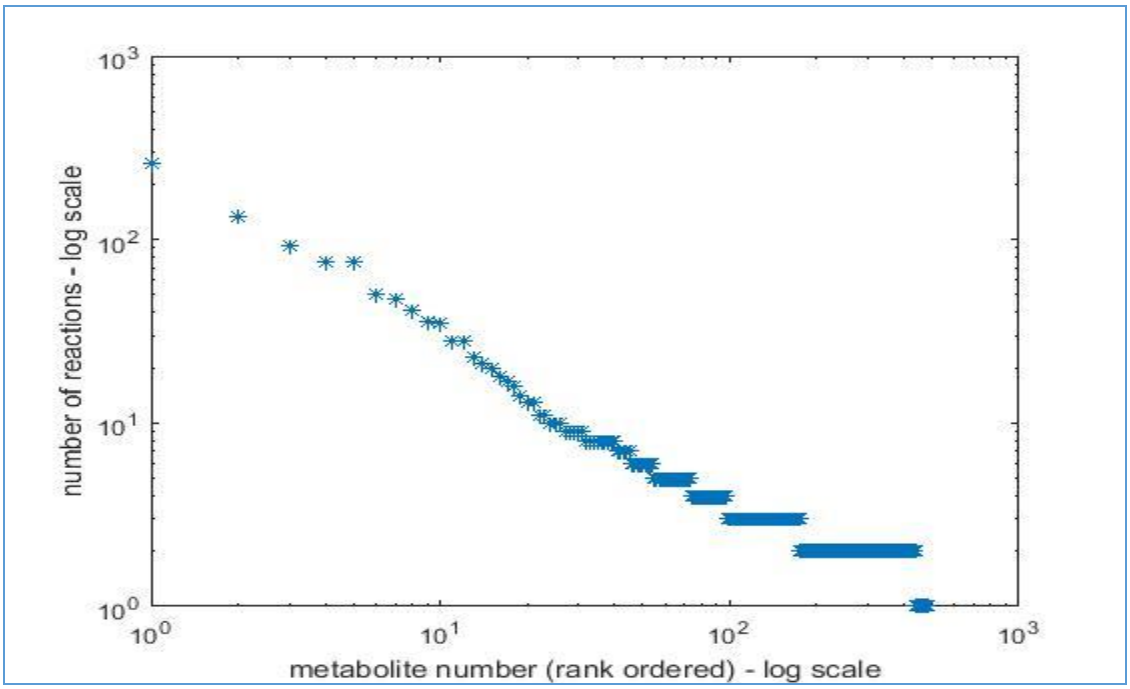

Figure 19 Metabolite connectivity of *Helicobacter pylori* ((loglog plot)

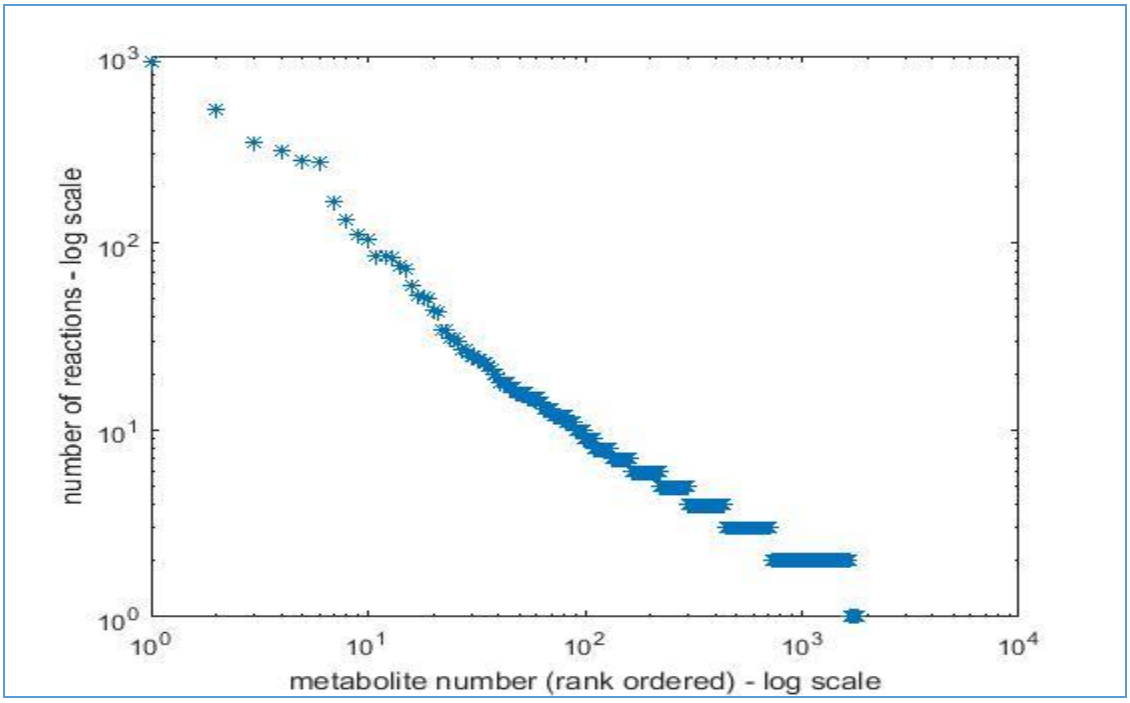

Figure 20 Metabolite connectivity of *Salmonella typhimurium* ((loglog plot)

Reaction participation for each of the reaction in the network was also calculated. It is nothing but the number of metabolites per reaction. The most common type of reaction in the metabolic

network of each of the organism is the bi-linear reaction involving two substrates and two products. Average of the reaction participation in each of the network is shown in the table 10.

Table 10 Average Reaction Participation.

Organism	**Average reaction participation**
Mycobacterium tuberculosis	4.6576
Staphylococcus aureus	5.1427
Helicobacter pylori	3.8523
Salmonella typhimurium	4.1796

5.7 Reaction Essentiality and Metabolite Connectivity

In order to correlate the reaction essentiality and the metabolite connectivity in each of the metabolic networks, metabolite associated reactions were found out and finally they are deleted. After then FBA was used to predict the growth if possible. This step was performed for the metabolic networks of each of the four organisms. The results were plotted on the semi-log scale with metabolite connectivity and lethality of the network. The results for the metabolic network of four different organisms show that some less connected metabolites have a higher lethality fraction than highly connected metabolites. For the *H. pylori* core model, the average lethality fraction lies between 0.3 and 0.6 for the majority of the metabolites, regardless of their connectivity as shown in figure 23. For the *M. tuberculosis* core model, the average lethality fraction lies between 0.3 and 0.6 for the majority of the metabolites, regardless of their connectivity as shown in figure 21. For the *S. aureus* core model, the average lethality fraction lies between 0.2 and 0.6 for the majority of the metabolites, regardless of their connectivity as shown in figure 22. For the *S. typhimurium* core model, the average lethality fraction lies between 0.2 and 0.5 for the majority of the metabolites, regardless of their connectivity as shown in figure 24.

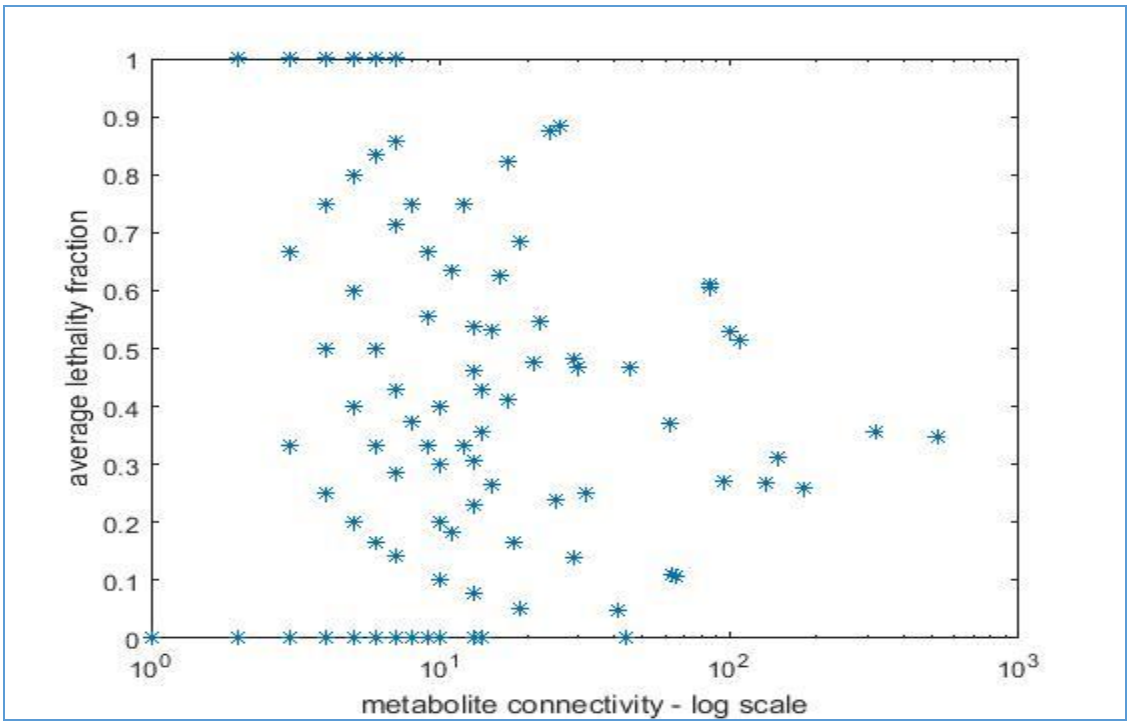

Figure 21 Correlation between average lethality fraction and metabolite connectivity for *Mycobacterium tuberculosis.*

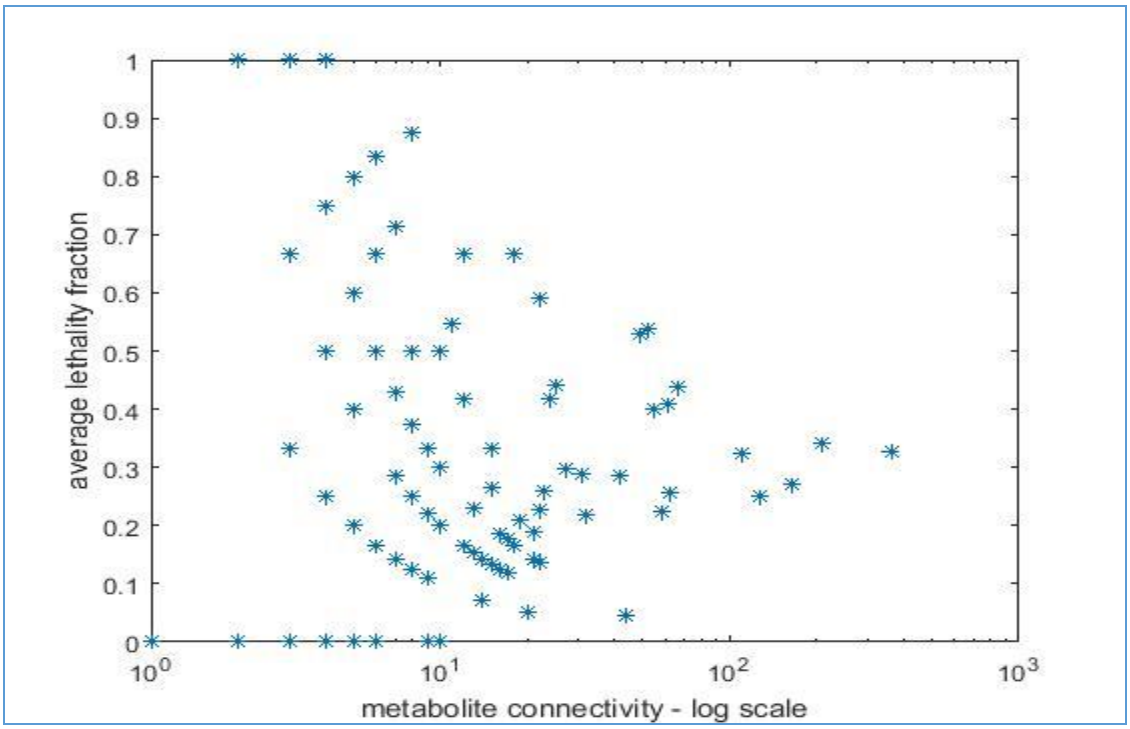

Figure 22 Correlation between average lethality fraction and metabolite connectivity for *Staphylococcus aureus*

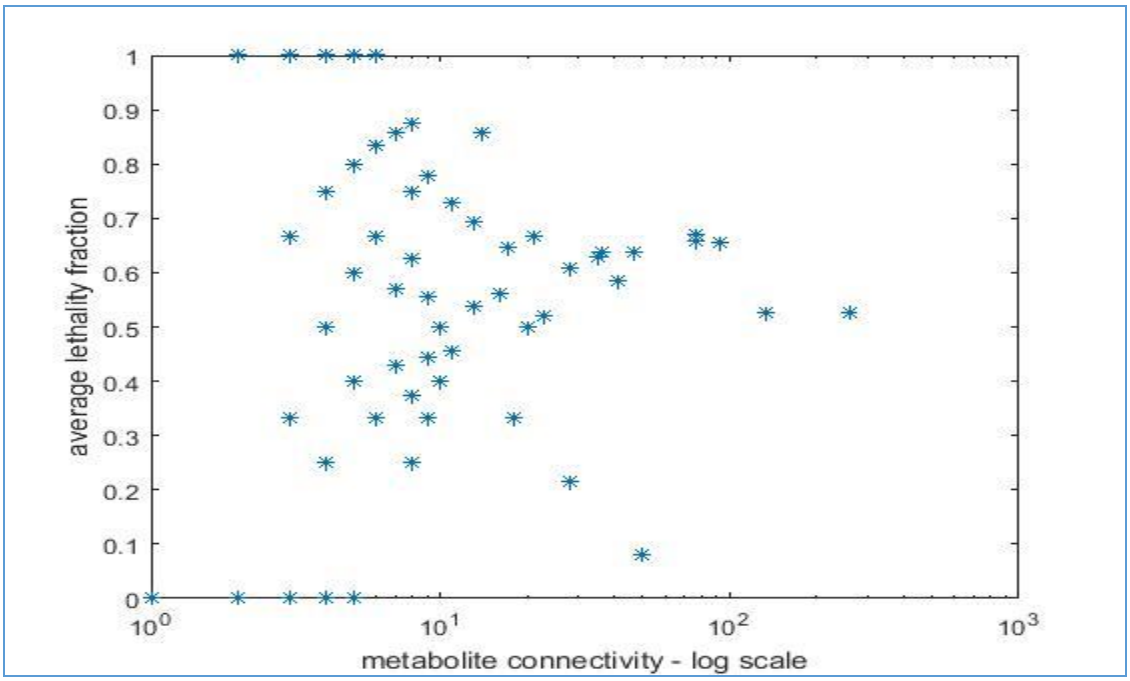

Figure 23 Correlation between average lethality fraction and metabolite connectivity for *Helicobacter pylori.*

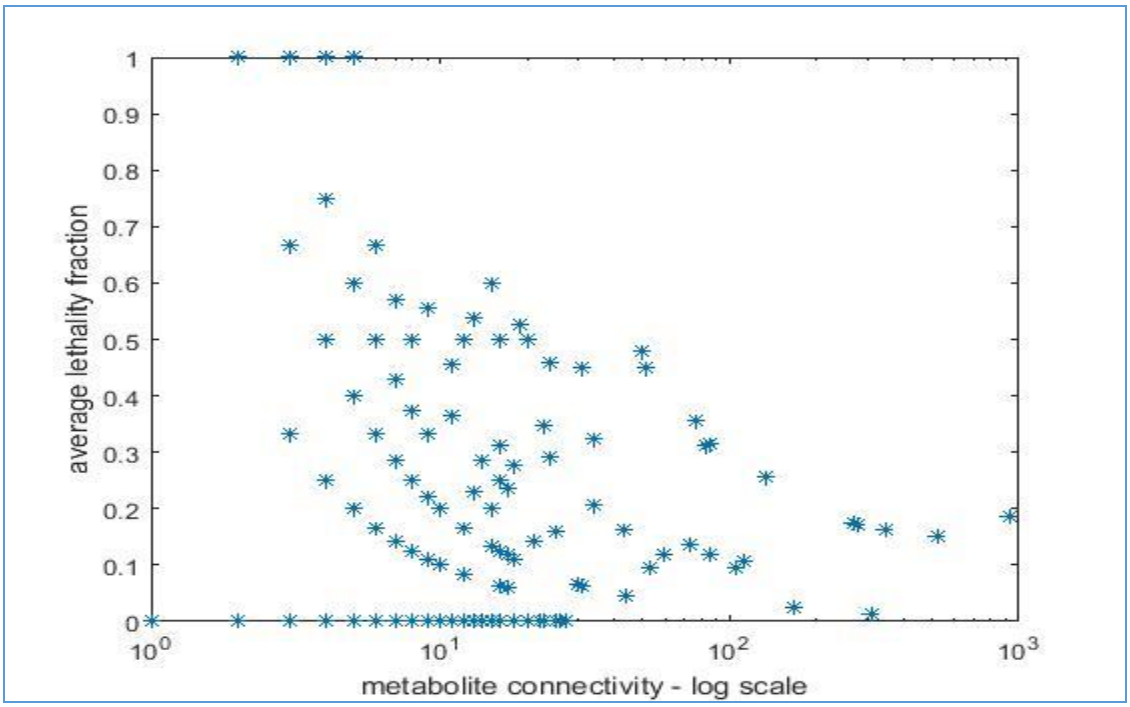

Figure 24 Correlation between average lethality fraction and metabolite connectivity for *Salmonella typhimurium*

6. Conclusion

All the four metabolic network files are the reconstructed models of these organisms. Maximization of growth is very important compartmental objective to find out good FBA estimation using the kind of objective function explored. While performing the FBA analysis, growth was possible both aerobically and anaerobically for each of the four organisms. In this study, the variation in flux values were generally associated with important metabolic pathways like Glycolysis, Oxidative Phosphorylation, and Amino Acid Biosynthesis etc. In the flux variability analysis, sub-system or pathways comes out to have the high flux variability and essential sub-systems resulted from the gene knockout analysis are similar. Single gene deletion and single reaction deletion functions estimate the amount of essential genes and reactions by deleting the certain genes and reactions and analyzing the impact of the deletion. These essential genes and reactions are the major players in the regulation of important pathways of the organisms. These genes and reactions are also involved in the signal transduction pathways. These genes and reactions serve as the candidate for the potential drug target against diseases caused by these organisms. The identification of synthetic lethals in organisms can be used to understand complex genetic interactions between genes and identify drug targets for combinatorial therapy. Fast – SL algorithm exploits the structure of the metabolic network better than previous algorithms, to eliminate combinations of reactions/genes that are guaranteed not to produce a lethal phenotype under the conditions considered. This algorithm also identifies synthetic lethal gene sets rigorously, by carefully considering the GPR associations in the metabolic model. In this study, it has been found that the highly connected metabolites contributed an uneven percentage of the enriched coupling interactions in that highly connected metabolites are more connected to each other than would be expected from their individual connectivities alone. Metabolite connectivity enables us to understand the basic structure of these metabolic networks of four different organisms. The correlation study between metabolite connectivity and reaction essentiality gives the major insight into the lethality fraction of the metabolic networks. The lethality fraction is computed as the ratio between metabolite connectivity and reaction essentiality. The average lethality fraction of all the four organisms studied, ranged from 0.2 to 0.6.

7. References

1. Adam M Feist and Bernhard Ø Palsson 2008, The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli,* Nature Publishing Group.
2. Aditya Pratapa, Shankar Balachandran and Karthik Raman 2015, Fast-SL: an efficient algorithm to identify synthetic lethal sets in metabolic networks, Bioinformatics.
3. Akhil Kumar, Patrick F Suthers and Costas D Maranas 2012, MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases, BMC Bioinformatics.
4. Balaji Veeramani and Joel S. Bader 2009, Metabolic Flux Correlations, Genetic Interactions, and Disease, Journal of Computational Biology.
5. Bates DG, Postlethwaite I, Ma L, Iglesias PA 2006, Robustness analysis of biochemical network models, Systems biology.
6. Becker SA, Palsson BØ 2005, Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation, BMC Microbiology.
7. Becker SA1, Price ND, Palsson BØ 2006, Metabolite coupling in genome-scale metabolic networks, BMC Bioinformatics.
8. Christophe H. Schilling, Markus W. Covert, Iman Famili, George M. Church, Jeremy S. Edwards, and Bernhard O. Palsson 2002, Genome-Scale Metabolic Model of *Helicobacter pylori* 26695, JOURNAL OF BACTERIOLOGY.
9. Jamshidi N, Palsson BØ 2007, Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets, BMC Systems Biology.
10. Jan Schellenberger *et al* 2011**,** Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0, Nature Publishing Group.
11. Jeffrey D Orth, Ines Thiele & Bernhard Ø Palsson 2010, What is flux balance analysis, Computational Biology.
12. Jeremy S Edwards and Bernhard O Palsson 2000, Metabolic flux balance analysis of Escherichia coli K-12 gene deletions, BMC Bioinformatics.

13. Jeremy S. Edwards, Ramprasad Ramakrishna, Bernhard O. Palsson 2001, Characterizing the metabolic phenotype: a phenotype phase plane analysis, Biotechnology and Bioengineering.
14. Joanna Masel and Mark L. Siegal 2010, Robustness: mechanisms and consequences, Trends Genetics.
15. Joshua J. Hamilton, Jennifer L. Reed 2012, Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint - Based Models, PLoS ONE.
16. Karthik Raman and Nagasuma Chandra 2009, Flux balance analysis of biological systems: applications and challenges, Oxford Journals.
17. Karthik Raman, Preethi Rajagopalan and Nagasuma Chandra 2005, Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs, PLoS Computational Biology, 2005.
18. Matthias Heinemann, Anne Kummel, Reto Ruinatscha, Sven Panke 2005, In Silico Genome-Scale Reconstruction and Validation of the Staphylococcus aureus Metabolic Network, Wiley Interscience.
19. McClelland M *et al* 2001, Complete genome sequence of Salmonella enterica serovar Typhimurium LT2, Nature.
20. Patrick F Suthers, Zomorrodi A, Maranas CD 2009, Genome-scale gene/reaction essentiality and synthetic lethality analysis, Molecular System Biology.
21. Ratul Chowdhury, Anupam Chowdhury and Costas D. Maranas 2015, Using Gene Essentiality and Synthetic Lethality Information to Correct Yeast and CHO Cell Genome-Scale Models, Metabolites.
22. Sebastian M.B. Nijman 2011, Synthetic lethality: General principles, utility and detection using genetic screens in human cells, FEBS Letter.
23. Thiele I *et al* 2011, A community effort towards a knowledge-base and mathematical model of the human pathogen Salmonella Typhimurium LT2, BMC Systems Biology.
24. Thiele I, Vo TD, Price ND, Palsson BØ 2005, Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants, Journal of Bacteriology.

25. Vashisht *et al* 2014, Systems level mapping of metabolic complexity in Mycobacterium tuberculosis to identify high-value drug targets, Journal of Translational Medicine.

26. Zachary A. King *et al* 2015, BiGG Models: A platform for integrating, standardizing and sharing genome-scale models, Oxford Journals.