UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA MOLECULAR E CELULAR

ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD)

Gustavo Tamasco

Ribeirão Preto

Brasil

2022

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA MOLECULAR E CELULAR

ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization

Gustavo Tamasco

Ribeirão Preto

Brasil

2022

Gustavo Tamasco

ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization

Dissertação apresentada ao Programa de Biologia Molecular e Celular da Faculdade de Medicina de Ribeirão Preto - USP, como requisito parcial para obtenção do título de Mestre do curso de pós-graduação em Biologia Molecular e Celular.

Orientadores:
Prof. Dr. Rafael Silva-Rocha
Prof. Dr. Ricardo Roberto da Silva

Ribeirão Preto

Brasil

2022

Tamasco G, da Silva RR, Silva-Rocha R. ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization. bioRxiv. 2021 Jan 1.

78 f. : il. ; 30 cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Biologia Celular e Molecular.

Orientador: Prof. Dr. Rafael Silva-Rocha e Prof. Dr. Ricardo Roberto da Silva

1. Biologia de Sistemas 2. FBA 3. GSMR 4. Reconstrução de Vias Metabólicas 5. Visualização de Vias Metabólicas. 6. Engenharia Metabólica

Aluno: Gustavo Tamasco

Título: ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências. Área de concentração: Biologia Celular e Molecular. Orientador: Prof. Dr. Rafael Silva Rocha; Prof. Dr. Ricardo Roberto da Silva

Aprovado em: ___/___/___

Banca Examinadora

Prof(a).Dr(a).:_____
Instituição:_____
Assinatura:_____

Prof(a).Dr(a).:_____
Instituição:_____
Assinatura:_____

Prof(a).Dr(a).:_____
Instituição:_____
Assinatura:_____

Acknowledgement

Primeiramente gostaria de agradecer a minha esposa Daniela que me deu suporte durante toda essa jornada. Mantendo minha mente no lugar mesmo durante o período da pandemia. Sou grato por todo carinho e confiança depositados em mim, bem como pela parceria em todos os momentos. Também agradeço pela paciência, compreensão, por ouvir minhas reclamações e muitas vezes sem nem entender, e por nunca medir esforços pra me ver feliz. E ao meu cachorro Togo por manter minha mente e espírito sempre felizes com suas demonstrações de companheirismo.

Agradecer também aos meus pais que me deram todas as oportunidades necessárias para que eu me tornasse uma pessoa melhor, mesmo que a muito custo. Pelo amor, carinho e afeto tenho eterna gratidão.

Agradeço também aos meus dois grandes orientadores Rafael e Ricardo que fizeram parte de toda essa jornada. Sem a mentoria de vocês, com certeza não seria fácil executar essa etapa ainda mais durante tempos de pandemia. Agradeço a Euge também por ceder um espaço no grupo dela e pelas orientações que me forneceu ao longo da jornada.

Agradeço ao Matheus, Edson, Beatriz e Daniela por testarem a minha ferramenta e me fornecerem os feedbacks necessários para que eu pudesse criar algo realmente útil para meus colegas. Além disso, agradeço de maneira geral a todos os colegas de laboratório, que apesar do pouco tempo, sempre foram muito legais comigo.

Agradeço a CAPES pelo apoio financeiro durante essa jornada (Processo 2017/18934-0). Também agradeço ao Departamento de Biologia Molecular e Celular por todo apoio

.

“When it is obvious that goals can’t be reached, don’t adjust the goals, but adjust the action steps.”

- Confucius

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001”

GENERAL INDEX

RESUMO

Tamasco, Gustavo. ChiMera: Uma ferramenta para Reconstrução, Visualização e Avaliação de Redes Metabólicas Bacterianas. [dissertação]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2022.

Uma série de ferramentas de reconstrução metabólica em escala genômica foram desenvolvidas nas últimas décadas. Essas ferramentas auxiliaram na reconstrução de modelos metabólicos, que contribuíram para uma variedade de campos, por exemplo, engenharia genética, descoberta de drogas e previsão de fenótipos. No entanto, o uso desses programas requer um alto nível de habilidades em bioinformática, e a maioria deles não é escalável para múltiplos genomas. Além disso, as funcionalidades necessárias para a construção de modelos geralmente estão espalhadas por várias ferramentas, exigindo conhecimento de sua utilização. Aqui, apresentamos a ChiMera, que combina as ferramentas mais eficientes em reconstrução, predição e visualização de modelos. A ChiMera usa a abordagem top-down da ferramenta CarveMe, baseada em evidências genômicas, para editar um modelo global com alto nível de curadoria, gerando uma reconstrução preliminar capaz de produzir previsões de crescimento usando análise de fluxo de balanço de metabólitos, tanto para para bactérias gram-positivas como gram-negativas. A ChiMera também contém dois módulos para visualização da rede metabólica. O primeiro módulo gera mapas para as vias mais importantes, por exemplo, metabolismo central, oxidação e biossíntese de ácidos graxos, biossíntese de nucleotídeos e aminoácidos e glicólise. O segundo módulo produz um mapa metabólico de todo o genoma, que pode ser usado para recuperar informações das vias metabólicas usando informações do banco de dados KEGG para cada composto no modelo. Um módulo para investigar a essencialidade e nocaute gênico também está presente.No geral, a ChiMera combina criação de modelo, preenchimento de etapas nas vias metabólicas (*gap-fill)*, análise de balanço de fluxo (FBA) e visualização de rede metabólica para criar um modelo em escala de genoma pronto para simulação, ajudando projetos de engenharia genética, predição de fenótipos e outras descobertas orientadas por modelos. Tudo isso sem exigir alto nível de habilidades de bioinformática.

Palavras-chave: Reconstrução de Metabolismo em Escala Genômica, GSMR, Engenharia Metabólica, Visualização de Metabolismo, Análise de Fluxo de Balanço de Metabólitos.

ABSTRACT

Tamasco, Gustavo. ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization. [dissertação]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2022.

Several genome-scale metabolic reconstruction tools have been developed in the last decades. They have helped to reconstruct various metabolic models, which have contributed to a variety of fields, e.g., genetic engineering, drug discovery, prediction of phenotypes, and other model-driven discoveries. However, the use of these programs requires a high level of bioinformatic skills, and most of them are not scalable for multiple genomes. Moreover, the functionalities required to build models are generally scattered through multiple tools, requiring knowledge of their utilization.Here, we present ChiMera, which combines the most efficient tools in model reconstruction, prediction, and visualization. ChiMera uses CarveMe top-down approach, based on genomic evidence, to prune a global model with a high level of curation, generating a draft genome able to produce growth predictions using flux balance analysis for gram-positive and gram-negative bacteria. ChiMera also contains two modules for metabolic network visualization. The first module generates maps for the most important pathways, e.g., core-metabolism, fatty acid oxidation and biosynthesis, nucleotides and amino acids biosynthesis, and glycolysis. The second module produces a genome-wide metabolic map, which can be used to retrieve KEGG pathway information for each compound in the model. A module to investigate gene essentiality and knockout is also present.Overall, ChiMera combines model creation, gap-filling, flux balance analysis (FBA), and metabolic network visualization to create a simulation-ready genome-scale model, helping genetic engineering projects, prediction of phenotypes, and other model-driven discoveries without requiring high-level of bioinformatic skills.

Keywords: Genome-Scale Metabolic Reconstruction; GSMR; Metabolic Engineering; Metabolic Network Visualization; Flux Balance Analysis.

I. INTRODUCTION

1. Contextualizing Genome Scale Reconstruction

All living organisms are complex creatures of nature, especially when one decides to investigate how the small compartments of life work together and how the organism can survive and thrive in complex and challenging environments. Our knowledge of all biochemical reactions inside metabolism is far from complete, even for the well-studied organisms, making it even more challenging to understand the cellular metabolism.

The building blocks of the cell comprise metabolites, genes, cofactors, enzymes, and other molecules. The interaction of these blocks is responsible for cellular functioning under certain environmental conditions. Therefore, to correctly define the cellular state, the tracking of thousands of compounds, interactions, and their variation in time and space, needs to be done.

To perform such a task, computational power is fundamental. However, even with the advancement in bioinformatics, the lack of biological information, e.g. enzyme kinetics, demands alternative strategies for the correct implementation of simulations (1). The recent advancement in sequencing technologies has provided high-quality genomic information in abundance, under the reduced cost of data acquisition. Genomic data can help researchers to understand the basics of biological components, especially when accompanied by other omics. Multi-omic data can provide a detailed understanding of the dynamicity of cellular components and their behavior inside the cell (2).

In the last decades, a variety of techniques were developed to simulate cellular behavior based on biological data. One of these techniques is called Constraint-Based Modeling, which is a mathematical approach that evaluates the flux inside of the metabolism based on physical, mechanical, and biological constraints. This technique relies on relational information, which compounds are present in the organism and in what reactions they play a role. Hence,

facilitating the overall process of determining the metabolic state of the cell at the cost of accuracy (3). This approach strongly relying on reactions stoichiometry, which is plenty available in multiple databases, such KEGG (4), BiGG (5), Model SEED (6), MetaCyc (7), BREDA (8). These databases play a fundamental role in model creation, providing the foundation knowledge necessary to create metabolic models.

The constraint-based approach compiles biological and biochemical information into a mathematical formulation, enabling a logical evaluation of the organism's metabolism and physiology. The use of this formulation has helped researchers over the last 30 years to understand microbial evolution, network interaction, genetic engineering, drug discovery, prediction of phenotypes, and model-driven discoveries (3).

2. Evolution of DNA sequencing and its contribution to Genome Scale Metabolic Reconstruction (GSMR)

Since the discovery of DNA structure and the development of methods to detect the sequence of DNA, a new era of biology has started. Recent advancements have led sequencing technology to the Era of next-generation sequencing (NGS). Although being more than a decade old, this term still holds due to its correlation to techniques that can produce very high-throughput genomic information (9). The hasty evolution of instruments, techniques, and chemistry supports the application of sequencing technologies on a large scale. Hence, producing an enormous volume of genomic data. Therefore, resulting in new challenges, e.g., the development of bioinformatics approaches that enables the evaluation of such data, producing valuable insights into its content (10).

The advancements in sequencing technologies has also provided a vast amount of data for researchers focusing on metabolic models. Genomic data has supported increasing our knowledge about the molecular process carried by the organism and how the flow from the information contained in the gene is processed until an enzyme catalyzes a reaction. In the last

decades, researchers have compiled our knowledge of molecular processes in databases, which are valuable assets for metabolic model reconstruction (4–8).

3. Biological information Databases

As mentioned before, multiple databases provide biological and biochemical information. However, this information alone doesn't provide all the data necessary to build a metabolic model. The genomic sequence is necessary to elucidate the group of genes present in an organism. Using these databases, we can annotate putative proteins and their participation in metabolic reactions based on the genomic information, a process that is also known as the gene-protein-reaction rule (GPR) (11). Below, some of the most important databases for metabolic modeling are briefly described.

3.1. KEGG

KEGG is a repository of genes and genomes. It provides a curated set of information regarding gene functionality and characterization in different levels of detail. It also provides information on molecular interactions in the form of KEGG pathway maps (more than 963.800 entries), or pathway-specific submodules (4). This information, together with gene composition (more than 42.702.900 entries), can be used to infer the overall structure of the metabolic pathway that an organism harbors (12).

3.2. BiGG

Biochemical, Genetic and Genomic Knowledge database, also known as BiGG database was designed to address an issue in the modeling community, the absence of a repository of curated metabolic models that follows the standards of production. The database has more than 75 highly-curated models, and also provides external links to other databases, allowing researchers to investigate the genomes they originate from. It also keeps track of the nomenclature of each metabolite and reaction across different databases (5).

3.3. MetaCyc

MetacyC is a reference database of metabolic pathways and enzymes from all domains. They have the largest curated collection of metabolic pathways, derived from more than 60 000 publications. This database is used in many of the modeling tools in the literature (13), including Pathway Tools, proprietary software that provides many services ranging from the development of organism-specific databases to comparative genome and pathway analysis. The database contains more than 2900 pathways, 17.700 reactions and 18.100 metabolites (14).

3.4. ModelSEED

The ModelSEED has been designed to help researchers in the construction of genome scale metabolic reconstruction (GSMR), including more than 33.900 compounds and 36.600 reactions. The database also includes a diverse resource of biochemical information, ranging from compartmentalization, transcription reactions and stoichiometry reactions (6). All the hosted information is also integrated with external databases such as KEGG, BiGG and MetaCyc.

3.5. BRENDA

Host a comprehensive relational database that relies on literature experiments. There are more than 40.000 enzymes included in the database, providing information for thousands of organisms. BRENDA is fundamental for metabolic pathway identification in an organism, therefore helping the reconstruction process (8).

Even with the availability of many different databases, there is a problem regarding metabolic reconstruction. Each of them has a singular identifier for each metabolite and reaction. Therefore, increasing the complexity to build a model based on relational information of multiple databases. MetRxn, a knowledgebase was designed to standardize the information

across multiple databases (15). However, the database is no longer maintained. Hence, tools designed to reconstruct metabolic models tend to use a single reference database during the process (13). We will cover this in the next section.

4. Model Creation

4.1. Conversion of genome information into a Draft model

Models can be built using two different perspectives. The first one is the bottom-up approach, which relies upon a compilation of biological information based on genomic evidence. Gene-Protein-Reaction rule is applied to connect annotated genes of an organism to its related proteins and their EC (Enzyme Commission) number, which provides insights about the reactions that the enzymes play a role (Figure 1). Therefore, the database reconstruction represents the organism biochemically and genetically (16).

However, there are downsides to this approach. There is a strong dependence on manual evaluation of the model, e.g., adding missing reactions, correcting elemental balance, detecting wrong fluxes, fixing blocked reactions, and others. All of these problems can result in huge differences during model prediction (17).

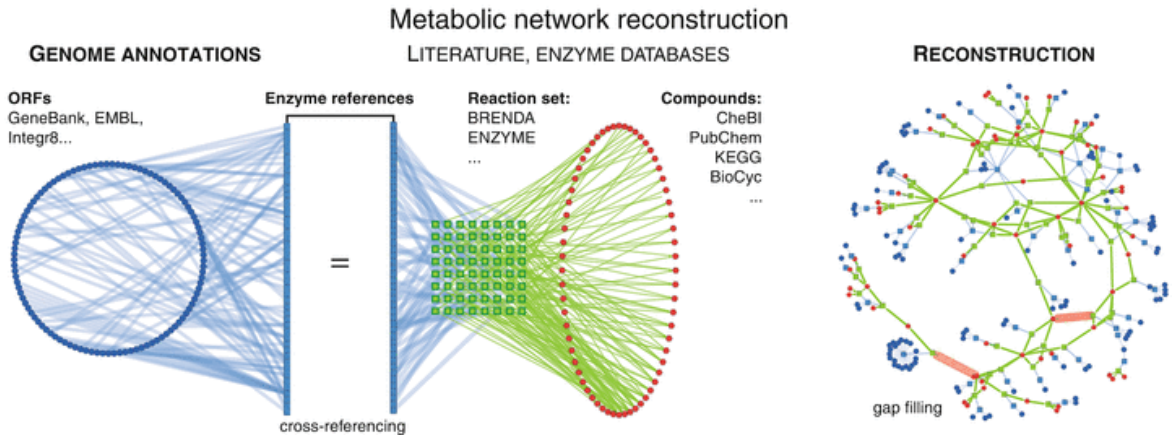


Figure 1: Bottom-up Metabolic Network Reconstruction schematics**.** Genomic evidence is used as a base for integration of EC numbers and reactions to the model. **From:** (18).

Conversely, the second approach uses a top-down mechanism to build GSMR. A curated universal model is used as the basis of the reconstruction. Genomic evidence of the target organism provides evidence for a pruning algorithm that removes reactions with no evidence from an universal model, generating an organism-specific version of it. It's a process independent of context, which infers metabolic paths based on the genomic evidence. The advantages of this approach comprise the addition of manually curated reactions to the model, reducing the necessity of manual evaluation. And also, the production of models ready to perform predictions (Figure 2) (17).

In both approaches, the use of a manually curated database is critical. They help to check important reactions and also provide secure evidence for the metabolic model reconstruction. The use of automated datasets - generated by algorithm predictions - can offer inaccurate information for the final model. Therefore, the use of the most up-to-date databases, KEGG, BIGG, or MetaCyc, can also increase the precision and efficiency of the predictions generated by the model.

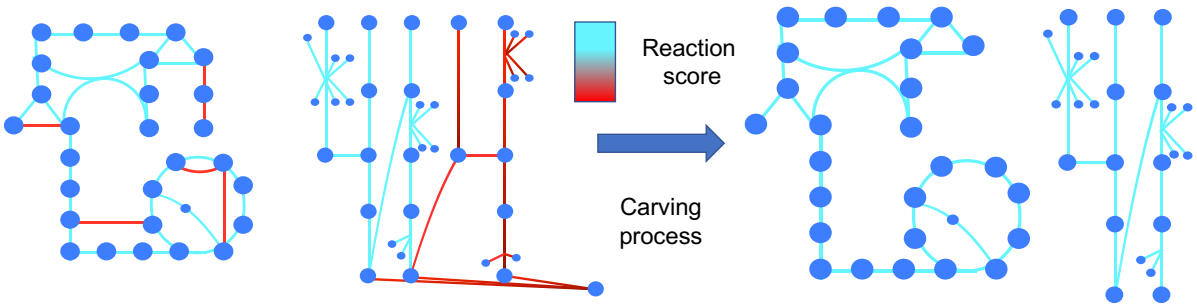


Figure 2: Top-down Metabolic Network Reconstruction. The organism specific model is generated by a pruning process of the universal model. Reaction in red had low evidence detected in the reference genome, being removed during the pruning of the universal model.

5. Genome-Scale Models and the understanding of biological data

Genome-scale metabolic models (GEMs) enable the conversion of biological information into a stoichiometric model, describing the metabolism with linear equations. Hence, enabling the

simulation of the metabolic state of a living organism. The GEMs can identify relationships between genotype and phenotype, using different omics as a source of information (13).

As mentioned before, there are two possible approaches to create the GSMR, and we will give a brief introduction to the most used tools in the field, their advantages and disadvantages. The reconstruction tools differ from each other based on some parameters like the database used in the process, presence of gap-filling module, annotation of transport and exchange reactions, the cellular compartments detected, biomass reaction flexibility, visualization module, programming language and more (Table 1).

Table 1: List of Genome Scale Metabolic Reconstruction Tools and their functionality.

Reconstruction Tool	Mapping Method	Reaction origin	Associated Database	Version	Visualization	Knockout	Type of Software
AureMe	Pantograph (Inparanoid and OrthoMCL)	Template model(s)	BiGG-MetaCyc	1.2.4	No	No	Command line
CarveMe	Diamond, eggNOG-mappera	Template model	BiGG	1.5.1	No	No	Command line
ChiMeRa	Diamond	Template model User defined model	BiGG	1.0	Yes	Yes	Command line
Merlin	Mapping from annotation with BLAST or HMMER	Template model(s)	KEGG	4.0	Yes	Yes	Stand Alone Interface
ModelSEED	Annotation ontology map from RAST data	Template model(s)	ModelSEED	2.2-2.4	Yes	No	Online service
Pathway Tools	Pathologic	Database	MetaCyc	22.0	Yes	Yes*	Stand Alone Interface
RAVEN	Autograph-type method from BLASTP and Bidirectional BLASTPb	Database	KEGG-MEtaCyc	2.0.1	Yes	No	Command line

*means that the tool has dependency on external software to produce the knockouts, in this case Web-MetaFlux Modeling Tool.

5.1. AureMe

AureMe (Automatic Reconstruction of Metabolic Models) is a command-line tool that has good traceability of the reconstruction process. It has compatibility with different databases such as BiGG (5) and MetaCyc (7). This tool is ideal for those that want to keep track of all the steps during the process. However there is no visualization module included. (19).

5.2. CarveMe

CarveMe uses a top-down approach based on a BiGG universally curated model. It also automatically includes the gap-fill of the model, generating ready-to-simulate drafts. Users can also use their own curated model for the reconstruction. This tool also doesn't include visualization (17).

5.3. Merlin

Merlin is an application for genome-scale reconstruction based on the KEGG database. Users can annotate the genome using BLAST or HMMER and the information is used to produce the model. The tool also offers a very flexible customization of parameters, enabling researchers to annotate enzymatic and transport genes, as well as subcellular localization. This tool also has a visualization module that identifies all reactions in the model, helping the process of gap-fill. These characteristics are organized to help a manual curation process (20).

5.4. ModelSEED

ModelSEED is a database that also provides the service of GSMR. The web application is responsible to perform all the steps of creation, from annotation of the genome to the gap-fill based on a user defined media composition. However, they don’t provide any customization of the steps and no visualization module (6).

5.5. Pathway Tools

Pathway Tools is software that allows the user to provide a genome and build an organism-specific database that includes genes, reactions, and metabolites (PGDB - Pathway Genome Database). The database also helps the process of gap-fill. The tool also has a visualization module that enables the observation of whole metabolic network composition (14).

5.6. RAVEN

RAVEN (Reconstruction, Analysis, and Visualization of Metabolic Networks) is a tool developed in MATLAB. They are compatible with KEGG and MetaCyc databases, or one can provide

template models for the reconstruction. The tool also provides information about transport and spontaneous reactions (21).

5.7. ChiMera

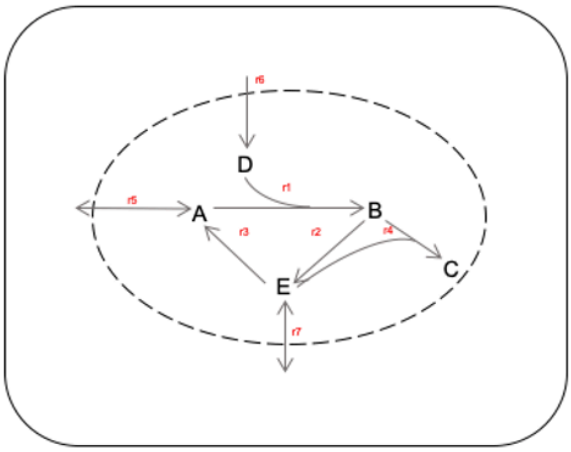
ChiMera is the tool developed in this research. This tool combines model creation, gap-filling, flux balance analysis (FBA), and metabolic network visualization to create a simulation-ready genome-scale model, based on CarveMe (17), COBRApy (22), Escher (23), and PSAMM (24) algorithms. All the steps are done automatically for the user, reducing the need for bioinformatics skills (25).

6. System of Linear equations (Mathematical formulation) behind the GSMR

6.1. From DNA to FBA

Once the draft model of a target organism is reconstructed, we need to convert all biological and biochemical information into a mathematical representation, a stoichiometric matrix (S matrix) (Figure 3). The stoichiometric matrix provides the first level of constraining to the network – it only contains the reactions and compounds supported by the genomic data, therefore reducing the space of possible solutions for the target organism phenotype (26).The compounds are represented in the rows of the S matrix, and the reactions by the columns. In each cell of the matrix is stored the stoichiometric coefficient of the compound in a specific reaction – zeros in the matrix indicate no association with the reaction, negative values indicate consumption, and positive values its production (27). The cellular compartment annotated in the biological draft also provides another level of constraint. It provides a mechanism to identify required compounds inside the cell, secreted substances and compounds consumed from the medium. The most common cell compartments are the cytosol, extracellular matrix and periplasm. However, other compartments as mitochondria, chloroplast, and organelles can also be implemented, to the expanse of manual evaluation and curation (28).

A)



r1: A + D → B
r2: 3B → 1E
r3: E → A
r4: B + E → 2C
r5: ↔ A
r6: → D
r7: ↔ E
Transport reactions

B)

r1	r2	r3	r4	r5	r6	r7
-1	0	1	0	1	0	0
1	-3	0	-1	0	0	0
0	0	0	2	0	0	0
-1	0	0	0	0	1	0
0	1	-1	-1	0	0	1

$v =$

v1
v2
v3
v4
v5
v6
v7

Figure 3: Schematics of biological dataset to a mathematical formulation. (A) Is a toy-model with a few reactions representing a metabolic pathway, here cytosol and extracellular matrix are displayed as main compartments. (B) Stoichiometric matrix based on the compounds and reactions presented in the toy-model. The $\boldsymbol{v}$ represents the flux vector, produced during flux balance analysis to check which reactions are carrying flux in the given condition.

The S-matrix is the basic structure used by many algorithms developed to perform Flux Balance Analysis (FBA) in metabolic models. The process of FBA consists of the conversion of the biological information into a series of linear equations, representing every reaction in the metabolism. The objective function (OF) also plays an important role in the process. The algorithm will use the objective function as the maximization point of the system of equations. As a rule of thumb, the reaction responsible to produce biomass (BOF) is selected as the OF of the process, using the assumption that organisms have evolved to optimize growth (12,27).

There are a number of tools that can be used to perform FBA analysis in GSM. There are commercially available softwares, e.g. MATLAB (29) and IBM ILOG CPLEX (30). There are also Open Source solvers such as Open Source Gnu Linear Programming Lit (GLPK) (31). These solvers are used in third party tools that were developed to apply those FBA algorithms in the metabolic models, e.g., COBRA ToolBox (32), COBRApy (22), ModelSEED (6) and KBase (33).

7. Applied use of Genome-Scale Metabolic Network Reconstruction (GSMR)

Along the last decades, genome-scale metabolic network reconstruction has evolved dramatically, especially due to the advancements in high-throughput techniques, leading to enormous data production. Together with the advancements in computing science and algorithms, omics data are fundamental for model creation. GSMR is built to facilitate the use of biological knowledge to perform computational qualitative and quantitative analysis in the network, helping to answer questions about survival capabilities, fitness, phenotype features, and more (34). *Escherichia coli* is probably the most important organism for the field of model creation and metabolic engineering, studies about the organism's metabolic network started in the early ’90s, using the available information to build a simplified metabolic reconstruction (35). The first model has evolved to different versions in the last two decades. Integration of new data concerning metabolic interactions contributed to new and more precise subsystems inside the model, creating more efficient predictions (36). Further modifications and enhancements to the model are yet to come. Many missing parts of the metabolism are being studied, especially enzyme promiscuity and protein synthesis, transport, and modification. Future integration of new omics data will help to issue these limitations in the current models.

Studies using *E. coli* for GSMR can be divided in six different fields of application: (A) metabolic engineering, (B) model-driven discoveries, (C) prediction of cellular phenotype, (D)

analysis of the network property, (E) evolutionary process and (F) interspecies interaction (Figure 4).

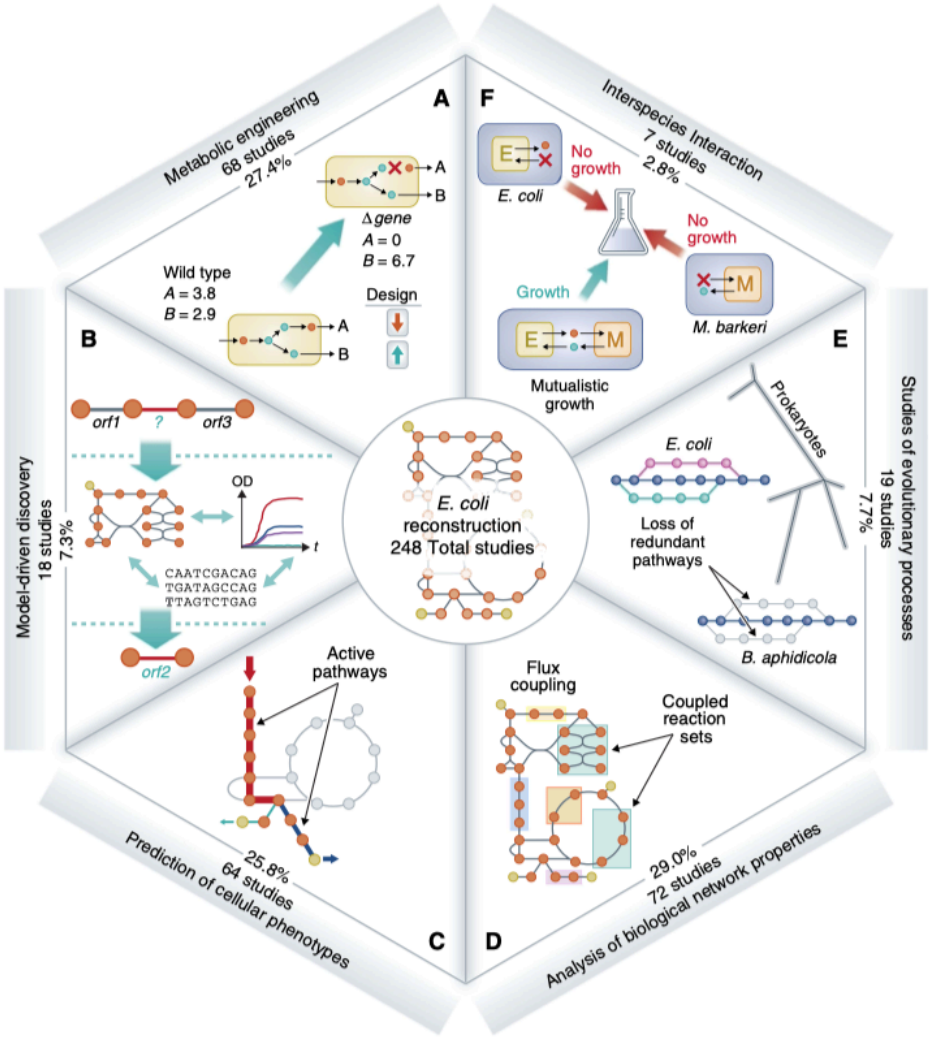


Figure 4: Six categories of uses and number of studies for each use of E. coli GSMR. (A) metabolic engineering, (B) model-driven discoveries, (C) prediction of cellular phenotype, (D) analysis of the network property, (E) evolutionary process and (F) interspecies interaction. From: (37).

7.1. Metabolic engineering

Currently, advancements in GSMR lifted the development of novel and more sustainable approaches to produce relevant compounds in industrial plants, using modified organisms with higher yields. *E. coli* models have a role in this process, guiding metabolic

engineers toward fine-tuning the production of a variety of compounds, including organic acids, enzymes, alcohols, and others (38).

One great example is Optknock, which is a tool that aims to develop knockout strategies in industrial lineages. This computational workflow was used to identify possible gene deletions that could lead to the overproduction of chemical or biochemical compounds in *E. coli*. The engineering strategies through mathematical formulations helps strains to accomplish their theoretical maximum growth, achieved by the maximization of growth in association with minimization of the metabolic adjustment (MOMA). The workflow can suggest multiple solutions to the overproduction, they include classical competing pathways silencing, but also include novel and non-intuitive solutions, both leading to optimal production of a desired compound. Its implementation for overproduction of succinate, lactate, and 1,3-propanediol in *E. coli* produced results in agreement with experimental data (39).

Another great example is OptORF, an approach that tries to minimize no biologically feasible solutions. The integration of metabolic networks and transcriptional regulation allows evaluation of the effects of gene knockout and transcriptional factors overexpression. This approach was tested in *E. coli* focusing on the production of ethanol and higher alcohols (40).

These examples demonstrate not only the capability of model-driven discoveries to study and identify the results of gene knockouts but also to predict engineering strategies that result in the overproduction of specific compounds. Despite great precision, there is still a need for manual curation to evaluate if the results are biologically meaningful.Further complementation of our biochemical and physicochemical knowledge will greatly improve metabolic engineering. There is still little information about expression, post transcriptional and post translational modification. Further development of experimental technologies and algorithms will allow more robust workflows and integration of different databases (41).

7.2. Biological discoveries using model-driven approaches

Even with the current efforts to uncover how the biological system works, there are still a great number of gaps in our knowledge. The implementation of GSMR can lead to new theories about how the interactions inside a metabolic network occur. These hypotheses can be experimentally tested, hence, contributing to the discovery of novel biological phenoms. One great example is a transcriptional regulatory model of *E. coli built* considering genomic information in association with data from the literature, specifically focusing on regulatory genes. The model was able to predict growth phenotype and also depicted gaps in the regulatory network, suggesting a new hypothesis over gene interaction and its effects on expression levels (42).

Model-driven approaches are also applied in the field of drug discoveries (43). The application of GSMR using essentiality analysis (44) leads to the identification of core metabolites associated with cellular growth. The results can be further analyzed, narrowing the list of metabolites. Those not detected, are screened using structural analogs to silence enzymes that consume them, resulting in the "silencing" of associated reactions. This approach was validated in *Vibrio vulnificus*, showing more efficiency than the current antibiotics in use (45).

7.3. Phenotypic functions

The complexity of bacterial metabolism has enabled its survival in a variety of places. The use of GSMR provides valuable insight into the biological processes that take place in different situations. Previous research demonstrated that environmental conditions and transcription regulation have an overall impact in the final phenotype (46).

Optimal phenotype predictions are obtained through the use constraints in the model – due to the use of a system of multiple linear equations, there are more than one possible solution

to phenotype prediction, the use of constraints reduce the variety of results, leading to more meaningful predictions and reducing the possible outcomes – this can accomplish by incorporating information about the compartmentalization of metabolites, thermodynamics, or even through proteins interactions (47,48) .

More recent models tried to implement a probabilistic approach based on gene essentiality to add new constraints. Therefore, increasing the efficiency of phenotype prediction. Results suggest a precision of around 38%, showing that gene essentiality can also play a role in phenotyping. Hence, there is a need for the complementation of other constraints for optimal model precision (49).

7.4. Biological network analysis

Network analysis using GSMR has been considered as an *in silico* only approach. However, recent progress is enabling a more practical application. Despite *E. coli* being one of the most studied organisms, more than one-third of its pathways are not annotated. Recent works used network-assisted analysis to support the identification of novel gene functioning based on network evidence. These approaches have a high predictive capability due to the combination of computational and experimental evidence (50).

The evaluation of the biological network can also be taken in a more mathematical approach, using graph theory to better understand the topology of the network, hence enabling the study of coupled gene deletion, or the metabolic readjustment to compensate for perturbation in the network (2,51).

7.5. Adapting to environmental selective pressure

The complexity of bacterial metabolism has evolved over the years due to the natural selection that has forced organisms to adapt. The evolution comes at the cost of gene loss, indels – adding or deleting a base in a gene –, alterations in expression, or even through horizontal gene transfer. *E. coli* models can help to better understand the process that led to the current species. These models can also identify the course of evolution based on controlled conditions, predicting novel phenotypes giving a list of environmental constraints (50,52).

Researchers evaluated the effects of gene deletion in pathway perturbation – the detour in pathway due to the absence of a gene –. Changes in gene diversity were made under constant environmental selective pressure, resulting in a minimal network, representing the minimal set of genes that enables survival in the environmental condition (53), this information can be applied for organism design in synthetic biology, especially for industrial strain development (54).

7.6. GSMR for microbiomes exploration

The understanding of how the microbial community communicates and keeps its balance in natural environments is one of the most challenging questions in biology. This knowledge can help us in a variety of fields, e.g., understanding how the human gut microbiome works and its impact on our health (55), how host and pathogen cells communicate, and even, how we can manipulate the microbiome of a specific environment to help us to achieve a goal (56) .

An example of how GSMR can be used for environmental purposes is the prediction of optimal strains for bioremediation of heavy metals. The identification of core genes associated with heavy metals conversion can help researchers to optimize their synthesis through model iterations, therefore producing strains with greater capabilities to solve environmental problems (57,58).

As mentioned before, these models can also elucidate the understanding of how the human gut works. GSMR can provide valuable insights toward the minimal media necessary for the microbial community, and how the environment composition can affect the population, resulting in a different phenotype condition. This knowledge can enable more targeted intervention in the gut microbiome composition, restoring or protecting humans against a variety of diseases (59).

8. Future prospects of GEMs

Even with all the advancements in sequencing technologies and computational algorithms to build genome-scale metabolic reconstruction, there is still a gap between *in silico* models and the realistic behavior of cell components and their interaction. Even the most advanced models can't accurately predict all the interactions inside a cell. The next advancement in GSMR technology will come from the combination of multi-omics, structural information, transcriptional networks, and other regulatory systems. All this data will have the potential to closely replicate cell behavior. Nonetheless, the current state of the art of GSMR is far from that. New algorithms capable of integrating multiple data into a single mathematical model are yet to exist, and there is a great chance that machine learning will play a major role in this next revolution (2,13).

II. OBJECTIVES

General Objective

Orchestrate state-of-art tools associated with GSMR, facilitating their use by non-expert users. ChiMera aims to provide valuable information about model and non-model bacteria metabolism in a fast and simple manner.

Specific Objectives:

- Implementation of reconstruction module using CarveMe
- Implementation of flux balance analysis and growth prediction based on Cobrapy
- Implementation of metabolic pathway visualization
    - Automatize Escher with ten predefined maps
    - Uses PSAMM to convert SBML to graph-based structure compatible with Cytoscape
- Implementation of an algorithm that translates BIGG IDs to KEGG IDs and pathway information to the model
- Implementation of the knockout module to perform single/double or complete silencing of reactions and genes

III. RESULTS

ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization

This chapter was published as pre-print:

Tamasco G, da Silva RR, Silva-Rocha R. ChiMera: An easy to use pipeline for Bacterial Genome Based Metabolic Network Reconstruction, Evaluation and Visualization. bioRxiv. 2021 Jan 1.

Current Status: Submitted to BMC

1. Background

Genome-Scale Metabolic Reconstructions (GSMR) are essential tools in System Biology (37). The use of GSMR helped researchers over the last 30 years to understand microbial evolution, network interaction, genetic engineering, drug discovery, prediction of phenotypes, and model-driven discoveries (3). However, the generation of a precise model is very complex and time-consuming, requiring several steps (12). The process starts with genome annotation and assembly of all associated known metabolites and reactions, which creates an initial metabolic reconstruction to build a draft model. Several rounds of manual curation and evaluation of the present genes, reactions, and compounds are necessary to create a high-quality metabolic model. After these steps, one needs to set a Biological Objective Function in the model (e.g., biomass function) followed by the conversion to a mathematical formulation known as Stoichiometric matrix (S-matrix), which is a computer-readable core part of the model. The S-matrix is used to simulate the models performing Flux Balance Analysis (FBA) and growth predictions (12). Other steps, such as gap-filling and stoichiometric balance, may also be necessary, increasing the complexity of the process.

Recently, several tools such as AureMe (19), Pathway Tools (14), RAVEN (21), Model SEED (6)and Merlin (20) were developed to address model creation (60). A few of those tools were designed to handle specific processes. CarveMe (17) is a command-line tool that deals with the initial phase of model creation and gap-filling. Cobrapy (22) can convert draft models into an S-matrix and perform FBA analysis using optimized algorithms. Escher (23) offers a fully customizable suite for pathway visualization. However, these tools require familiarity with command-line interfaces and programming (13). They also have their peculiarities, demanding time and knowledge from users to perform the analysis. Therefore, the use of those tools by non-bioinformatics can be challenging, and the number of steps required to build initial models precludes their usage in large-scale projects, which may include hundreds of genomes.

Here, we present a novel tool named ChiMera, which compiles the most used tools for genome-scale metabolic modeling in a pipeline that does not require programming skills. ChiMera uses a genome as

input (*.faa file), performing the model creation based on a highly curated universal model (17). The resulting draft model is used for FBA and growth predictions, knockout simulations, and pathway visualization (Figure 5). To evaluate ChiMera, we compared several aspects of model completion with manually curated models from the literature. We also compared the predicted growth values with experimental data to ensure the production of realistic values.

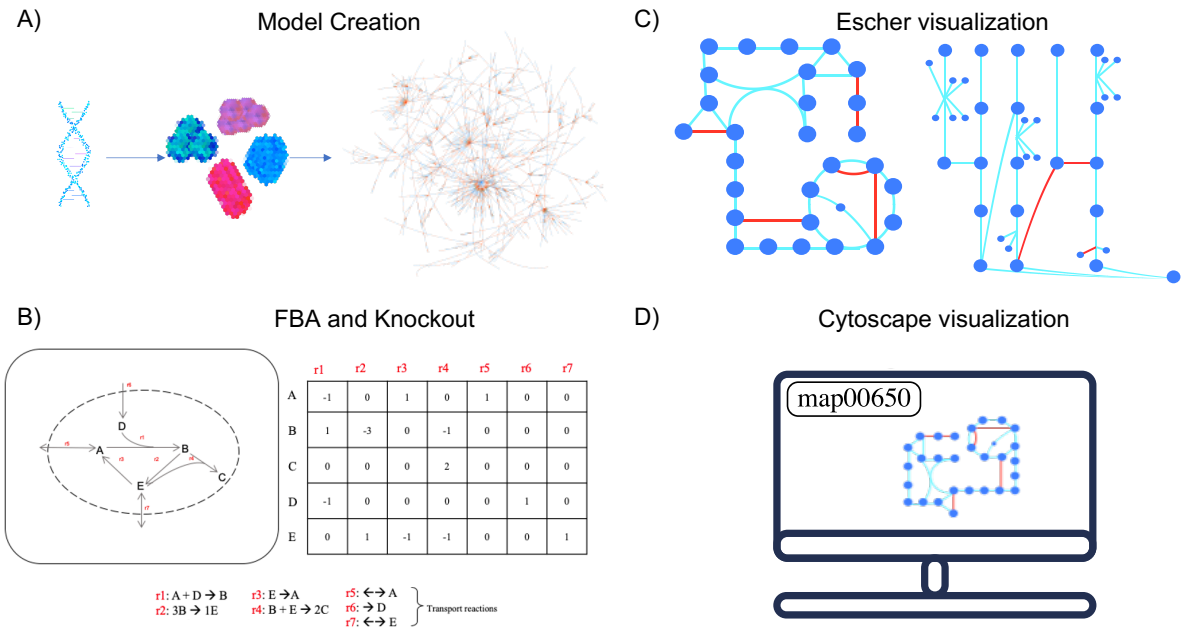


Figure 5: Schematic of processes automated by ChiMera. We use automation to compile several tasks and create an organism specific model based on the genomic data. A) Model creation based on CarveMe pruning algorithm, users just need to provide a protein annotation file based on the genome, B) Model conversion to a Stoichiometric Matrix to perform Flux Balance Analysis and Gene and Reaction Knockouts. C) Representation of pre-defined Escher maps, reactions depicted in blue were detected in the model, red ones were absent. D) Users can use the Cytoscape search bar to select specific pathways.

2. Material and Methods

General ChiMera structure.

ChiMera uses automation algorithms to combine three main steps in GSMR, model creation and gap-filling, FBA, and pathway visualization. Along with a knockout prediction module based on FBA, which enables gene essentiality evaluation. The tool has a modular design compatible with further module expansions (Figure 6).

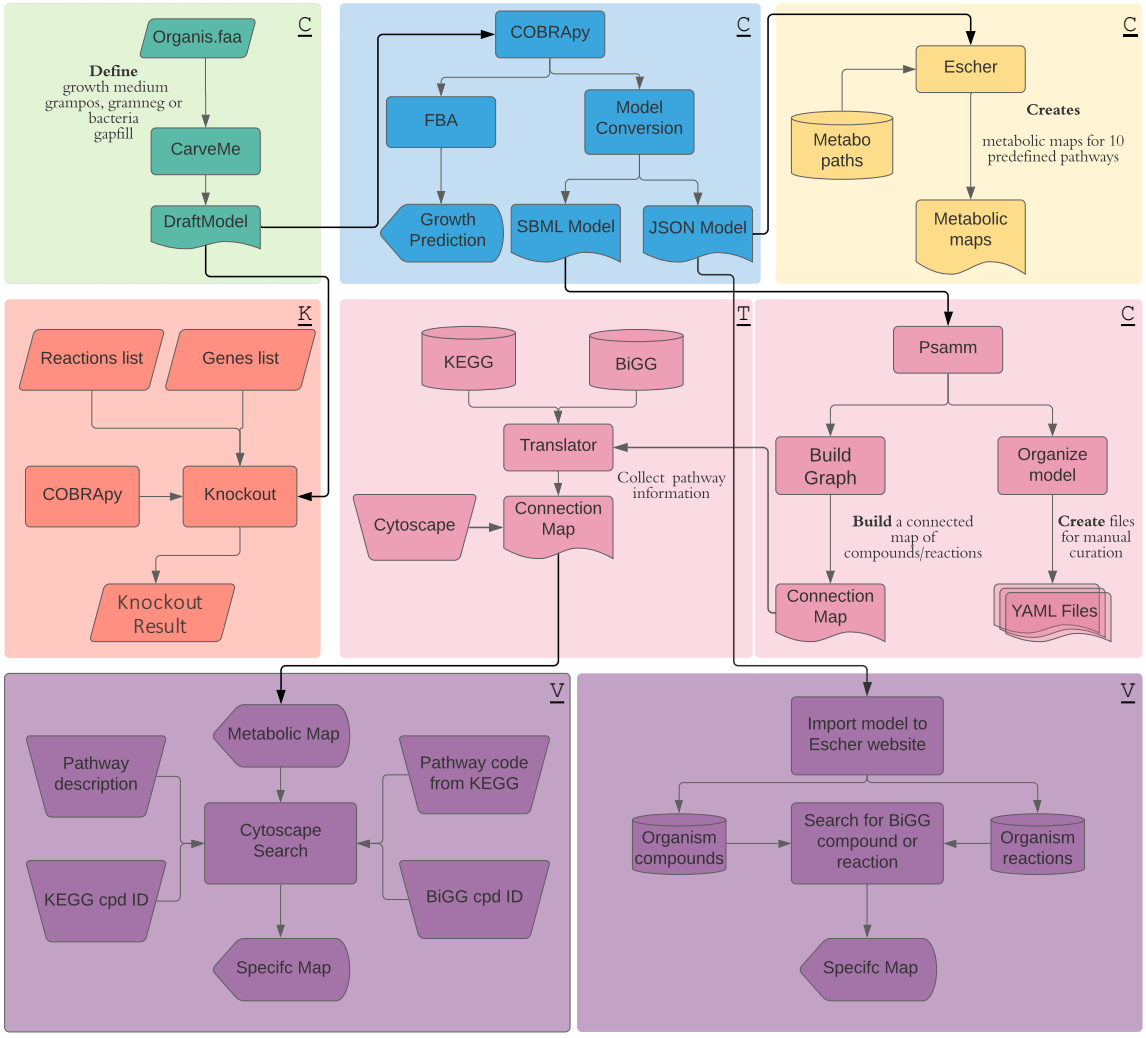


Figure 6: Flow chart of ChiMera processes. ChiMera has 3 submodules that can be used separately. The ones signed with "C" are part of the core module, which performs model creation, evaluation and creation of visualization files. The "T" represents the translator module that adds KEEG pathway information to the compounds in the edges file. The file can be loaded into Cytoscape or Gephi for visualization. The "K" represents the knockout module, which performs gene and reaction knockouts. "V" represents the use of outputs created by ChiMera to create custom maps by third-party tools based on user needs.

Model Creation and gap-filling

We automated the utilization of CarveMe (v1.5.1) in the reconstruction module of ChiMera. The initial draft model is created based on the protein file (*.faa file) provided by the user. During the reconstruction process, ChiMera also performs a gap-filling based on the medium definition, using genomic evidence to ensure that the model will produce growth under the given conditions. CarveMe uses a top-to-bottom approach in a pre-built reference and manually curated universal model . It applies a pruning algorithm that removes reactions not supported by the genomic evidence, generating an organism-specific model based on highly curated data (17). CarveMe comes with five predefined media: LB, anaerobic LB, M9, anaerobic M9, and M9 using glycerol as a carbon source. We added a new submodule to ChiMera that enables the user to update CarveMe with any desired media. Users need to provide the new media composition as a tab-separated file (Supplementary Table 1). After the media database update, users can use the new media in the model reconstruction.

S-matrix construction and initial FBA

We used COBRApy (v0.22.1) to convert the initial draft into an S-matrix and perform an FBA analysis (22). ChiMera uses COBRApy in the main module. Growth, uptake, and secretion metrics are displayed in the command line for the user. The tool is also used in the knockout module, enabling users to perform targeted single or double gene/reaction knockout. A file with the gene name or reaction name needs to be provided by the user (Supplementary Table 2). We also include an option to perform a single gene/reaction knockout in the whole model.

Visualization of the metabolic maps

ChiMera converts the initial XML model to 3 different model formats: SBML, JSON, and YAML. These model formats are compatible with the majority of the tools in the literature. We perform transformations in the JSON model to enable compatibility between Escher maps and the user model. We developed *in-house* algorithms to automate the generation of metabolic maps based on Escher (v1.7.3) (23). Ten predefined pathway maps are pre-loaded in this module. Users can also provide

custom JSON maps of desired pathways to check if they are present in the target organism. A video demonstration is provided for users' benefit to understand how to add new Escher maps to ChiMera (61,62). The pipeline uses the model data to evaluate reactions and compounds present in the organism, creating customizable HTML maps that can be edited by the user.

We also developed a second visualization module that creates files compatible with Cytoscape, Gephi, and other graphical visualization tools. ChiMera automated the use of PSAMM (v1.1.2), converting the model to a graphical representation (24). The graphical representation only contains information about the connection between nodes (compounds) and edges (reactions). Users can use the "harvest path" submodule to convert the BiGG ids to KEGG ids. This submodule also collects information on the pathways that the compounds participate in. This approach creates a graphical representation file with pathway information that can be loaded into Cytoscape (63). The pathway information can be used to select specific maps from the whole network (64).

Genome Selection

For demonstrating the functions of ChiMera, we selected two well-studied Gram-negative (*Pseudomonas putida* KT2440 and *Escherichia coli*) and one Gram-positive bacteria (*Bacillus subtilis*). Protein sequence files were downloaded from NCBI under the accession numbers NC_002947.4, NZ_CP020543.1, and AL009126.3. These genomes were used to generate models from ChiMera. Further, these ChiMera models were compared with manually curated models from the BiGG database, iJN1463 (*P. putida*), iEC1344_C (*E. coli*), and iYO844 (*B. subtilis*).

Model evaluation

We performed basic tests to check the correctness of the models produced by ChiMera using MEMOTE, which benchmarks the model using consensus tests based on model annotation, biomass composition, network topology, stoichiometry, and biomass composition and consistency (65). We also performed a gene essentiality benchmark to assay the effect of a single-gene deletion. The media composition was defined as M9 minimal medium for all the organisms. To calculate the performance metrics we measured the ChimMera's ability to correctly assign a gene as non-essential or essential. Predicted

outcomes were compared to the curated models (Supplementary Table 3). Experimental data was used to evaluate the predictions (36,66,67). To examine the prediction capabilities of ChiMera models, we simulated models using different carbon sources that were previously experimentally tested in the laboratory for growing *B. subtilis*, *E. coli*, and *P. putida* (68–73). Except for the carbon source, the update rates of other nutrients were kept constant in each simulation. Each carbon source was constrained using lower and upper bounds -10 and 0, respectively. The list of carbon sources can be seen in Supplementary Table 4, 5, and 6.

Performance metrics

We used 6 different performance metrics to compare the genes essentiality predictions from ChiMera to highly curated models.

Precision: TP/(TP+FP)
Sensitivity: TP/(TP+FN)
Specificity: TN/(TN+FP)
Accuracy: (TP+TN)/(TP+FP+FN+TN)
Negative Predictive Value(NVP): TN/(TN+FN)
F score:2 * ((Precision *Sensitivity)/(Precision+Sensitivity))

Where TP = True Positive, TN = True Negative, FP = False positive and FN = False Negative predictions.

ChiMera environment and user interface

ChiMera is a portable command-line-based tool. The source code, along with complete documentation of its utilization and examples of inputs are available (https://github.com/tamascogustavo/chimera, https://sourceforge.net/projects/chimera-gsmr/) (25).

3. Results

Key Capabilities of ChiMera

ChiMera was implemented in python v3.7, and its dependencies are freely available. There are four main functions of ChiMera: model creation, flux balance analysis and growth prediction, metabolism visualization, and knockout evaluation. ChiMera relies on CarveMe to create an organism-specific model. A curated model is pruned to produce a draft model containing thermodynamic balanced reactions and elemental balanced metabolites, which interact across three compartments, cytosol, periplasm, and extracellular space. During the reconstruction, the user can select one of the five predefined media, or one can build a specific media composition, and perform a gap-filling based on the genomic evidence to ensure that the organism-specific model can grow on the provided or experimentally-tested growth conditions. If the model is not able to grow in the given medium, a message is displayed, informing that the gap-fill has failed to enable growth. To ensure that the model can grow in the defined media, ensure that all the necessary nutrients are present in the media. We recommend the use of M9 minimal media as a base for the creation of a novel media, avoiding missing precursors.

Next, the organism-specific model is automatically converted to a S-matix, using COBRApy. The biomass-producing reaction, which contains the precursors like carbohydrate, protein, lipids, and energy molecules balanced for producing one gram of biomass, is set as the Biological Objective Function for performing FBA. The fluxes of uptake and secretion based on the media, along with growth value are displayed to the user (Supplementary Figure 1). Subsequently, the model is converted to a JSON format, used to produce predefined metabolic maps based on another tool, Escher. These maps include carbohydrate, central carbon, fatty acid oxidation and biosynthesis, glycolysis, inositol, and tryptophan metabolism (Supplementary Figure 2). However, users can design specific maps and add them to ChiMera pipeline (Supplementary Figure 3). The model is also converted to YAML format, which is used by PSAMM *findprimalpairs* algorithm to break down the GSMR into connections between metabolites (nodes) and reactions (edges).

The output of PSAMM can be directly loaded into Cytoscape, producing a visualization of the entire reconstruction. Users can also use the ChiMera translator submodule, to add pathway information to the file, enabling a targeted search of pathways in Cytoscape (Supplementary Figure 4).

To allow ChiMera's flexibility and modularity, users can also provide a pre-built model with the protein file, which should hold the same prefix, directly performing FBA analysis and construction of the pathway maps. Documentation is provided to ensure that the annotations of the model or the presence of extra compartments are compatible with PSAMM, to generate the Cytoscape compatible file. We also provide tutorials on how to use ChiMera output files to build custom maps for any organism.

The knockout module of ChiMera is dependent on COBRApy. Here, we implemented a function that enables the user to provide a file (.txt) containing a list of genes or reactions to be silenced. This module can perform single or double targeted deletions. Results are displayed in the command line for the user (Supplementary Figure 5).

The user can also perform gene essentiality analysis for the whole model, identifying the impact of silencing the genes/reactions on the growth under given growth conditions (Supplementary Table 7).

Comparison with manually curated models

We compared sets of metabolites and reactions included in ChiMera models with those present in manually curated models. iChiMera1182 (*B. subtilis*) shared 50% of its metabolites and 44% of its reactions with iYO844. iChiMera1657 (*E. coli*) and iEC1344_C models along with ChiMera1716 (*P. putida*) and iJN1463 contain similar features. In both comparisons, models shared 68% of their metabolites and 60% of their reactions (Figure 7A).

We also performed a more comprehensive comparison of model features based on MEMOTE metrics (65). The overall score of ChiMera models is comparable to the manually curated models. Moreover, ChiMera has a lower number of blocked reactions, orphan and dead-end metabolites. Curated models had a higher presence of missing essential precursors in the Biomass Function, which can lead to unrealistic growth predictions (Table 2).

Table 2: MEMOTE evaluation metrics.Parameters that can influence the precision of the predictions were selected to assay Chimera and BiGG curated models. Values in the range of 1 ± 10^{-3} in Biomass Constitution are necessary to indicate a realistic biomass function.

Model ID	Balance Metrics			Biomass Constitution		Network Topology		
	Stoichiometric	Mass	Charge	Biomass Constitution	Missing precursors in Biomass	Blocked Reactions	Orphan Metabolites	Dead-end Metabolites
P. putida **iChiMera1716**	**99.8**	**99.9**	**80.2**	**1.00**	**1**	**19**	**0**	**0**
iJN1463	**0**	**99.6**	**99.7**	**0.98**	**2**	**247**	**56**	**85**
E. coli **iChiMera1657**	**99.6**	**99.9**	**83.2**	**1.00**	**1**	**18**	**0**	**1**
iEC1344_C	**100**	**100**	**74**	**1.54**	**30**	**0**	**0**	**1**
B. subtilis **iChiMera1182**	**100.0**	**99.9**	**82.1**	**1.03**	**1**	**54**	**1**	**1**
iYO844	**100.0**	**94.4**	**98.9**	**1.04**	**6**	**50**	**122**	**21**

Next, we also examined the prediction capabilities of ChiMera models by comparing the predicted growth with experimentally measured growth rates. The prediction capabilities of ChiMera models were also compared with manually curated models. Both sets of models were simulated using 46, 50, and 70 different carbon sources for *B. subtilis* (27,28), *E. coli* (23,24), and *P. putida* models (25), respectively (Supplementary Table 4, 5 and 6). This analysis suggested that ChiMera models were able to perform comparably to manually-curated models. In comparison with manually-curated models, ChiMera models predicted 96 to 100% accurate growth on different carbon sources (Figure 7B).

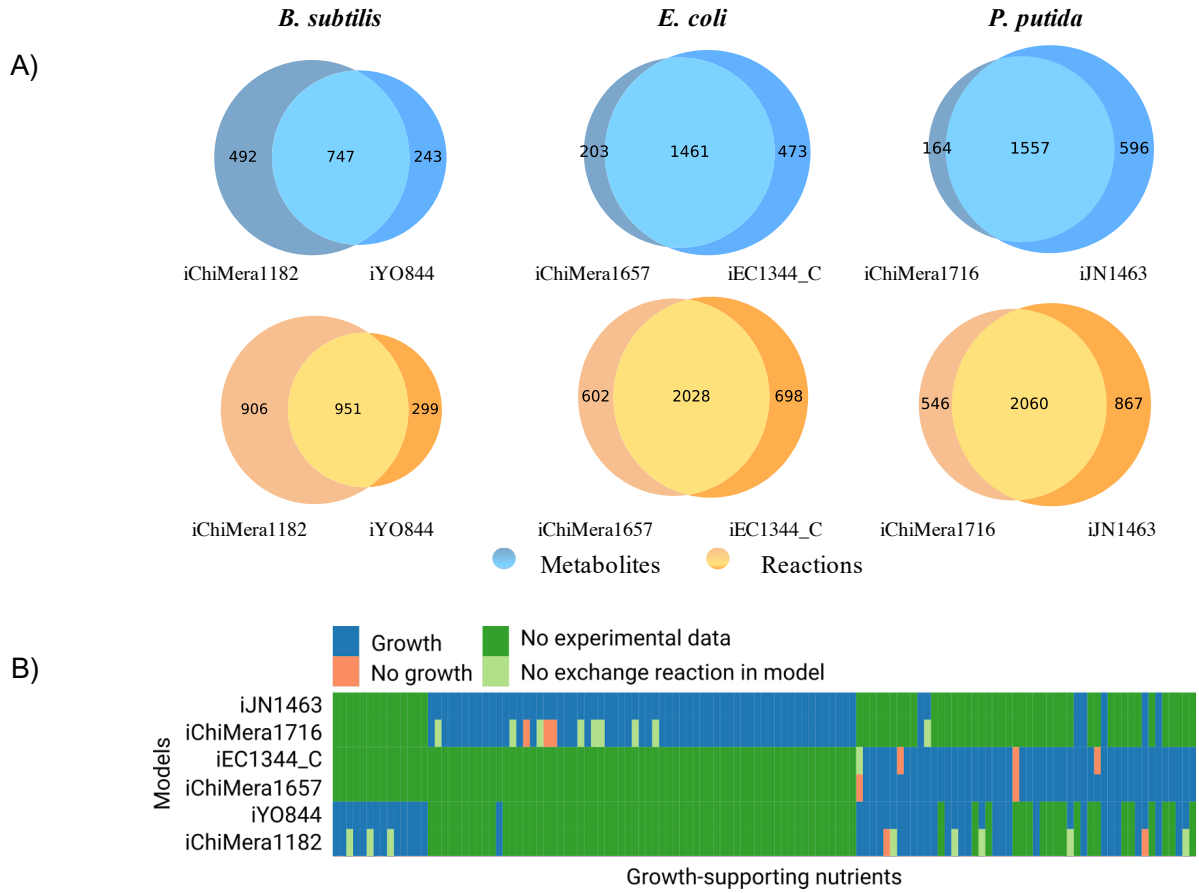


Figure 7: Comparison between ChiMera and manually curated models of *B. subtilis*, *E. coli*, and *P. putida*. A) Venn diagram of reactions and metabolites sets. Reactions and compound sets from ChiMera (iChiMera1182, iChiMera1657, and iChiMera1716) and manually curated models (iYO844, iEC1344_C, and iJN1463) were compared to identify the intersection of the model features. Model-specific information is also depicted. B) Heatmap illustrates predicted growth using ChiMera and manually curated models on different experimentally tested growth environments. The models of *B. subtilis*, *E. coli*, and *P. putida* were used to simulate growth on 46, 50, and 70 carbon sources, respectively. These carbon sources and *in silico* growth rates can be seen in Supplementary Table 4, 5, and 6. The data in heatmap was clustered based on rows and columns.

Gene essentiality metric evaluation

Before we evaluate the predictions of each model, gene datasets for each organism were normalized based on the weighted average of hits in the model (Figure 8B). Model performance statistics were calculated by the ability to detect essential genes and non-essential genes, respectively (Supplementary Table 3).

The gene essentiality predictive metrics were higher in manually curated models. For *E. coli*, the iEC1344_C had a perfect prediction on the dataset. The iChiMera1657 model was outperformed in sensitivity, negative predictive value, accuracy, and F1 score. For *B. subtilis*, we observed a better performance at specificity, negative predictive value, accuracy, and F1 score for iYO844 (Figure 8A).

When comparing *P. putida* models, iChiMera1716 and iJN1463, we observed that the curated model had worse specificity and precision. On the contrary, they had better performance at the sensitivity and negative predictive value (Figure 8A). ChiMera's models were outperformed in sensitivity and negative predictive value in all the comparisons. Metadata indicates that our models had a higher mislabeling of essential genes (Supplementary Table 4).

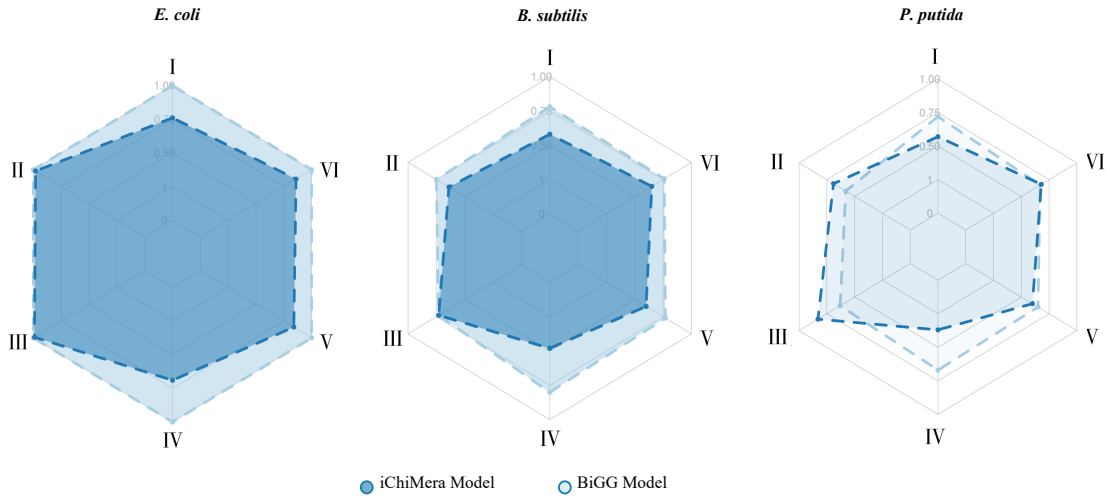


Figure 8: Gene essentiality metrics. Six metrics were selected to compare prediction capability from ChiMera and manually curated models. A) Radarplot of Gene Essentiality metrics. Non-essential genes were used as true positive and essential genes as true negative. B) Stacked bar plot of gene essentiality classification according to presence in the genome.

4. Discussion

We introduce ChiMera, an automated, well-documented and easy-to-use command-line tool that enables researchers with less or no knowledge of bioinformatics and computational biology to produce Genome-Scale Metabolic Reconstructions. These reconstructions can be great tools to explore the metabolic potential of the target organisms. Gene essentiality modules within ChiMera can help researchers to understand the behavior of the organisms under diverse experimental conditions. The visualization modules facilitate the exploration of essential pathways, as well as the identification of unique pathways for non-model organisms. Collectively, the outcomes provided by ChiMera assists researchers in understanding non-model organism metabolism and even developing metabolic engineering approaches for model organisms.

ChiMera has similar ambitions to AuReMe and Merlin. These tools offer a custom workspace for the user, hence facilitating the construction of Genome-Scale Models. AuReMe has its own data structure based on PADMet, and focuses on traceability of the reconstruction process, performing at its best if highly curated models are available (68). There are several steps that the user can process, but it lacks visualization and knockout modules (Supplementary Table 8). Its performance was comparable to CarveMe in model creation (17). Merlin offers a vast workspace for its users. Its graphical interface allows users to re-annotate genomes using BLAST or HMMER, and also integrate data from NCBI and KEGG to its draft model (20,74). This tool is preferable for those focusing on manual curation of single organisms with expertise in metabolic engineering and model creation (60).

ChiMera inherits some pros and cons from CarveMe. Generating networks that share coverage of reactions and metabolites above 60% compared to highly curated models (Figure 3A). Which shows great potential for the first model draft, prior to manual curation. The ready to simulate models of ChiMera are also valuable assets for those working with hundreds of genomes due to the easiness and speed of a draft construction, enabling researchers to evaluate multiple candidate models and choosing the best option for a manual curation if needed. Here we demonstrated the ChiMera models can predict

organism's phenotypes comparable to manually curated models (Figure 3B). We also observed good agreement in gene essentiality detection. Manually curated models mostly had higher prediction capabilities compared to ChiMera models (Figure 4A). However, the differences were more accentuated for sensibility and negative predictive value where the metrics between ChiMera models and curated ones agreed 76% and 61% respectively. For accuracy, ChiMera achieved 84% of the curated model prediction. Specificity and precision metrics were similar, with marginal advantage to ChiMera predictions. These inferences are held with a F-score of 91%. These results show how ChiMera can be used to provide initial models, speeding the manual curation process.

ChiMera complements the reconstruction module based on CarveMe, by adding a new visualization module that allows the user to have a comprehensive overview of the organism's metabolism. One can rely on the predefined maps or design specific maps using ChiMera outputs to suit their research needs. We also provide models in different formats that enable compatibility with most of the tools used to create GSMR. We further complemented CarveMe by adding a new module to ChiMera that enables users to define any custom media during the reconstruction process. Users only need to create a file with the media composition, and it will be automatically loaded into the CarveMe media database. Once the database is updated, the new media can be used in any further reconstruction, enabling a more customized model creation.

Finally, the FBA and knockout modules can help to elucidate ecological niches and the planning of knockout strategies. These modules can also perform pathway engineering, identifying the best silencing strategies to deflect the metabolic flux to the desired metabolite. ChiMera archives all these functionalities in a modular and easy-to-use pipeline.

5. Conclusion

ChiMera is a novel command-line tool that automatizes the usage of state-of-art GSMR tools, enabling biologists with no prior experience in model reconstruction to create ready-to-simulate genome-scale models. ChiMera contains submodules that enable users to investigate the metabolic pathways present in the target organism. And also perform gene or reaction knockout, facilitating the development of engineering strategies. To demonstrate the benefits of ChiMera, we compared gene essentiality and growth prediction capability of ChiMera models against curated models. As a result, ChiMera provides automatization of a unique set of tools, for biologists who are interested in Genome-Scale Models as well as for those interested in a more comprehensive understanding of an organism's metabolism.

IV. GENERAL CONCLUSION

The results of this dissertation have given rise to the following conclusions:

1. ChiMera is a novel command-line tool that automatizes the usage of state-of-art GSMR tools, making it easier for researchers to build their own model based on any genome.
2. ChiMera contains submodules that enable users to investigate the metabolic pathways present in the target organism, and also perform gene or reaction knockout, facilitating the development of engineering strategies.
3. ChiMera capabilities are on pair with curated models

Altogether, the current work has provided to researchers a new tool that can be used to evaluate sequenced organisms. Providing a deep understanding of their metabolic pathways and how they may interact with the surrounding environment.

V. REFERENCES

1. Smallbone K, Simeonidis E, Broomhead DS, Kell DB. Something from nothing − bridging the gap between constraint-based and kinetic modelling. FEBS J. 2007;274(21):5576–85.
2. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol Syst Biol. 2013;9:661.
3. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet. 2014 Feb;15(2):107–20.
4. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.
5. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016 Jan 4;44(D1):D515-522.
6. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Res. 2021 Jan 8;49(D1):D1555.
7. 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 Jan 4;44(D1):D471-480.
8. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, et al. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 2011 Jan;39(Database issue):D670-676.
9. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. Annu Rev Genomics Hum Genet. 2016;17(1):95–115.
10. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31–46.
11. Filippo MD, Damiani C, Pescini D. GPRuler: Metabolic gene-protein-reaction rules automatic reconstruction. PLOS Comput Biol. 2021 Nov 8;17(11):e1009550.
12. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 2010 Jan;5(1):93–121.
13. Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data. Metabolites. 2022 Jan;12(1):14.
14. Karp PD, Paley SM, Midford PE, Krummenacker M, Billington R, Kothari A, et al. Pathway Tools version 24.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. ArXiv151003964 Q-Bio [Internet]. 2020 Nov 12 [cited 2022 Feb 7]; Available from: http://arxiv.org/abs/1510.03964
15. Kumar A, Suthers PF, Maranas CD. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics. 2012 Jan 10;13(1):6.
16. O'Brien EJ, Monk JM, Palsson BO. Using Genome-scale Models to Predict Biological Capabilities. Cell. 2015 May 21;161(5):971–87.
17. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 2018 Sep 6;46(15):7542–53.
18. Chalancon G, Kruse K, Babu MM. Metabolic Networks, Reconstruction. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. Encyclopedia of Systems Biology [Internet]. New York, NY: Springer; 2013 [cited 2022 Mar 21]. p. 1259–63. Available from: https://doi.org/10.1007/978-1-4419-9863-7_1238
19. Aite M, Chevallier M, Frioux C, Trottier C, Got J, Cortés MP, et al. Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale

metabolic models. PLoS Comput Biol. 2018 May 23;14(5):e1006146.
20. Dias O, Rocha M, Ferreira EC, Rocha I. Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Res. 2015 Apr 30;43(8):3899–910.
21. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J. The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput Biol. 2013;9(3):e1002980.
22. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013 Aug 8;7:74.
23. King ZA, Drager A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLoS Comput Biol. 2015 Aug;11(8):e1004321.
24. Steffensen JL, Dufault-Thompson K, Zhang Y. PSAMM: A Portable System for the Analysis of Metabolic Models. Dandekar T, editor. PLOS Comput Biol. 2016 Feb 1;12(2):e1004732.
25. Tamasco G. ChiMera: An easy to use pipeline for Genome based Metabolic Network reconstruction, evaluation and visualization [Internet]. Available from: https://doi.org/10.5281/zenodo.5720515
26. Raškevičius V, Mikalayeva V, Antanavičiūtė I, Ceslevičienė I, Skeberdis VA, Kairys V, et al. Genome scale metabolic models as tools for drug design and personalized medicine. PLoS ONE. 2018 Jan 5;13(1):e0190636.
27. Cuevas DA, Edirisinghe J, Henry CS, Overbeek R, O'Connell TG, Edwards RA. From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model. Front Microbiol [Internet]. 2016 [cited 2022 Mar 22];7. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2016.00907
28. Seaver SMD, Gerdes S, Frelin O, Lerma-Ortiz C, Bradbury LMT, Zallot R, et al. High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. Proc Natl Acad Sci. 2014 Jul;111(26):9645–50.
29. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction - PubMed [Internet]. [cited 2022 Mar 23]. Available from: https://pubmed.ncbi.nlm.nih.gov/24084808/
30. IBM ILOG. Cplex Optimization Studio. [Internet]. 2014. Available from: http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer.
31. MAKHORIN A. GLPK (GNU Linear Programming Kit). http://www.gnu.org/s/glpk/glpk.html [Internet]. 2008 [cited 2022 Mar 23]; Available from: https://ci.nii.ac.jp/naid/10031083954/
32. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0 | Nature Protocols [Internet]. [cited 2022 Mar 23]. Available from: https://www.nature.com/articles/s41596-018-0098-2
33. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED. In: Alper HS, editor. Systems Metabolic Engineering: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2022 Mar 23]. p. 17–45. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-62703-299-5_2
34. Palsson BØ. Systems biology: properties of reconstructed networks. Cambridge university press; 2006.
35. Majewski RA, Domach MM. Simple constrained-optimization view of acetate overflow in E. coli. Biotechnol Bioeng. 1990 Mar 25;35(7):732–8.
36. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. Mol Syst Biol. 2011 Oct 11;7:535.
37. Monk J, Nogales J, Palsson BO. Optimizing genome-scale network reconstructions. Nat Biotechnol. 2014 May;32(5):447–52.
38. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol. 2012 Jun;8(6):536–46.

39. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 2003 Dec 20;84(6):647–57.
40. Kim J, Reed JL. OptORF: Research article Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. 2010;19.
41. Bernstein DB, Sulheim S, Almaas E, Segrè D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. Genome Biol. 2021 Feb 18;22(1):64.
42. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature. 2004 May 6;429(6987):92–6.
43. Oberhardt MA, Yizhak K, Ruppin E. Metabolically re-modeling the drug pipeline. Curr Opin Pharmacol. 2013 Oct;13(5):778–85.
44. Kim P-J, Lee D-Y, Kim TY, Lee KH, Jeong H, Lee SY, et al. Metabolite essentiality elucidates robustness of Escherichia coli metabolism. Proc Natl Acad Sci U S A. 2007 Aug 21;104(34):13638–42.
45. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, et al. Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. Mol Syst Biol. 2011 Jan 18;7:460.
46. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol Syst Biol. 2007 Apr 17;3:101.
47. Ederer M, Gilles ED. Thermodynamically Feasible Kinetic Models of Reaction Networks. Biophys J. 2007 Mar 15;92(6):1846–57.
48. Pe´rez-Bercof A, Aoife M, Gavin CC. Patterns of indirect protein interactions suggest a spatial organization to metabolism. 2011;
49. Galardini M, Koumoutsi A, Herrera-Dominguez L, Cordero Varela JA, Telzerow A, Wagih O, et al. Phenotype inference in an Escherichia coli strain panel. eLife. 2017;6:e31035.
50. Kim H, Shim JE, Shin J, Lee I. EcoliNet: a database of cofunctional gene network for Escherichia coli. Database J Biol Databases Curation. 2015 Feb 2;2015:bav001.
51. Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. BioData Min. 2011 Apr 28;4(1):10.
52. Pang TY, Lercher MJ. Each of 3,323 metabolic innovations in the evolution of E. coli arose through the horizontal transfer of a single DNA segment. Proc Natl Acad Sci U S A. 2019 Jan 2;116(1):187–92.
53. Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature. 2006 Mar 30;440(7084):667–70.
54. Smith KM, Liao JC. An evolutionary strategy for isobutanol production strain development in Escherichia coli. Metab Eng. 2011 Nov;13(6):674–81.
55. Walter J, Ley R. The human gut microbiome: ecology and recent evolutionary changes. Annu Rev Microbiol. 2011;65:411–29.
56. Lebeis SL, Kalman D. Aligning antimicrobial drug discovery with complex and redundant host-pathogen interactions. Cell Host Microbe. 2009 Feb 19;5(2):114–22.
57. Al Hasin A, Gurman SJ, Murphy LM, Perry A, Smith TJ, Gardiner PHE. Remediation of chromium(VI) by a methane-oxidizing bacterium. Environ Sci Technol. 2010 Jan 1;44(1):400–5.
58. Singh JS, Abhilash PC, Singh HB, Singh RP, Singh DP. Genetically engineered bacteria: an emerging tool for environmental remediation and future research perspectives. Gene. 2011 Jul 1;480(1–2):1–9.
59. van der Ark KCH, van Heck RGA, Martins Dos Santos VAP, Belzer C, de Vos WM. More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. Microbiome. 2017 Jul 14;5(1):78.
60. Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biol. 2019 Aug 7;20(1):158.
61. Tamasco G. Build your own metabolic map with Chimera outputs [Internet]. [cited 2022 Jan 7]. Available from: https://www.youtube.com/watch?v=XQRbSkvMpN4

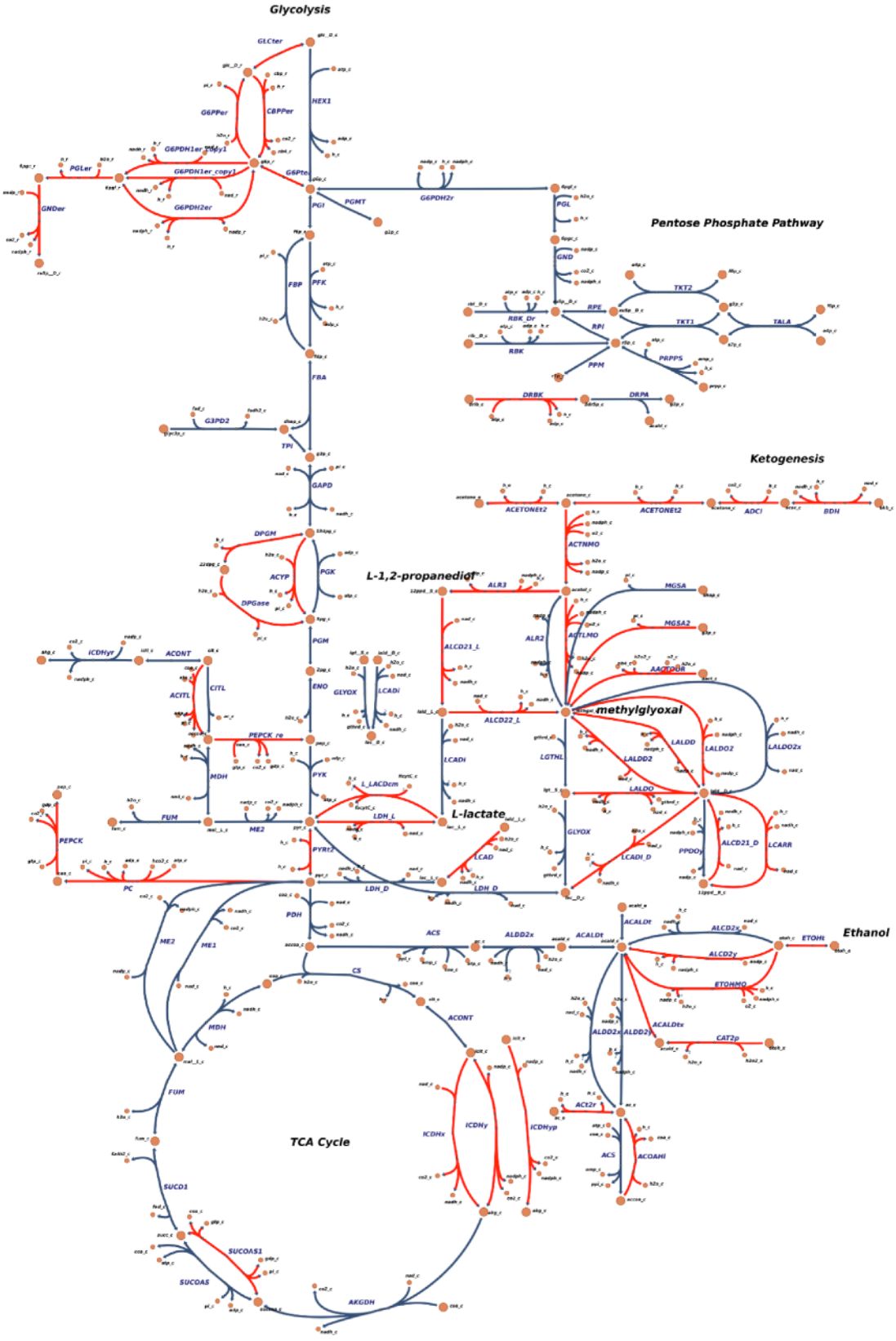
62. Tamasco G. How to add new Escher maps to ChiMera [Internet]. [cited 2022 Jan 7]. Available from: https://www.youtube.com/watch?v=YeAczYRWLTI
63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov;13(11):2498–504.
64. Tamasco G. How to visualize ChiMera metabolic maps using Cytoscape [Internet]. [cited 2022 Jan 7]. Available from: https://www.youtube.com/watch?v=M7SNCnPwqF0
65. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, et al. MEMOTE for standardized genome-scale metabolic model testing. Nat Biotechnol. 2020 Mar;38(3):272–6.
66. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc Natl Acad Sci. 2015 Mar 31;112(13):4110–5.
67. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential Bacillus subtilis genes. Proc Natl Acad Sci. 2003 Apr 15;100(8):4678–83.
68. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, et al. Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. Cell Syst. 2016 Sep 28;3(3):238-251.e12.
69. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes Escherichia coli traits. Nat Biotechnol. 2017 Oct;35(10):904–8.
70. Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, et al. High-quality genome-scale metabolic modelling of Pseudomonas putida highlights its broad metabolic capabilities. Environ Microbiol. 2020;22(1):255–69.
71. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol. 2010 Sep;28(9):977–82.
72. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data *. J Biol Chem. 2007 Sep 28;282(39):28791–9.
73. Henry CS, Zinner JF, Cohoon MP, Stevens RL. iBsu1103: a new genome-scale metabolic model of Bacillus subtilisbased on SEED annotations. Genome Biol. 2009 Jun 25;10(6):R69.
74. Capela J, Lagoa D, Rodrigues R, Cunha E, Cruz F, Barbosa A, et al. *merlin* v4.0: an updated platform for the reconstruction of high-quality genome-scale metabolic models [Internet]. Bioinformatics; 2021 Feb [cited 2021 Nov 3]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.02.24.432752

VI. SUPPLEMENTARY DATA

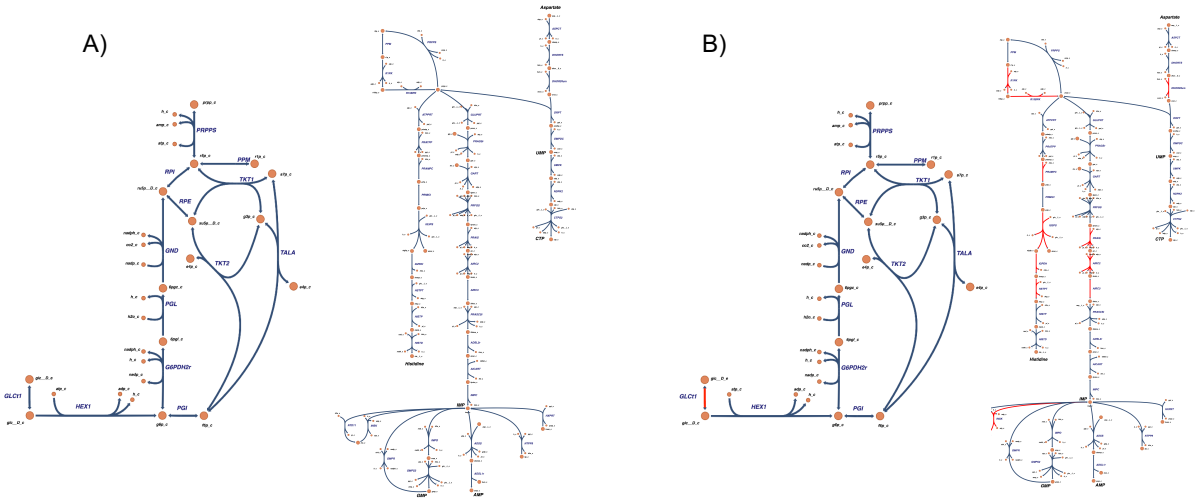
Supplementary Figure 1: ChiMera core module growth prediction output. The core module shows in the user screen the growth rate given the conditions provided to Chimera.

```
-----------------------------------------------------
Building Model
-----------------------------------------------------
p_putida.xml already exists
Basic info of the model
----------------------
model has 2606 reactions
model has 1721 metabolites
model has 1716 genes
The default flux bounds are:(-1000.0, 1000.0)
Maximize
1.0*Growth - 1.0*Growth_reverse_699ae
max
objective expression 1.0*Growth - 1.0*Growth_reverse_699ae
<Solution 0.495 at 0x7fc8e4b05310>
Objective
=========
1.0 Growth = 0.49524025041046654

Uptake
------
 Metabolite       Reaction      Flux  C-Number  C-Flux
    M_ca2_e     R_EX_ca2_e  0.002578         0   0.00%
     M_cl_e      R_EX_cl_e  0.002578         0   0.00%
M_cobalt2_e R_EX_cobalt2_e 4.952E-05         0   0.00%
    M_cu2_e     R_EX_cu2_e 0.0003511         0   0.00%
    M_fe2_e     R_EX_fe2_e  0.003546         0   0.00%
    M_fe3_e     R_EX_fe3_e  0.003867         0   0.00%
 M_glc__D_e  R_EX_glc__D_e        10         6 100.00%
      M_k_e       R_EX_k_e   0.09667         0   0.00%
    M_mg2_e     R_EX_mg2_e  0.004296         0   0.00%
    M_mn2_e     R_EX_mn2_e 0.0003422         0   0.00%
    M_nh4_e     R_EX_nh4_e     9.158         0   0.00%
     M_o2_e      R_EX_o2_e        10         0   0.00%
     M_pi_e      R_EX_pi_e    0.4774         0   0.00%
    M_so4_e     R_EX_so4_e    0.1241         0   0.00%
    M_zn2_e     R_EX_zn2_e 0.0001689         0   0.00%

Secretion
---------
Metabolite      Reaction       Flux  C-Number C-Flux
  M_4hbz_e   R_EX_4hbz_e  0.0001104         7  0.00%
   M_co2_e    R_EX_co2_e      16.78         1 42.29%
   M_h2o_e    R_EX_h2o_e      33.67         0  0.00%
     M_h_e      R_EX_h_e       8.37         0  0.00%
M_leu__L_e R_EX_leu__L_e      3.817         6 57.71%
```

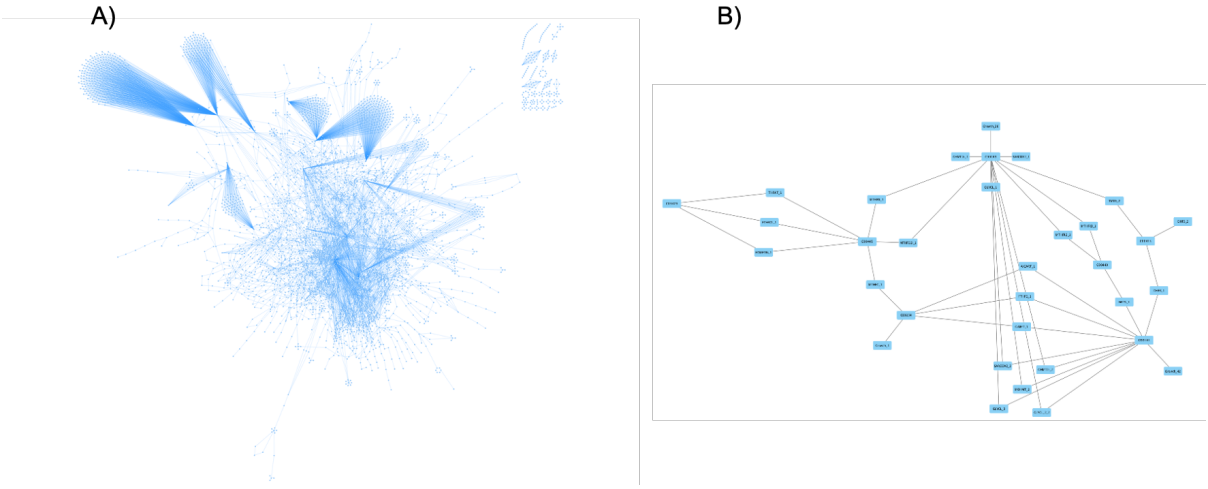
Supplementary Figure 2: Pre-defined map of carbon central metabolism in *E. coli*. Blue reactions were identified in the model. The reaction in red had no evidence detected in the model.



Supplementary Figure 3: Custom designed Escher maps for pentose phosphate and histidine biosynthesis. (A) These maps were designed based on the *E. coli* model inside Escher website. (B) The maps were included in the core module of ChiMera and used along the other ten maps. Blue and red indicate presence and absence of reactions in the model, respectively.



Supplementary Figure 4: ChiMera output visualized in Cytoscape. (A) Metabolic network of the entire organism. (B) Resulting Compounds and reactions associated with a search for one carbon by folate pathway.



Supplementary Figure 5: ChiMera output of single gene knockout. The pipeline prompts to the user the growth rate before and after the knockout. It also informs if the gene was not detected in the faa file.

```
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eeaf90>
tolQ knocked out:  <Solution 0.495 at 0x7fb541eea810>
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eeae90>
tolR knocked out:  <Solution 0.495 at 0x7fb541eeac90>
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eeacd0>
Gene tolA was not incoporated to the model automatically
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eeaf50>
Gene tolB was not incoporated to the model automatically
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eeaf50>
dapA knocked out:  <Solution 0.495 at 0x7fb541eef150>
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eef090>
dapE knocked out:  <Solution 0.000 at 0x7fb541eef5d0>
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eef150>
pdxB knocked out:  <Solution 0.495 at 0x7fb541eef110>
<<<Complete model>>>:  <Solution 0.495 at 0x7fb541eea510>
ccmA knocked out:  <Solution 0.495 at 0x7fb541ef3310>
```

Supplementary Table 1: Western Diet media composition Example of a new media composition to be provided to ChiMera..

medium	description	compound	name
western_diet_gut	average European diet adjusted for adsorption in the small intestine	26dap__M	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ac	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	acgam	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ala__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	amp	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	arab__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	arg__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	asn__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	asp__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ca2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cgly	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cl	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cobalt2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cu2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cys__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	dcyt	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	drib	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fe2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fe3	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fol	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	for	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fru	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gam	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gln__L	no

western_diet_gut	average European diet adjusted for adsorption in the small intestine	glu__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gly	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	glyc	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	glyc3p	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gua	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	h	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	h2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	h2o	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	h2s	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	his__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	hxan	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ile__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	indole	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	k	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	leu__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	lys__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	malt	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	man	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	met__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	mg2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	mn2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	mnl	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	nac	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	o2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	orn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	phe__L	no

western_diet_gut	average European diet adjusted for adsorption in the small intestine	pheme	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pi	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pro__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ptrc	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pydam	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pydx	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pydxn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	rib__D	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ribflv	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ser__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	so4	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	sucr	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	thm	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	thr__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	tre	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	trp__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	tyr__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ura	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	val__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	xyl__D	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	zn2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	2obut	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ade	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cit	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	csn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cytd	no

western_diet_gut	average European diet adjusted for adsorption in the small intestine	dad_2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	dgsn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fum	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ocdca	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	spmd	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	thymd	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ttdca	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	uri	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	xan	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	4abz	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ddca	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	hdca	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	nmn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pnto__R	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	arab__D	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	chol	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	glcn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gthrd	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	no2	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	acnam	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fuc__L	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gal	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	lcts	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pullulan1200	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	4hbz	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ncam	no

western_diet_gut	average European diet adjusted for adsorption in the small intestine	ocdcea	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	adn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	ala__D	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	chor	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	oxa	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gthox	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	rmn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	raffin	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	cellb	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	melib	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	gsn	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	mobd	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	meoh	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	amylose300	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	lmn30	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	acmana	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	starch1200	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	fald	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	starch	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	amylose	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	glc__D	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	pect	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	xylan4	no
western_diet_gut	average European diet adjusted for adsorption in the small intestine	xylan8	no

Supplementary Table 2: Gene/Reaction file for ChiMera knockout module. Users just need to provide a single gene or reaction name per line, no header should be informed.

dnaA
dnaB
dnaC
dnaD
dnaE
dnaG
dnaI
dnaN
dnaX
yqeN
holB
ligA

Supplementary Table 3: Confusion Matrix metadata used for model performance calculation. The original values were normalized based on the total number of observations.

Organism	Model	True Positive	False Positive	True Negative	False Negative
P. putida	iChiMera1716	0.83	0.62	0.37	0.16
	iJN1463	0.71	0.41	0.58	0.28
E. coli	iChiMera1657	0.98	0.30	0.70	0
	iEC1344_C	1	0	1	0
B. subtilis	iChiMera1182	0.72	0.52	0.47	0.27
	iYO844	0.73	0.20	0.79	0.26

Supplementary Table 4: List of carbon sources used to simulate growth from ChiMera and manually curated models of *B. subtilis*. Growth unit mmol/g[CDW]/h.

Reaction identifier	Metabolite identifier	Metabolite name	Predicted growth (iYO844)	Predicted growth (iChiMera1182)
EX_acgam_e	acgam	N-Acetyl-D-Glucosamine	1.10867078	0.15968736
EX_glcr_e	glcr	D-Saccharic Acid	0.59509134	0.16234508
EX_succ_e	succ	Succinic Acid	0.43279032	1.09E-16
EX_gal_e	gal	D-Galactose	0.62418481	0.03965804
EX_tre_e	tre	D-Trehalose	0.62418481	0.50655701
EX_man_e	man	D-Mannose	0.62418481	0.27627166
EX_galt_e	galt	Dulcitol	0.62418481	0.2732395
EX_glyc_e	glyc	Glycerol	0.4749642	0.14655855
EX_glcur_e	glcur	D-Glucuronic Acid	0.62418481	0.15763817
EX_glyc3p_e	glyc3p	D,L-a-Glycerol- Phosphate	0.5119264	0.03019911
EX_mnl_e	mnl	D-Mannitol	0.62418481	0.27627166
EX_g6p_e	g6p	D-Glucose-6-Phosphate	0.62418481	absent
EX_rmn_e	rmn	L-Rhamnose	0.62418481	0.03820638
EX_fru_e	fru	D-Fructose	0.62418481	0.27627166
EX_malt_e	malt	Maltose	0.62418481	0.60930755
EX_melib_e	melib	D-Melibiose	0.62418481	absent
EX_thymd_e	thymd	Thymidine	0.62781591	0.25711999
EX_akg_e	akg	a-Keto-Glutaric Acid	0.53539147	0.04811836
EX_sucr_e	sucr	Sucrose	0.62418481	0.50655701
EX_uri_e	uri	Uridine	0.64379081	0.35614481
EX_g1p_e	g1p	D-Glucose-1-Phosphate	0.62418481	absent
EX_f6p_e	f6p	D-Fructose-6-Phosphate	0.62418481	absent

EX_mbdg_e	mbdg	b-Methyl-D-Glucoside	0.62418481	0.27503957
EX_malttr_e	malttr	Maltotriose	0.62418481	0.63656605
EX_dad_2_e	dad_2	2`-Deoxy-Adenosine	1.06607664	0.07248535
EX_adn_e	adn	Adenosine	1.0603293	0.10389427
EX_cit_e	cit	Citric Acid	0.61001049	0.04824346
EX_fum_e	fum	Fumaric Acid	0.38615343	0.02841946
EX_ppa_e	ppa	Propionic Acid	0.37610019	5.36E-16
EX_glyclt_e	glyclt	Glycolic Acid	0.10534226	0.01831982
EX_glx_e	glx	Glyoxylic Acid	0.10534226	absent
EX_cellb_e	cellb	D-Cellobiose	0.62418481	0.50655701
EX_ins_e	ins	Inosine	1.04949897	0.12826859
EX_acmana_e	acmana	N-Acetyl-b-D-Mannosamine	1.18514342	0.15968736
EX_pyr_e	pyr	Pyruvic Acid	0.30342383	0.03442503
EX_galur_e	galur	D-Galacturonic Acid	0.62418481	0.15763817
EX_dextrin_e	dextrin	Dextrin	0.62418481	0.50037948
EX_glycogen_e	glycogen	Glycogen	0.62418481	absent
EX_acnam_e	acnam	N-Acetyl-Neuraminic Acid	1.57894108	0.27750557
EX_arbt_e	arbt	Arbutin	0.62418481	0.27841858
EX_drib_e	drib	2-Deoxy-D-Ribose	0.62418481	0.03665744
EX_pala_e	pala	Palatinose	0.62418481	absent
EX_raffin_e	raffin	D-Raffinose	0.62418481	0.41048088
EX_salcn_e	salcn	Salicin	0.62418481	0.27841859
EX_gam_e	gam	D-Glucosamine	0.85796633	0.27569252
EX_dha_e	dha	Dihydroxy-Acetone	0.44724256	absent

*absent = exchange reaction is not present in the model

Supplementary Table 5: List of carbon sources used to simulate growth from ChiMera and manually curated models of *E. coli*. Growth unit mmol/g[CDW]/h.

Reaction identifier	Metabolite identifier	Metabolite name	Predicted growth (iEC1344_C)	Predicted growth (iChiMera1657)
EX_5dglcn_e	5dglcn_e	5-Dehydro-D-gluconate	0.81648605	0.63409099
EX_dha_e	dha_e	Dihydroxyacetone	0.48466546	0.49481489
EX_pyr_e	pyr_e	Pyruvate	0.35990656	0.37007788
EX_thr__L_e	thr__L_e	L-Threonine	0.59620423	0.46385392
EX_ala__D_e	ala__D_e	D-Alanine	0.43463998	0.40478115
EX_fru_e	fru_e	D-Fructose	0.98239635	0.67435074
EX_ser__L_e	ser__L_e	L-Serine	0.36605028	0.38578525
EX_succ_e	succ_e	Succinate	0.49249687	0.40946345
EX_gal_e	gal_e	D-Galactose	0.97202695	0.64415593
EX_glc__D_e	glc__D_e	D-Glucose	0.98239635	0.67435074
EX_xyl__D_e	xyl__D_e	D-Xylose	0.80611666	0.60389618
EX_man_e	man_e	D-Mannose	0.98239635	0.67435074
EX_galur_e	galur_e	D-Galacturonate	0.79156522	0.5582789
EX_glcur_e	glcur_e	D-Glucuronate	0.79156522	0.5582789
EX_fuc__L_e	fuc__L_e	L-Fucose	-0.0239467	0.5063052
EX_rib__D_e	rib__D_e	D-Ribose	0.77500847	0.51331175
EX_glyc_e	glyc_e	Glycerol	0.56280423	0.54853903
EX_ac_e	ac_e	Acetate	0.24720539	0.26417919
EX_g1p_e	g1p_e	D-Glucose 1-phosphate	0.98239635	0.67435074
EX_mal__D_e	mal__D_e	D-Malate	0.42847885	0.39249163
EX_mal__L_e	mal__L_e	L-Malate	0.44942762	0.47210664
EX_lac__D_e	lac__D_e	D-Lactate	0.42847885	0.40478115

EX_lac__L_e	lac__L_e	L-Lactate	0.39909073	0.33562933
EX_akg_e	akg_e	2-Oxoglutarate	0.6179377	0.5027076
EX_tartr__D_e	tartr__D_e	D-tartrate	0.36328912	0.39699876
EX_tartr__L_e	tartr__L_e	L-tartrate	0.37405643	0.39699876
EX_acgal_e	acgal_e	N-Acetyl-D-galactosamine	-0.0239467	0
EX_acnam_e	acnam_e	N-Acetylneuraminate	1.64668289	0.59006791
EX_glyc3p_e	glyc3p_e	Glycerol 3-phosphate	0.60324825	0.60858629
EX_adn_e	adn_e	Adenosine	1.29463431	0.80275901
EX_all__D_e	all__D_e	D-Allose	0.94091877	0.5535715
EX_f6p_e	f6p_e	D-Fructose 6-phosphate	1.01350453	0.77913215
EX_galct__D_e	galct__D_e	D-Galactarate	0.68830263	0.46062413
EX_glcr_e	glcr_e	D-Glucarate	0.68830263	0.46062413
EX_gam_e	gam_e	D-Glucosamine	0.98278151	0.6751212
EX_g6p_e	g6p_e	D-Glucose 6-phosphate	1.01350453	0.77913215
EX_mnl_e	mnl_e	D-Mannitol	1.05484847	0.66931827
EX_sbt__D_e	sbt__D_e	D-Sorbitol	1.05484847	0.66931827
EX_dad_2_e	dad_2_e	Deoxyadenosine	1.34537539	0.7152439
EX_fum_e	fum_e	Fumarate	0.44942762	0.47210664
EX_glyclt_e	glyclt_e	Glycolate C2H3O3	0.18498823	0.19717356
EX_ins_e	ins_e	Inosine	1.28290133	0.76815272
EX_malt_e	malt_e	Maltose C12H22O11	1.97785813	0.91590921
EX_malttr_e	malttr_e	Maltotriose C18H32O16	-0.0239467	0.92718303
EX_melib_e	melib_e	Melibiose C12H22O11	1.96748873	0.8857144
EX_ppa_e	ppa_e	Propionate (n-C3:0)	0.42847885	0.2922571
EX_thymd_e	thymd_e	Thymidine C10H14N2O5	0.88424604	0.64423604

EX_tre_e	tre_e	Trehalose	1.97785813	0.91590921
EX_uri_e	uri_e	Uridine	0.87293509	0.73732594
EX_dextrin_e	dextrin_e	Dextrin C12H20O10	absent	-6.84E-15

*absent = exchange reaction is not present in the model.

Supplementary Table 6: List of carbon sources used to simulate growth from ChiMera and manually curated models of *P. putida*. Growth unit mmol/g[CDW]/h.

Reaction identifier	Metabolite identifier	Metabolite name	Predicted growth (iJN1463)	Predicted growth (iChiMera1716)
EX_etoh_e	etoh	Ethanol	0.46534778	0.23749025
EX_ocdca_e	ocdca	octadecanoate-n-C180-	2.81844533	0.33967394
EX_tag180_e	tag180	Triacylglycerol-octadecanoate-	2.81844533	0.5359456
EX_12dgr180_e	12dgr180	1-2-Diacyl-sn-glycerol-dioctadecanoyl-n-C180-	2.81844533	0.5359456
EX_1ag180_e	1ag180	1-Acyl-sn-glycerol-octadecanoate-	2.81844533	0.5359456
EX_12dgr160_e	12dgr160	1-2-Diacyl-sn-glycerol-dihexadecanoyl-n-C160-	2.81844533	0.53640677
EX_hdca_e	hdca	Hexadecanoate-n-C160-	2.81844533	0.33851224
EX_tag160_e	tag160	Triacylglycerol-hexadecanoate-	2.81844533	0.53640677
EX_1ag160_e	1ag160	1-Acyl-sn-glycerol-hexadecanoate-	2.81844533	0.53640677
EX_vacc_e	vacc	vaccenic-acid	2.81844533	0.34333965
EX_1ag181d9_e	1ag181d9	1-Acyl-sn-glycerol-nC181d9-	2.81844533	0.53609923
EX_tag181d9_e	tag181d9	Triacylglycerol-nC181d9-	2.81844533	0.53609923
EX_ocdcea_e	ocdcea	octadecenoate-n-C181-	2.81844533	0.34333965
EX_ptrc_e	ptrc	Putrescine	0.82991177	0.29061392
EX_ptsla_e	ptsla	petroselaidic-acid	2.81844533	0.34445068
EX_ttdca_e	ttdca	tetradecanoate-n-C140-	2.81844533	0.33205143
EX_hdcea_e	hdcea	Hexadecenoate-n-C161-	2.81844533	0.34261337

EX_ddca_e	ddca	Dodecanoate-n-C120-	2.47216244	0.33015271
EX_ttdcea_e	ttdcea	tetradecenoate-n-C141-	2.81844533	0.33656786
EX_1ag182d9d12_e	1ag182d9d12	1-Acyl-sn-glycerol-nC182d9d12-	2.81844533	0.53062324
EX_dag182d9d12_e	dag182d9d12	1-2-Diacyl-sn-glycerol-nC182d9d12-	2.81844533	0.53062324
EX_tag182d9d12_e	tag182d9d12	Triacylglycerol-nC182d9d12-	2.81844533	0.53062324
EX_lnlc_e	lnlc	linoleic-acid-all-cis-C182-n-6	2.81844533	0.34180734
EX_dca_e	dca	Decanoate-n-C100-	2.03107589	0.32753069
EX_glyc_e	glyc	Glycerol	0.60177534	0.53062324
EX_nona_e	nona	Nonanoate	1.79320397	0.3389072
EX_octa_e	octa	octanoate-n-C80-	1.58758076	0.32367484
EX_hxa_e	hxa	Hexanoate-n-C60-	1.14601299	0.31744628
EX_pta_e	pta	Pentanoate	0.91111239	0.29191405
EX_chol_e	chol	Choline	0.89299386	0.25581841
EX_25dkglcn_e	25dkglcn	2,5-diketo-D-gluconate	0.77609236	absent
EX_fru_e	fru	D-Fructose	1.05059347	0.63467713
EX_3mb_e	3mb	3-Methylbutanoic-acid	0.86843683	0.28411076
EX_4abut_e	4abut	4-Aminobutanoate	0.68760671	0.36291464
EX_5aptn_e	5aptn	5-Aminopentanoate	0.82555787	absent
EX_peamn_e	peamn	Phenethylamine	1.30927788	0.23841109
EX_pac_e	pac	Phenylacetic-acid	1.25455682	0.27134712
EX_glcn_e	glcn	D-Gluconate	0.93440471	0.52352918
EX_glyb_e	glyb	Glycine-betaine	0.75250446	0.25581841
EX_phpyr_e	phpyr	Phenylpyruvate	1.35108609	absent
EX_2dhglcn_e	2dhglcn	2-Dehydro-D-gluconate	0.88917261	absent
EX_acac_e	acac	Acetoacetate	0.58541781	0.32683134

EX_catechol_e	catechol	Catechol	0.86931493	absent
EX_3oxoadp_e	3oxoadp	3-Oxoadipate	0.86905722	0.40745465
EX_confrl_e	confrl	Coniferol	1.44060841	0.2197711
EX_glutar_e	glutar	Glutarate	0.7164033	0.17994066
EX_m_xyl_e	m_xyl	m-Xylene	1.14266143	0
EX_p_xyl_e	p_xyl	p-methyltoluene	1.09855685	0
EX_hgentis_e	hgentis	Homogentisate	1.09039365	absent
EX_ac_e	ac	Acetate	0.26886955	0.19353295
EX_34dhcinm_e	34dhcinm	3-4-Dihydroxy-trans-cinnamate	1.22172419	0.2560464
EX_tol_e	tol	Toluene	0.94670333	-1.38E-15
EX_glcur_e	glcur	D-Glucuronate	0.80790279	0.42402465
EX_5oxpro_e	5oxpro	5-Oxoproline	0.6649254	0.34325615
EX_succ_e	succ	Succinate	0.52883751	0.36111582
EX_fer_e	fer	Ferulate	1.29833367	absent
EX_T4hcinnm_e	T4hcinnm	trans-4-Hydroxycinnamate	1.14511471	0.20392944
EX_cit_e	cit	Citrate	0.75377703	0.46044428
EX_icit_e	icit	Isocitrate	0.75377703	0.46044428
EX_bz_e	bz	Benzoate	0.86931493	0.19850469
EX_34dhbz_e	34dhbz	3-4-Dihydroxybenzoate	0.85771656	0.2638147
EX_fum_e	fum	Fumarate	0.48347489	absent
EX_ga_e	ga	Gallic-acid	0.81235394	0.27912114
EX_4hbz_e	4hbz	4-Hydroxybenzoate	0.77833197	0.18868567
EX_chols_e	chols	Choline-sulfate	0.52885536	0.16613172
EX_glyclt_e	glyclt	Glycolate	0.18172591	0.15423271
EX_6hnac_e	6hnac	6-Hydroxynicotinate	0.49481554	0.15491929

EX_nac_e	nac	Nicotinate	0.49487047	0.13278708
EX_xan_e	xan	Xanthine	0.29735752	absent
EX_6atha_e	6atha	6-acetylthiohexanoic-acid	0.43260141	0.22822339

*absent = exchange reaction is not present in the model.

Supplementary Table 7: Parcial ChiMera output when knocking out all genes/reactions from the model. Only the first 100 genes were displayed.

Gene	ids	growth	status
0	{'G_lcl_NZ_CP027599_1_prot_WP_001305111_1_708'}	0.6744	optimal
1	{'G_lcl_NZ_CP027599_1_prot_WP_000357740_1_5606'}	0.6744	optimal
2	{'G_lcl_NZ_CP027599_1_prot_WP_001065885_1_696'}	0.6744	optimal
3	{'G_lcl_NZ_CP027599_1_prot_WP_001292353_1_3112'}	0.6744	optimal
4	{'G_lcl_NZ_CP027599_1_prot_WP_001069997_1_2756'}	0.6744	optimal
5	{'G_lcl_NZ_CP027599_1_prot_WP_000748261_1_1893'}	0.6744	optimal
6	{'G_lcl_NZ_CP027599_1_prot_WP_000440782_1_1079'}	0.6744	optimal
7	{'G_lcl_NZ_CP027599_1_prot_WP_000253975_1_347'}	0.6744	optimal
8	{'G_lcl_NZ_CP027599_1_prot_WP_001094499_1_1340'}	0.0	optimal
9	{'G_lcl_NZ_CP027599_1_prot_WP_000555854_1_3549'}	0.6744	optimal
10	{'G_lcl_NZ_CP027599_1_prot_WP_000187442_1_674'}	0.6744	optimal
11	{'G_lcl_NZ_CP027599_1_prot_WP_000172466_1_3236'}	0.6744	optimal
12	{'G_lcl_NZ_CP027599_1_prot_WP_001272796_1_706'}	0.6744	optimal
13	{'G_lcl_NZ_CP027599_1_prot_WP_001021161_1_4718'}	0.6744	optimal
14	{'G_lcl_NZ_CP027599_1_prot_WP_000965518_1_1895'}	0.6744	optimal
15	{'G_lcl_NZ_CP027599_1_prot_WP_000692135_1_2729'}	0.6744	optimal
16	{'G_lcl_NZ_CP027599_1_prot_WP_000438245_1_566'}	0.6744	optimal
17	{'G_lcl_NZ_CP027599_1_prot_WP_001010707_1_2697'}	0.6744	optimal

18	{'G_lcl_NZ_CP027599_1_prot_WP_001081870_1_763'}	0.6744	optimal
19	{'G_lcl_NZ_CP027599_1_prot_4820'}	0.6744	optimal
20	{'G_lcl_NZ_CP027599_1_prot_WP_000055075_1_5384'}	0.6744	optimal
21	{'G_lcl_NZ_CP027599_1_prot_WP_000044066_1_5015'}	0.6744	optimal
22	{'G_lcl_NZ_CP027599_1_prot_WP_000017694_1_764'}	0.6744	optimal
23	{'G_lcl_NZ_CP027599_1_prot_WP_000986029_1_1352'}	0.6744	optimal
24	{'G_lcl_NZ_CP027599_1_prot_WP_000550422_1_4475'}	0.6744	optimal
25	{'G_lcl_NZ_CP027599_1_prot_WP_000513775_1_4202'}	0.6744	optimal
26	{'G_lcl_NZ_CP027599_1_prot_WP_001307577_1_5016'}	0.6744	optimal
27	{'G_lcl_NZ_CP027599_1_prot_WP_000374004_1_5718'}	0.6744	optimal
28	{'G_lcl_NZ_CP027599_1_prot_WP_001043882_1_2595'}	0.6744	optimal
29	{'G_lcl_NZ_CP027599_1_prot_WP_001307427_1_407'}	0.6744	optimal
30	{'G_lcl_NZ_CP027599_1_prot_WP_000191073_1_3107'}	0.6744	optimal
31	{'G_lcl_NZ_CP027599_1_prot_WP_000062997_1_1915'}	0.6744	optimal
32	{'G_lcl_NZ_CP027599_1_prot_WP_000254843_1_1647'}	0.6744	optimal
33	{'G_lcl_NZ_CP027599_1_prot_WP_000534635_1_4134'}	0.6744	optimal
34	{'G_lcl_NZ_CP027599_1_prot_WP_000172749_1_1922'}	0.6744	optimal
35	{'G_lcl_NZ_CP027599_1_prot_WP_000057977_1_2999'}	0.6744	optimal
36	{'G_lcl_NZ_CP027599_1_prot_WP_001276671_1_1423'}	0.6744	optimal
37	{'G_lcl_NZ_CP027599_1_prot_WP_001307229_1_2789'}	0.6744	optimal
38	{'G_lcl_NZ_CP027599_1_prot_WP_000137048_1_509'}	0.6744	optimal
39	{'G_lcl_NZ_CP027599_1_prot_WP_001285544_1_3199'}	0.6744	optimal
40	{'G_lcl_NZ_CP027599_1_prot_WP_001140652_1_5097'}	0.0	optimal
41	{'G_lcl_NZ_CP027599_1_prot_WP_000767718_1_1141'}	0.6744	optimal
42	{'G_lcl_NZ_CP027599_1_prot_WP_000813860_1_1912'}	0.6744	optimal
43	{'G_lcl_NZ_CP027599_1_prot_WP_001033352_1_3574'}	0.6744	optimal

44	{'G_lcl_NZ_CP027599_1_prot_WP_000651298_1_1611'}	0.6744	optimal
45	{'G_lcl_NZ_CP027599_1_prot_WP_001173673_1_1130'}	0.6744	optimal
46	{'G_lcl_NZ_CP027599_1_prot_WP_001148478_1_515'}	0.6744	optimal
47	{'G_lcl_NZ_CP027599_1_prot_WP_000963792_1_419'}	0.6744	optimal
48	{'G_lcl_NZ_CP027599_1_prot_WP_001130266_1_1111'}	0.6744	optimal
49	{'G_lcl_NZ_CP027599_1_prot_WP_000012163_1_959'}	0.6744	optimal
50	{'G_lcl_NZ_CP027599_1_prot_WP_000057149_1_4123'}	0.6388	optimal
51	{'G_lcl_NZ_CP027599_1_prot_WP_001055435_1_1464'}	0.6744	optimal
52	{'G_lcl_NZ_CP027599_1_prot_WP_000183940_1_4488'}	0.6744	optimal
53	{'G_lcl_NZ_CP027599_1_prot_WP_000722571_1_3256'}	0.6744	optimal
54	{'G_lcl_NZ_CP027599_1_prot_WP_000187594_1_4425'}	0.6744	optimal
55	{'G_lcl_NZ_CP027599_1_prot_5653'}	0.6744	optimal
56	{'G_lcl_NZ_CP027599_1_prot_WP_000937432_1_5024'}	0.6744	optimal
57	{'G_lcl_NZ_CP027599_1_prot_WP_000110493_1_972'}	0.6744	optimal
58	{'G_lcl_NZ_CP027599_1_prot_WP_000477811_1_5193'}	0.6744	optimal
59	{'G_lcl_NZ_CP027599_1_prot_WP_000235203_1_4095'}	0.6744	optimal
60	{'G_lcl_NZ_CP027599_1_prot_WP_000129551_1_2188'}	0.6744	optimal
61	{'G_lcl_NZ_CP027599_1_prot_WP_000276149_1_2827'}	0.6744	optimal
62	{'G_lcl_NZ_CP027599_1_prot_WP_001297158_1_641'}	0.6744	optimal
63	{'G_lcl_NZ_CP027599_1_prot_WP_000812724_1_2589'}	0.6744	optimal
64	{'G_lcl_NZ_CP027599_1_prot_WP_001299058_1_1603'}	0.6744	optimal
65	{'G_lcl_NZ_CP027599_1_prot_WP_001275861_1_4794'}	0.6744	optimal
66	{'G_lcl_NZ_CP027599_1_prot_WP_000034372_1_915'}	0.6744	optimal
67	{'G_lcl_NZ_CP027599_1_prot_WP_000082693_1_5948'}	0.6744	optimal
68	{'G_lcl_NZ_CP027599_1_prot_WP_000617097_1_3182'}	0.6744	optimal
69	{'G_lcl_NZ_CP027599_1_prot_WP_001158929_1_5017'}	0.6744	optimal

70	{'G_lcl_NZ_CP027599_1_prot_WP_001195240_1_4181'}	0.6744	optimal
71	{'G_lcl_NZ_CP027599_1_prot_WP_000218360_1_5419'}	0.6744	optimal
72	{'G_lcl_NZ_CP027599_1_prot_WP_000015837_1_5473'}	0.6744	optimal
73	{'G_lcl_NZ_CP027599_1_prot_WP_000347235_1_4712'}	0.0	optimal
74	{'G_lcl_NZ_CP027599_1_prot_WP_000517439_1_1635'}	0.6744	optimal
75	{'G_lcl_NZ_CP027599_1_prot_WP_000531601_1_3655'}	0.6744	optimal
76	{'G_lcl_NZ_CP027599_1_prot_WP_000082101_1_323'}	0.6744	optimal
77	{'G_lcl_NZ_CP027599_1_prot_WP_000189036_1_216'}	0.6744	optimal
78	{'G_lcl_NZ_CP027599_1_prot_WP_000009269_1_5765'}	0.6744	optimal
79	{'G_lcl_NZ_CP027599_1_prot_WP_000954872_1_2257'}	0.0	optimal
80	{'G_lcl_NZ_CP027599_1_prot_WP_000736013_1_5069'}	0.6744	optimal
81	{'G_lcl_NZ_CP027599_1_prot_WP_001234767_1_2218'}	0.6744	optimal
82	{'G_lcl_NZ_CP027599_1_prot_WP_000084573_1_1654'}	0.6744	optimal
83	{'G_lcl_NZ_CP027599_1_prot_WP_001018496_1_3857'}	0.6744	optimal
84	{'G_lcl_NZ_CP027599_1_prot_WP_000227958_1_5831'}	0.6744	optimal
85	{'G_lcl_NZ_CP027599_1_prot_WP_000192245_1_4474'}	0.6744	optimal
86	{'G_lcl_NZ_CP027599_1_prot_WP_001170683_1_2813'}	0.6744	optimal
87	{'G_lcl_NZ_CP027599_1_prot_WP_000378946_1_871'}	0.6744	optimal
88	{'G_lcl_NZ_CP027599_1_prot_WP_000102485_1_5052'}	0.674	optimal
89	{'G_lcl_NZ_CP027599_1_prot_WP_001154187_1_2716'}	0.6744	optimal
90	{'G_lcl_NZ_CP027599_1_prot_WP_001251332_1_3156'}	0.6744	optimal
91	{'G_lcl_NZ_CP027599_1_prot_WP_001102289_1_1579'}	0.6744	optimal
92	{'G_lcl_NZ_CP027599_1_prot_WP_000865576_1_2052'}	0.6744	optimal
93	{'G_lcl_NZ_CP027599_1_prot_WP_000437935_1_1904'}	0.6744	optimal
94	{'G_lcl_NZ_CP027599_1_prot_WP_000784004_1_873'}	0.6744	optimal
95	{'G_lcl_NZ_CP027599_1_prot_WP_000548675_1_2448'}	0.6744	optimal

96	{'G_lcl_NZ_CP027599_1_prot_WP_000771770_1_4651'}	0.6744	optimal
97	{'G_lcl_NZ_CP027599_1_prot_WP_000854640_1_3240'}	0.6744	optimal
98	{'G_lcl_NZ_CP027599_1_prot_WP_000073591_1_329'}	0.6744	optimal
99	{'G_lcl_NZ_CP027599_1_prot_WP_001250644_1_5742'}	0.6744	optimal