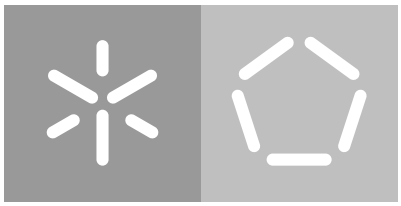


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Miguel Ângelo Fernandes da Silva

Genome-Scale Metabolic Modelling of the pathogen *Xylella fastidiosa*

October 2019



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Miguel Ângelo Fernandes da Silva

Genome-Scale Metabolic Modelling of the pathogen *Xylella fastidiosa*

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Oscar Dias
Cristiana Faria

October 2019

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição
CC BY

https://creativecommons.org/licenses/by/4.0/

ACKNOWLEDGEMENTS

I would like to state that during this work I overcame several challenges that allowed me to extend my professional skills and experience. Without the help of everyone that accompanied and made part of my life throughout this year this would not be possible, therefore I am forever grateful to you all.

In the first place, I would like to express my sincere gratitude to my supervisors for giving me the opportunity to study such an emergent issue of our society and for all the help and guidance throughout the development of this work. A huge word of appreciation to Doctor Oscar Dias, for sharing all of his knowledge in the Systems Biology field of study during this work and for his availability whenever I requested. Also, I would like to acknowledge Doctor Cristiana Faria for her wonderful support and effort in this work. I look forward to keep working with you both and thank you for this amazing opportunity.

I would like to acknowledge everyone from the computational biology laboratory for the great working environment and amazing cooperation. A special thanks to Emanuel, as I am sure that our mutual help benefited both of our works. Also, I would like to acknowledge Fernando for his 'mentorship' for the past 2 years. Last but not least, I extend my gratitude to Davide for his exceptional help during this work.

A huge thank you to all my family, in particular to my mother, Maria, and brother, César, as all this would be impossible without you. Thank you for your restless support and effort, which allowed me to pursue my dreams and professional aims with the least of worries. This success also belongs to you, thank you!

I would like to address a word of appreciation to all my friends. Thank you to Alexandre and Freitas for all the lively and priceless moments we had during this year. Thank you to Manel, Luís, Marcelo, Ricardo, Tiago for being extraordinary friends and for your constant 'bullying' that brings the best of me (I know, it is counter-intuitive).

Finally, I have to thank the person that accompanied me the most throughout the past years, my best friend and girlfriend, Catarina. I am forever grateful for your amazing help, friendship and advice. Thank you for all the time we have together and for always listening to my concerns. Your 'randomess' and our nonsense conversations always bright up my day. Thank you for always cheering me up, encourage me to keep going and improve myself. I am sure that your hard-working personality and quick-witted mind will make very successful in the near future and I'm here to support you whenever you need. I am eternally grateful to you, Catarina!

This work is a result of the project 22231/01/SAICT/2016: Biodata.pt Infraestrutura Portuguesa de Dados Biológicos, supported by the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).



STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

ABSTRACT

Xylella fastidiosa is a phytopathogenic bacteria that causes disease in hundreds of different plant species. Increased reports of plants infected by these xylem-limited bacteria are alarming as this pathogen continues to attack crops of relevant economic power such as citrus, grapes, olives and almonds, with considerable economical losses for the producers. The current employed strategy to contain this epidemic is radical in action as it destroys the infected plant and surrounding area. For this reason, it became urgent to develop new ways to eliminate these bacteria with therapeutics that are more pathogen oriented.

Genome-Scale Metabolic (GSM) models contain genomic and metabolic information of a given organism and can be used to discover new potential drug targets. Thus, a GSM model of *X. fastidiosa* may unveil new ways to control these bacteria.

In this work, we developed a high-quality GSM model for *X. fastidiosa* subsp. *pauca* De Donno, using the user-friendly software Metabolic Models Reconstruction Using Genome-Scale Information (merlin). This strain was chosen for its importance in the national economy, as it causes Olive Quick Decline Syndrome. The reconstructed model of *X. fastidiosa* comprises a set of 1280 reactions and 524 genes.

The genome of *X. fastidiosa* subsp. *pauca* De Donno was functionally annotated in order to identify the metabolic potential of the phytopathogenic organism. Metabolic functions identified in the genome were used to assemble the initial draft metabolic network. Manual curation procedures were made in order to correctly represent organism's capabilities and biomass related reactions, based on literature and experimental data, were added to model. The reconstructed model was then validated using experimental data, regarding the aerobic metabolism, carbon flux pattern, carbon usage, amino acid auxotrophies and growth in several media developed for *X. fastidiosa*.

In silico simulations revealed interesting metabolic properties of *X. fastidiosa*. The usage of carbon through the Entner-Doudoroff pathway seems to be a way of generating redox potential for defensive mechanism against the host plant. An absence of auxotrophies and the presence of catabolic routes for amino acids shows the metabolic and adaptive potential of the organism, as expected since it grows on nutrient-limited environments.

This reconstructed model can be used to explore the metabolism and provide information for potential drug targets for the agricultural-destructive phytopathogen *X. fastidiosa*.

Keywords: *Xylella fastidiosa* subsp. *pauca* De Donno, Phytopathogen, Systems Biology, Genome-Scale Metabolic Model

RESUMO

Xylella fastidiosa é uma bactéria fitopatogénica que causa doenças em centenas de espécies de plantas. O aumento de relatos de plantas infetadas por esta bactéria limitada ao xilema é alarmante, uma vez que o patogéneo ataca cultivações de significante valor económico, como o caso de citrinos, uvas, azeitonas e amêndoas, resultando em perdas económicas extensivas para produtores. A atual estratégia de contenção da epidemia é radical, uma vez que resulta na destruição da planta infetada e da área circundante. Desta forma, é urgente desenvolver novas formas de eliminar esta bactéria com terapias direcionadas ao patogéneo.

Modelos metabólicos à escala-genómica contêm informação genómica e metabólica de um organism e podem ser utilizados para descobrir novos potenciais alvos terapêuticos. Assim, um modelo metabólico à escala-genómica de *X. fastidiosa* poderá desvendar novas formas de controlar a bactéria.

Nesta tese, foi desenvolvido um modelo de alta qualidade para *X. fastidiosa* subsp. *pauca* De Donno, recorrendo à ferramenta *merlin*. Esta estirpe foi escolhida pela sua importância na economia nacional, visto que causa Síndrome de declínio rápido da oliveira. O modelo reconstruído de *X. fastidiosa* contém um conjunto de 1280 reações e 524 genes.

O genoma de *X. fastidiosa* subsp. *pauca* De Donno foi funcionalmente anotado de forma a identificar o potencial metabólico do organismo fitopatogénico. Funções metabólicas identificadas no genoma foram utilizadas para agregar uma rede metabólica. Posteriormente, curação manual da rede metabólica foi feita para representar corretamente as capacidades do organismo, e reações associadas à biomassa, baseadas em literatura e dados experimentais, foram adicionadas ao modelo. O modelo reconstruído foi de seguida validado usando dados experimentais em relação ao metabolismo aeróbico, padrão do fluxo de carbono, utilização de carbono, auxotrofias de aminoácidos e crescimento em diversos meios criados para o crescimento de *X. fastidiosa*.

Simulações *in silico* revelaram propriedades metabólicas interessantes de *X. fastidiosa*. A utilização de carbono pela via Entner-Doudoroff parece ser uma estratégia para gerar potencial redutor para mecanismos defensivos contra a planta hospedeira. A ausência de auxotrofias e a presença de vias catabólicas de aminoácidos mostram o potencial metabólico e adaptativo do organismo, como seria de esperar face ao seu crescimento natural em ambientes com restrições nutritivas.

Este modelo pode ser usado para explorar o metabolismo e fornecer informação de potenciais alvos terapéuticos para o fitopatogénico destrutivo para a agricultura, *X. fastidiosa*.

CONTENTS

1 INTRODUCTION 1
1.1 Context and Motivation 1
1.2 Goals 2
1.3 Structure of the Document 2
2 STATE-OF-THE-ART 4
2.1 Genome-Scale Metabolic Models 4
2.1.1 Background 4
2.1.2 Genome Annotation 7
2.1.3 Metabolic Network Reconstruction 8
2.1.4 Stoichiometric Model 12
2.1.5 Model Validation 14
2.2 Computational tools 16
2.2.1 *merlin* 16
2.2.2 Simulation tools 18
2.3 Xylella fastidiosa 19
2.3.1 Background 19
2.3.2 *In vivo* growth 22
2.3.3 *Xylella fastidiosa* subsp. *pauca* De Donno 23
3 MATERIALS AND METHODS 24
3.1 Software 24
3.2 Genome Files 24
3.3 Genome Annotation 24
3.4 Metabolic Network Reconstruction 26
3.4.1 Metabolic Data 26
3.4.2 Transport Reactions, Compartments and Exchange Reactions 27
3.4.3 GPR Associations 27
3.5 Biomass and Energy Requirements 27
3.6 Manual Curation 29
3.6.1 Reversibility and Directionality 29
3.6.2 Gap-Filling 29
3.6.3 Reaction's Balance 31
3.7 Model Validation 31
3.7.1 Aerobic Metabolism 32
3.7.2 Metabolic Flux Pattern 32

3.7.3 Specific Growth and Exopolysaccharide production rates 33
3.7.4 Metabolism under different carbon sources 34
3.7.5 Amino acid auxotrophies 34
3.7.6 Growth in different chemically defined media 35
3.8 Potential drug targets 37
4 RESULTS AND DISCUSSION 38
4.1 Genome Annotation 38
4.2 Transport reactions 39
4.3 Biomass and Energy Requirements 40
4.4 Model Curation 47
4.4.1 Reversibility and Directionality 47
4.4.2 Gap-Filling 48
4.4.3 Reaction's Balance 50
4.5 Model Validation 51
4.5.1 Aerobic Metabolism 51
4.5.2 Metabolic Flux Pattern 52
4.5.3 Specific Growth and Exopolysaccharide production rates 54
4.5.4 Metabolism under different carbon sources 55
4.5.5 Amino acid auxotrophies 57
4.5.6 Growth in different chemically defined media 58
4.6 Topological Analysis of the Model 60
4.7 Drug Targets 61
5 CONCLUSION AND FUTURE WORK 64

A SUPPORT MATERIAL 86
A.1 Generic Metabolites 86
A.2 Reaction Distribution 89
A.3 Critical Genes 93

LIST OF FIGURES

Figure 1	GSM model reconstruction iterative process.	5
Figure 2	Example of GPR associations and its assigned Boolean rules for *Escherichia coli*.	9
Figure 3	Illustration of a metabolic network.	15
Figure 4	*Xylella fastidiosa* life cycle.	21
Figure 5	Scheme of the annotation workflow followed in this work.	26
Figure 6	Workflow to calculate precursors' stoichiometric coefficients in macromolecular synthesis reactions.	28
Figure 7	Example of a KEGG pathway map section 'painted' by *merlin*'s 'Draw in browser' tool.	30
Figure 8	Schematic view of the carbon flux distribution predicted for *X. fastidiosa*.	53
Figure 9	Growth rates predictions under different carbon sources by *X. fastidiosa* and *X. campestris* models. Experimental data from *X. campestris* is also presented.	56

LIST OF TABLES

Table 1 Online data sources used in GSM model reconstruction 6
Table 2 Environmental conditions used to simulated aerated and anaerobic environments. 32
Table 3 Environmental conditions applied to *X. fastidiosa* model during validation of the glucose metabolic flux pattern. 33
Table 4 Environmental conditions applied to *X. fastidiosa* model during validation of specific growth and exopolysaccharide production rates. 34
Table 5 Environmental conditions applied during the amino acid auxotrophies validation phase. 35
Table 6 Metabolites used to simulate different chemically defined media. 36
Table 7 Genome annotation automatic workflow results. 38
Table 8 Distribution of metabolic genes among EC classes and their respective percentage. 39
Table 9 Number of generated and included in the model transport reactions organized by class. 40
Table 10 Biomass composition of *E. coli*, *X. campestris* and *X. fastidiosa* models. 41
Table 11 DNA synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 42
Table 12 RNA synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 42
Table 13 Protein synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 43
Table 14 Lipid synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 44
Table 15 Fatty acid synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 44
Table 16 Lipopolysaccharide synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 45
Table 17 Peptidoglycan synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 46
Table 18 Cofactor synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model. 46

Table 19 Output ratios of competing entities in both *X. campestris* and *X. fastidiosa* models. 47
Table 20 Summary of manually curated reactions, in terms of reversibility and directionality. 48
Table 21 Reactions added to the model during gap-filling analysis. 49
Table 22 Annotation corrections made during gap-filling analysis. 50
Table 23 Summary of corrections made during reaction balance manual curation. 51
Table 24 Reactions associated to metabolite breakdown errors and their respective correction. 51
Table 25 Predicted specific growth and exopolysaccharide production rates for *X. fastidiosa* and *X. campestris* models, as well as experimental data of *X. campestris* retrieved from *Letisse et al.* 54
Table 26 Production rate of metabolites predicted by a *in silico* simulation. 55
Table 27 Results of amino acids auxotrophies obtained for *X. fastidiosa* subsp. *pauca* De Donno. 58
Table 28 Qualitative evaluation of model predictions simulating several chemically defined media developed for *X. fastidiosa*. 59
Table 29 Summary of the reconstructed network of *X. fastidiosa*'s model. 60
Table 30 Pathways or metabolic functions associated to essential genes. 61
Table 31 Generic metabolites automatically removed from the metabolic network. Reactions associated with these compounds were also removed. 86
Table 32 Reaction distribution among the different pathways included in the model. 89
Table 33 Locus Tag of critical genes of the reconstructed model without similarities with *O. europaea*. Locus Tags represented in bold are exclusive to the biofilm predictions. 93

ACRONYMS

A

ABC ATP-Binding Cassette.

ACP Acyl-Carrier Protein.

ATP Adenosine Triphosphate.

B

BIGG Biochemical, Genetic and Genomic.

BLAST Basic Local Alignment Search Tool.

BRENDA BRaunschweig Enzyme Database.

C

COA Coenzyme A.

COBRA Constraint-Based Reconstruction and Analysis.

D

DNA Deoxyribonucleic Acid.

DSF Diffusible Signalling Factor.

E

EA Evolutionary Algorithms.

EC Enzyme Commission.

EFM Elementary Flux Modes.

EXPASY Expert Protein Analysis System.

F

FBA Flux Balance Analysis.

G

GPR Gene-Protein-Reaction.

GSM Genome-Scale Metabolic.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LPS Lipopolysaccharides.

M

MERLIN Metabolic Models Reconstruction Using Genome-Scale Information.

MFA Metabolic Flux Analysis.

MOMA Minimization of Metabolic Adjustment.

N

NCBI National Center for Biotechnology Information.

O

ORF Open Reading Frame.

P

PGDBS Pathway/Genome Databases.

PTS Phosphotransferase System.

R

RNA Ribonucleic Acid.

ROOM Regulatory On/Off Minimization.

RRNA Ribosomal Ribonucleic Acid.

S

SA Simulated Annealing.

SBML Systems Biology Markup Language.

T

TC Transport Classification.

TCDB Transporter Classification Database.

TMHMM TransMembrane Prediction using Hidden Markov Models.

TRANSYT Transport Systems Tracker.

TREMBL Translated EMBL Nucleotide Sequence Data Library.

TRIAGE Transport Proteins Annotation and Reactions Generation.

U

UNIPROT Universal Protein Resource.

UNIPROTKB UniProt Knowledgebase.

1

INTRODUCTION

The current chapter describes the context and motivation, goals and the structure of this thesis.

1.1 CONTEXT AND MOTIVATION

The development of novel high throughput technologies leads to an increasing number of sequenced genomes each year. Systems biology approaches are extremely meaningful nowadays, to keep in pace with the demand of organizing and extracting information from these genomes.

Genome-scale metabolic (*GSM*) models comprise both genomic and metabolic information of an organism and are used to predict its phenotypical behavior in a broad range of environmental and genetic conditions [1]. Therefore, these are valuable computational tools that improve metabolic engineering strategies formulated to optimize strains and discover candidate drug targets [1, 2].

Xylella fastidiosa is a xylem-limited bacterium known to cause a wide variety of diseases that affect plants with high economic value [3]. Pierce's Disease of grapevine and citrus variegated chlorosis have a high economic impact in the production of these crops and are caused by this gram-negative phytopathogen [4, 5]. Despite the similarities with *Xanthomonas* species and other related plant pathogens, *X. fastidiosa* has a distinct way of regulating its pathogenic machinery [6]. This phytopathogen is able to colonize both plants and xylem-fluid feeding insects, being the latter the mechanism of spreading the pathogen between plants [6, 7].

There have been reports of severe outbreaks of Olive Quick Decline Syndrome caused by this pathogen, leading to the death of thousands of olive trees in Southern Italy [8]. Also, according to the European Commission, *X. fastidiosa* has been detected in multiple plants across France and Spain [9, 10]. In the beginning of 2019, the phytopathogenic bacterium was detected for the first time in the north of Portugal [9]. Thus, the organism could impose as a major threat for various agricultural crops, including the Portuguese olive market, if it keeps spreading across Portuguese territory.

As *X. fastidiosa* is such a destructive pathogen and leads to huge economical losses, it is extremely relevant to build a highly curated GSM model of the organism that will unveil this pathogen metabolic traits, accurately predict its response to different environments and lead to the discovery of new ways to fight the spread of this pathogen. This thesis is inserted in the project Biodata.pt, a Portuguese distributed infrastructure for biological data.

1.2 GOALS

The main goal of this work is to build a genome-scale metabolic model of the bacterium *X. fastidiosa*. From this work it is expected that new knowledge regarding this pathogens metabolism is unveiled, helping to eradicate this pathogen.

The following steps need to be fulfilled to achieve this goal:

- Obtain a high-quality functional annotation of the organism's genome;
- Generate a draft of the metabolic network;
- Perform refinement and manual curation of the network using information from literature;
- Convert the metabolic network to a stoichiometric model;
- Validate the reconstructed model using experimental data;
- Analyze the organism's metabolic network to find potential drug targets.

1.3 STRUCTURE OF THE DOCUMENT

The remaining document has the following structure:

- **Chapter 2** - State-of-the-art:
  - Overview of GSM models reconstruction methodologies
  - Computational tools used for GSM model reconstruction
  - Description of *Xylella fastidiosa* metabolism, life cycle and *in vivo* growth
- **Chapter 3** - Materials and Methods:
  - Software used throughout the work
  - Description of genome annotation routine
  - Methodology used to assemble an initial draft metabolic network

  - Methods used to establish Biomass and Energy Requirements
  - Methods used to manually curate the developed network
  - Strategies used to validate the model
  - Methodologies employed to analyse potential drug targets

- **Chapter 4** - Results and Discussion:
  - Results obtained for the genome annotation procedure
  - Results of the generated transport reactions
  - Description of biomass and energy requirements
  - Changes made to the model during manual curation
  - Results obtained during model validation
  - Description of the developed metabolic network
  - Overview of the identified drug targets

- **Chapter 5** - Conclusion and Future Works:
  - Brief model analysis
  - Applications of the reconstructed model
  - Future work to improve model and its applications

2

STATE-OF-THE-ART

The following chapter provides state-of-the-art methods on GSM model reconstruction and computational tools that aid this process. Description of the metabolism, life cycle and *in vivo* growth of the pathogen *Xylella fastidiosa* are included in this chapter.

2.1 GENOME-SCALE METABOLIC MODELS

2.1.1 *Background*

Novel high-throughput technologies in genome sequencing allow us to gain comprehensive data on biological components of the cell. Interactions between these components enable the construction of a biochemical network that represents the cell as a system. Thus, systems biology's main focus is understanding such systems' structures and dynamics, which can provide useful insight of an organism or cell metabolism [11, 12].

Bioinformatics tools and systems biology approaches turned the reconstruction of *GSM* models into a straightforward process, even for organisms with less literature-based information [13]. The time necessary to reconstruct a high-quality GSM model varies depending on the organism's complexity, going from months, for well-studied bacteria, to years, in a human metabolism reconstruction [14].

A GSM network is described as a set of biological reactions obtained from the enzymes encoded in an organism's genome. Specific information needs to be collected to reconstruct a metabolic network. For each enzyme encoded in the genome, it is necessary to know which substrates and products the enzyme acts on, the stoichiometry of the reactions catalyzed by the enzyme as well as the reactions reversibility and localization [13]. Although GSM networks can help in the determination of numerous properties of cells, only GSM models are able to predict metabolic capabilities of an organism. These models should contain kinetic and regulatory information whenever possible. However, this data is usually only available for well-studied organisms. Nevertheless, GSM models usually include information on biomass composition and energy requirements [1].

The GSM model's reconstruction is an iterative process, as demonstrated by the refinement of the metabolic network of *Escherichia coli* [15] and *Saccharomyces cerevisiae* [16, 17, 18] over the years. The iterative process comprises several steps, as represented in figure 1 (adapted from [19]). Initially, the genome annotation should be performed, which gives the necessary information to build a draft metabolic network. From there on, network refinement is mandatory, and the conversion of the network into a GSM model is achieved by adding the biomass production equation and other constraints. Finally, the model is validated by comparing its predictions to experimental data. If simulations do not corroborate experimental data, further debugging is necessary. The final model should be exported in a format that most tools accept, for example the *Systems Biology Markup Language (SBML)* [20].

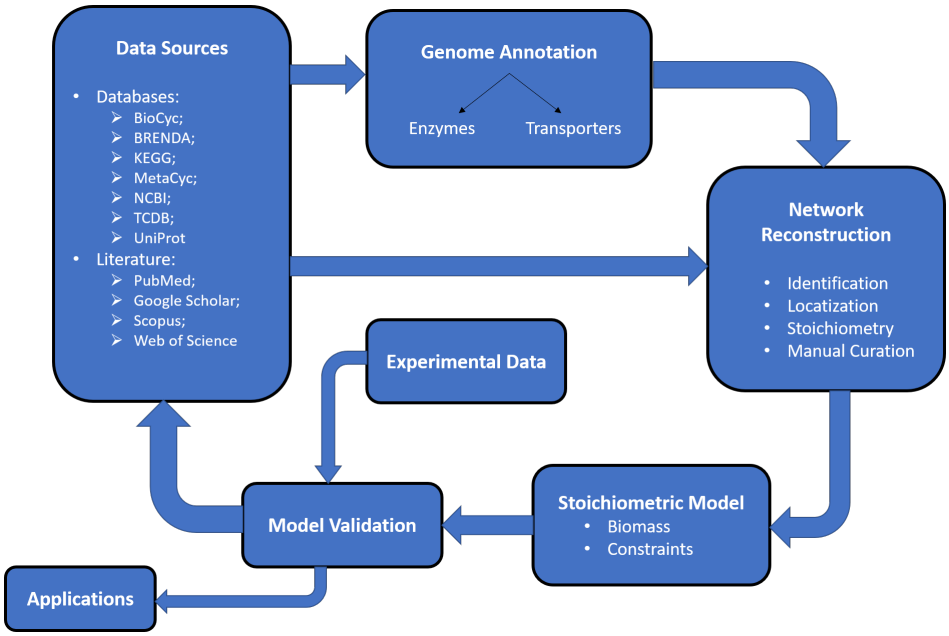


Figure 1.: GSM model reconstruction iterative process. The reconstruction is initiated with a review of available knowledge on the target organism from various sources. A metabolic network is reconstructed and reviewed before conversion into a stoichiometric model. At last, *in silico* predictions are compared to experimental data. If similar, the model can be used in wide range of applications (adapted from *Dias, O.* [19])

During the GSM model reconstruction process several databases are accessed to obtain up-to-date information of the target organism. Information from these sources, such as whole-genome sequences, whole-genome annotations, biochemical data of metabolic reactions and functional information of enzymes encoded in the genome, support the development of the metabolic network as well as its refinement.

Table 1 shows the most utilized online data sources used in GSM model reconstructions and a description of what information can be retrieved from each one.

Table 1.: Online data sources used in GSM model reconstruction

Database	Description	Reference
BiGG Models	*Biochemical, Genetic and Genomic (BiGG)* Models is a knowledgebase of genome-scale metabolic network reconstructions. Over 70 published GSM networks are included in this database. Additionally, a set of standardized identifiers called BiGG IDs are used.	[21]
BioCyc	BioCyc is a collection of organism-specifc *Pathway/Genome Databases (PGDBs)*. Each PGDB contains the genome and the predicted metabolic network of an organism. Also, it includes transport systems and pathway gap-fillers. The website of this database has tools to view and analyze PGDBs.	[22]
BRENDA	*BRaunschweig Enzyme Database (BRENDA)* is the main database of enzyme and enzyme-ligand information. A collection of manually curated enzymes is available within this database. In addition, the *Enzyme Commission (EC)* classification system is used on this database.	[23]
ExPASy	*Expert Protein Analysis System (ExPASy)* is one of the main bioinformatics resources for proteomics.	[24]
KEGG	*Kyoto Encyclopedia of Genes and Genomes (KEGG)* is an integrated database that comprises three sections. These are systems information, genomic information and chemical information. This repository can be used for biological interpretation and other high-throughput technologies.	[25]
MetaCyc	MetaCyc is a freely accessible database of curated metabolic pathways. It contains information on metabolic pathways and enzymes from a wide array of organisms.	[22]
NCBI	*National Center for Biotechnology Information (NCBI)* is an online repository composed by several databases. Through the NCBI website, biomedical and genomic information can be retrieved.	[26]

TCDB	The *Transporter Classification Database (TCDB)* contains structural, functional, mechanist, evolutionary and medical information of transporter proteins. This information is freely accessible and each transporter protein is assigned to a *Transport Classification (TC)* number.	[27]
UniProt	The *Universal Protein Resource (UniProt)* is a comprehensive resource for protein sequence and annotation information. UniProt is divided in three databases, being *UniProt Knowledgebase (UniProtKB)*, one of them. The UniProtKB is fragmented in two sections, TrEMBL (Translated EMBL Nucleotide Sequence Data Library), that contains automatically annotated and non-reviewed information, and Swiss-Prot, which comprises manually annotated and reviewed data.	[28]

In recent years, bioinformatics tools that support models reconstruction or provide network visualization features have been developed to automatize several steps of the reconstruction process, reducing the time necessary to reconstruct a model. Tools such as *merlin* [29], KBase [30], ModelSEED [31], CarveMe [32], Cytoscape [33], MicrobesFlux [34], Pathway Tools [35] and CellDesigner [36], among others, contain some of these features.

Currently, GSM models are used to identify the metabolic capabilities of organisms and are useful to foresee a microbial response to different environmental conditions. In addition, GSM models have applications within the metabolic engineering scope, such as strain optimization. Gene deletion analysis can also help in the discovery of new drug targets.

2.1.2 *Genome Annotation*

The reconstruction of a GSM model always starts with the genome annotation of the organism of interest. Reliable annotations are crucial to a high-quality reconstruction, thus retrieving updated genome annotations ensures that the latest corrections are accounted for [2]. Online databases, such as NCBI and KEGG can be used to download whole genome annotations.

Occasionally, annotations are neither available nor trustworthy. In these cases, a functional annotation has to be performed using sequence similarity tools. Sequence alignment algorithms such as *Basic Local Alignment Search Tool (BLAST)* [37] or HMMER [38] are largely

utilized in this stage of the reconstruction. These tools determine which genes from different organisms or strains are significantly analogous to a given gene. Usually, homologous genes have similar functions and such function should be assigned to the unannotated gene.

Important data such as gene or *Open Reading Frame (ORF)* name, cellular function(s) and EC or TC numbers (if available) should be retrieved for each gene annotation. EC and TC numbers are extremely useful to an automatic reconstruction. Enzyme and transport systems encoding genes, called metabolic genes, are indispensable for the model reconstruction. Still, other potential metabolic genes or regulation/signaling associated genes should be thoroughly revised as they can be associated with missing reactions or be useful for integration within the network.

2.1.3 *Metabolic Network Reconstruction*

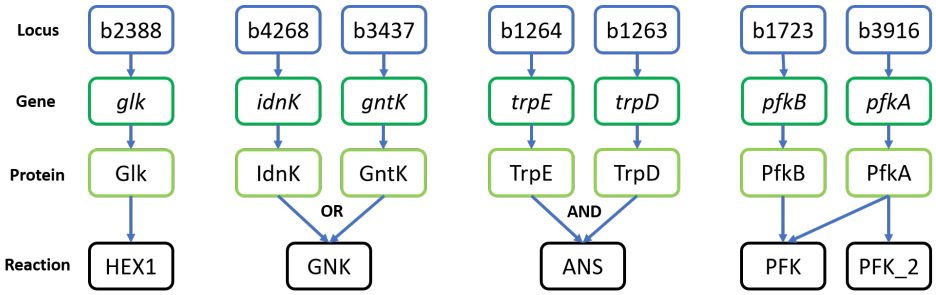
The second stage on a GSM model reconstruction consists in assembling a network of biochemical reactions. This stage comprises several steps, starting with building the backbone of the metabolic network and finishing with refinement of the latter.

2.1.3.1 *Genes, Proteins and Reactions Associations*

To initiate the metabolic network assembly process, each metabolic gene is associated with the corresponding protein(s) and reaction(s), in a process called *Gene-Protein-Reaction (GPR)* associations. Essentially, GPR associations indicate the function of the encoded enzyme and are usually represented using *AND* or *OR* Boolean rules. Different types of GPR associations are represented in figure 2 (adapted from [2]). Isozymes, promiscuous enzymes (multiple functions) and protein complex are special cases of GPR associations and should be revised carefully when setting its Boolean rule. For instance, the Boolean rule of an enzyme complex should comprise all the subunits using the operator *AND*. Another special case to consider are lumped reactions, that occur frequently in linear pathways. In this situation, the Boolean rule should be similar to an enzyme complex, reflecting the requirement of all genes associated with the reaction [2].

The EC and TC numbers previously assigned in genome annotation are highly useful in the GPR associations step, as these can be used to retrieve both enzymatic and transport reactions. Information regarding enzymatic reactions can be obtained in online databases, such as KEGG and BRENDA, through the EC numbers. Additionally, the TCDB database is used to collect data on transport reactions through the TC numbers.

A system for enzyme nomenclature was published in 1992 that classifies enzymes based on the reactions that they catalyze instead of their structure, as in previous approaches [39, 40]. This system, known as the EC system, identifies the enzymes with a classification number made up of four digit code separated by dots. The first digit can go from 1 to



Reaction	Reaction Name	EC number	GPR
HEX1	Hexokinase	2.7.1.1	(b2388)
GNK	Glucokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)

Figure 2.: Example of GPR associations and its assigned Boolean rules for *Escherichia coli*. These rules can represent different situations, as represented (from left to right), one enzyme one reaction, isoenzymes, enzymatic complexes and promiscuous enzymes (protein PfkA). Adapted from *Thiele, I. and Palsson, B.*[2]

7 and it represents the class of the enzyme. These classes are oxireductases, transferases, hydrolases, lyases, isomerases, ligases and the recently added translocases [23]. The remaining numbers progressively specify the enzyme function. Also, proteins involved in signal transfer, *Ribosomal Ribonucleic Acid (rRNA)* modification and *Deoxyribonucleic Acid (DNA)* methylation are associated with EC numbers but should not be considered to the reconstruction of the network.

A similar approach to the EC system uses a code to represent a transport protein. However, unlike the enzymes system, the TC system characterizes the transport proteins by its structure and function [27]. The code used in this system is composed by a five character identifier: N.L.N.N.N, where N represents a number and L a letter. As in the EC system, the first element of the identifier represents one of seven classes of transporters. Channels/pores, electrochemical potential driven transporter, primary active transporters, group translocaters, transmembrane electron carriers, auxiliary transport proteins and putative transport system that are incompletely characterized are the seven classes represented in the TC system.

After retrieving available information from the previously mentioned databases, data on genes names/identifiers, proteins names, EC and TC numbers, biochemical reactions substrates and products, the equation of these reactions as well as the stoichiometry associated with each metabolite, should be compiled, and if possible, automatically. Finally, the main set of reactions that build up the network is assembled.

2.1.3.2 *Completing the draft network*

To finish the initial draft of the network, spontaneous and non-enzymatic reactions should be included into the network. These reactions can be obtained from online public databases, such as KEGG, or published literature. Furthermore, the draft network should be complemented with organism-specific and literature-based reactions.

2.1.3.3 *Compartmentalization*

The reactions integrated in the network should be assigned to the compartment where they take place. Thus, the next step in a reconstruction is the compartmentalization of the reactions. Assigning enzymes to the organelle in which they work is an important part in the development of a metabolic network and, consequently, a GSM model.

Organism complexity represents a relevant characteristic when determining the location of an enzyme. Compartments in prokaryotic organisms are usually limited to extracellular space, cytoplasm and, occasionally, periplasmatic space. However, eukaryotic cells have higher complexity and reactions may take place in different organelles, namely: lysosome, Golgi apparatus, mitochondrion, peroxisome, nucleus and endoplasmic reticulum. In addition, it may be necessary to differentiate tissues in higher eukaryotes.

Another detail to consider when compartmentalizing the network is to distinguish similar reactions (same metabolites and stoichiometry), though present in different compartments, as distinct reactions. Thus, it is essential to assign distinct identifiers to each reaction, reflecting its correspondent compartment. Hence, a metabolite in distinct compartments should have different identifiers.

Bioinformatics tools, such as PSORTb [41], WoLF PSORT [42], TargetP [43], LocTree3 [44] and PASUB [45], are used to predict the subcellular location of enzymes based on their amino acid sequences and physiological characteristics of the respective organism. Additionally, information on protein localization can also be obtained in literature and online databases such as UniProt. When in doubt about the enzyme localization, the corresponding reaction(s) should be assigned to the cytoplasm to avoid connectivity issues within the network.

2.1.3.4 *Manual Curation and Network Refinement*

Despite the time saving and usefulness of automatic tools, such approaches are prone to errors [46]. Thus, the final step of a GSM network reconstruction is its manual curation, which unlike the automated processes in previous stages of reconstruction, can be a slow and tedious work. Thus, in this stage, an exhaustive revision of published literature is mandatory for the validation of the reaction set. Publications, textbooks and organism-specific databases can be useful assets in this process. If no organism-specific information

is available, data on phylogenetically close organisms can be used. Additionally, comparative genomics can accelerate curation and reconstruction procedure [47], by comparing the metabolic network with curated models of closely related organisms.

Several issues within the network are solved in the manual curation step, namely, the metabolic potential (presence or absence of reactions), reaction reversibility and directionality, gap filling, substrate/cofator usage and stoichiometry.

The reconstructed network should represent accurately the physiology of the organism, thus the presence or absence of pathways or reactions must be confirmed. Organism-specific reactions found in literature should be included and peculiar reactions with no evidence on literature should be removed. Furthermore, promiscuous enzymes have the ability to catalyze multiple reactions, but only reactions known to take place in the target organism should be incorporated in the metabolic network. Reactions assigned with ambiguous identifiers, such as partial EC numbers [46], should be corrected.

Incorrect reversibility and directionality may affect considerably the model's performance. An excessive number of reversible reactions can lead to futile cycles that can overcome proton gradient through metabolite and proton exchange among compartments [2]. Biochemical studies of enzymes or the recent approach of estimating the standard Gibbs free energy [48, 49, 50] could be a solution to overcome these issues.

A metabolic gap in a GSM network is a missing reaction in a given pathway [46]. The existence of these gaps can lead to the accumulation of compounds during a simulation. Consequently, reactions downstream of such pathway will be blocked, due to the absence of the enzyme required to produce its substrate. However, missing reactions should only be included in the model if supporting evidence regarding the reaction's presence in the organism is available, or if the reaction is required to ensure the model's performance [2]. KEGG pathways can be useful during the gap-filling analysis.

Enzymes' substrate and cofactor specificity of enzymes can vary among different organisms. Incorrect inclusion of cofactors and substrates can impair the model's predictive potential [2]. Non-organism-specific databases, such as KEGG, provides all substrates for an enzyme. If a protein is associated to a single reaction in these databases, no refinement is required. Contrarily, enzymes associated with multiple reactions require evaluation. Information on substrate and cofactor utilization can be retrieved from biochemical organism-specific studies or in organism-specific databases.

Unbalanced reactions lead to imprecise flux distribution predictions through the GSM network. Therefore, the reactions' stoichiometry must be revised to avoid this problem. The sum of all the elements in the compounds should be equal in both sides of the equation.

2.1.4 *Stoichiometric Model*

The final stage of a GSM model reconstruction is the conversion of the metabolic network to a stoichiometric model. In other words, the network is converted to a stoichiometric matrix and constraints are added. However, before the conversion, an equation representing the biomass formation must be included in the reaction set. This pseudo-reaction must describe the all known biomass components as well as their contribution to the cell biomass [2]. However, usually, the biomass precursors require biosynthesis reactions unavailable in the model. Thus, such reactions should be included in the reaction set.

Equation 1 represents the biomass formation equation.

$$\sum_{k=1}^{p} c_k.X_k \rightarrow biomass \tag{1}$$

Where:

- c_k represents the metabolite coefficient
- X_k represents a metabolite

The flux through this reaction represents the growth rate of the organism. To perform an accurate prediction on the growth rate of an organism, a detailed composition on biomass components should be obtained. This composition can be determined experimentally, as described in some studies [51, 52, 53], or estimated using organism-specific literature. If this data is unavailable, the biomass composition of closely related organisms can be used. Special care is required when building this reaction as it has an important role in *in silico* simulations.

Additionally, energy requirements must be included in the model. These requirements can be divided into two sections: growth-associated energy requirements and non-growth energy requirements.

Growth-associated energy is related to a variety of cellular processes, such as the polymerization of amino acids, proteins, nucleotides, among others. These requirements are represented in *Adenosine Triphosphate (ATP)* molecules needed per gram of biomass produced and can be determined by chemostat growth experiments or estimated by the energy required to synthesize macromolecules. Another option is to use published organism-specific data or from phylogenetic close organisms. Thus, these energy costs should be included in the biomass formation reaction.

Additionally, non-growth energy requirements also need to be represented. A simple ATP hydrolysis reaction should be included in the reaction set to represent the non-growth energy costs [2]. An example of these energy necessities is the maintenance of cell turgor

pressure [54]. Such as in the growth-associated energy requirements, non-growth energy requirements can be found in published literature or determined by experimental data.

After including the biomass equation and energy requirements in the metabolic network, the next step is the conversion of the reaction set into a stoichiometric matrix. The matrix is assembled according to chemical engineering classic principles, to represent the dynamic behavior of metabolite concentration, using ordinary differential equations. Equation 2 represents the rate of change of the metabolites concentration over time.

$$\frac{dX_i}{dt} = \sum_{j=0}^{N} S_{ij}.v_j + \mu X, i = 1, ..., M \tag{2}$$

Where:

- X_i represents the concentration of metabolite i
- v_j is the metabolic flux of reaction j
- S_{ij} is the stoichiometric coefficient of metabolite i in reaction j
- μX represents the growth rate of the system

Although dynamic models can predict concentration profiles over time, they require kinetic rates for all reactions. However, nowadays it is still not possible to determine and collect genome-wide kinetic rates and parameters. Hence these models employ the steady-state approximation approach , which assumes that metabolite concentration remains the same throughout time. Therefore, consumption rates are equal to production rates, as show in equation 3.

$$S.v = 0 \tag{3}$$

Where:

- S is the stoichiometric matrix, where each column represents a metabolite and each row a reaction
- v is the flux vector

As most metabolic networks are undetermined systems, an infinite number of solutions may satisfy the imposed constraints. This solution space is called the null space of S [55]. Therefore, it becomes impossible to compute a single solution. To overcome this, new constraints must be included to reduce the null space of S. With these restrictions, a set of feasible solutions called the flux cone of solutions is obtained.

These new constraints are related to the reactions' reversibility. Reversible reactions are constrained between minus infinity and plus infinity. Irreversible reactions can be restrained in one of two ways, depending on the directionality of the reaction. The constraint can be between zero and plus infinity or minus infinity and zero.

Likewise, transport fluxes should be constrained depending on the reversibility and directionality of the reaction. Additionally, exchange reactions already included in the model should be constrained based on specific uptake rates, medium availability and possible excretion. For instance, an exchange reaction associated to a limiting substrate should be constrained to a specific uptake, a metabolite unavailable in the medium should be constrained to zero and metabolites that can be produced should have an unconstrained upper bound.

Figure 3 is an example of the conversion of a metabolic network into a metabolic mathematical model.

The final stoichiometric model can be exported in the SBML format, which is a largely utilized format in simulation tools such as Optflux [56] or *Constraint-Based Reconstruction and Analysis (COBRA)* toolbox 3.0 [57].

2.1.5 *Model Validation*

The final step of a GSM model reconstruction consists in validating the previously generated stoichiometric model. The model is used to predict the organism behavior and it is compared to organism-specific experimental data. Although, if this data is not available, experimental data of closely related organisms can be used. As previously stated, if the predictions do not corroborate experimental data, a careful revision is necessary to find errors. Therefore, all steps of the reconstruction process should be performed carefully, preferably with knowledge on the organism's metabolism and genetics.

Flux Balance Analysis (FBA) is a common approach to study GSM models. This approach is based on linear programming, a methodology used to optimize a linear objective function subject to constraints. In GSM models, restrictions are imposed in two ways: by the stoichiometric matrix, which forces mass balance constraints on the system, and by reactions bounds, which define the minimal and maximum fluxes to each reaction. These constraints define a space of possible flux distributions of a system. The system of linear equations that represents mathematically the objective function and metabolic reactions, can be solved, in FBA, using linear programming. Therefore, FBA can be used to predict the growth rate of an organism by using a biomass production reaction as the objective function, as studies indicate that organisms under various carbon-limiting conditions maximize the biomass production [58, 1]. Other objective functions can be used in simulations, such as maximizing the production of a specific compound [59].

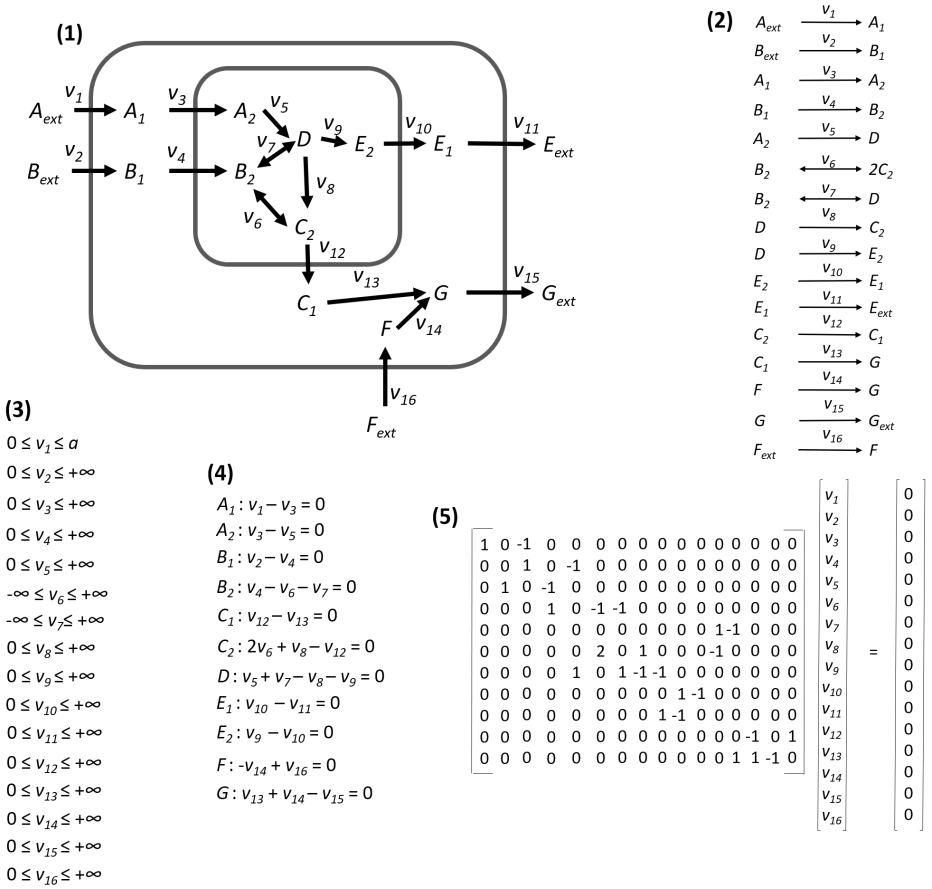


Figure 3.: Illustration of a metabolic network composed by seven metabolites (A-G) and 16 reactions (v1-v16). (1) is a schematic view of the system, where v1-v4, v10-v12 and v15-v16 represent exchange fluxes. Additionally,the reaction reversibility and direction are represented with the arrows (reversible reaction - double arrow; irreversible reaction - forward arrow). (2) shows the stoichiometry of reactions. (3) applied constraints to the fluxes. (4) steady-state mass balances. (5) stoichiometric matrix. Adapted from *Dias, O.* [19]

Model validation analyses include high-throughput growth phenotyping data and *in silico* simulations for various carbon, nitrogen and other nutrient sources. If predictions are not in accordance with experimental data, the model should be re-evaluated.

Another strategy in model validation is comparing predicted specific growth rates and by-product formation on different environmental conditions with information described in literature. Experimental environmental conditions can be simulated by constraining exchange reaction bounds. This information can also be used to calibrate the model, for instance regarding the energy requirements.

Gene deletion simulations are also used for the model validation. Although FBA can be used for knockout simulations, it is not the most correct approach. From an evolutionary standpoint, the optimality of a wild-type organism is credible, however, the same may not be valid for organisms that suffered gene mutations or knockouts. Therefore,

methods such as *Minimization of Metabolic Adjustment (MOMA)* [60] and *Regulatory On/Off Minimization (ROOM)* [61] are recommended for gene deletion simulations. In MOMA, it is assumed that the metabolic fluxes of a knockout are subjected to a minimal redistribution related to the flux vector obtained in a wild-type simulation, whereas in ROOM the goal is to minimize the number of significant flux changes with respect to the wild-type. The predicted phenotypical behavior of the organism, when a gene is deleted, should comply with experimental data, otherwise the model should be reviewed.

2.2 COMPUTATIONAL TOOLS

2.2.1 *merlin*

The Metabolic Models Reconstruction Using Genome-Scale Information (*merlin*) tool is an in-house developed application that supports all steps of the GSM models reconstruction process. This user-friendly framework is divided in the annotation module and the models reconstructions module.

merlin is the main tool used in this work, thus a detailed description of this software's capabilities is detailed in the following sections.

2.2.1.1 *Annotation Module*

merlin's annotation module allows performing enzymes, transporters and compartments annotations.

The enzyme's annotation section assigns enzymatic functions to proteins encoded in the genome, using homology search tools. Currently, *merlin* supports two algorithms, BLAST and HMMER. *merlin* is able to retrieve relevant information from each homologous gene retrieved from the similarity searches, such as the score, locus identifier, e-value and organism. Additionally, NCBI and UniProt are accessed to download, for each gene, the following information: taxonomy, organelle, chromosome, locus tag, EC number, protein name and molecular weight.

Using a specific algorithm, *merlin* calculates a confidence score and is able to assign to each gene an EC number and a product name. This algorithm considers two aspects:

- Frequency of each EC number or product name obtained from the set of homologous genes;
- The taxonomy of the organism in a given homologous gene.

Therefore, the annotation score is calculated using equation 4.

$$Score_{an} = \alpha \times Score_f + (1 - \alpha) \times Score_t \tag{4}$$

Where:

- $Score_{an}$ is the annotation score;
- $Score_f$ is the frequency score, calculated using the number of occurrences of an EC number/product name and the total number of homologous genes;
- $Score_t$ represents the taxonomy score, obtained from the phylogenetic proximity between the homologous gene source organism and the target organism;
- The α parameter changes the weights of the frequency and taxonomy scores. This value can be changed by the user.

A recently developed methodology to identify and annotate transport systems is implemented in *merlin*, the *Transport Proteins Annotation and Reactions Generation (TRIAGE)* tool [62]. This tool performs the automatic annotation of transport proteins with TC numbers and generates transport reactions for all metabolites carried by the transporters.

To annotate a transporter, TRIAGE follows three principles:

- Gene sequences must contain transmembrane domains;
- Gene sequences need to have similarities with TCDB records;
- The localization prediction of the protein must be a membrane.

The first step is to submit the genome amino acid FASTA file to the *TransMembrane Prediction using Hidden Markov Models (TMHMM)* [63, 64] or Phobius [65, 66] web servers, to identify proteins with transmembrane helices. Afterwards, protein sequences with transmembrane helices are compared to protein sequences available in TCDB, using the Smith-Waterman or BLAST algorithms [67]. For each homologous protein found in TCDB, relevant data is retrieved, such as the UniProt accession number, protein name, sequences length, amongst others. An annotation scoring system similar to the previously described for enzymes' annotation (Equation 4) is used to determine and assign metabolites and TC numbers. Finally, if the predicted compartment for a transport system is a membrane, all three criteria of TRIAGE are fulfilled and the transporter is annotated.

The compartments annotation is supported by three compartments' prediction tools, namely WoLF PSORT, PSORTb 3.0 and LocTree3. These tools predict the localization of proteins in eukaryotes, prokaryotes and both, respectively. A confidence score is provided for each compartment associated to a protein. *merlin* is able to automatically assign each protein to the main compartment . However, if prediction scores between different compartments are inferior to a user defined percentage (default 10%), the protein is assigned to multiple compartments and reactions catalyzed by such enzyme are associated with all compartments.

2.2.1.2 *Models Reconstructions Module*

merlin's second module allows loading metabolic data into the internal database, which is used to build the model. KEGG is currently the main source of metabolic data used to reconstruct the model, and along with the enzyme annotation it creates a draft of the metabolic network.

Transport reactions and associated GPR associations generated in the transporter's annotation are also integrated in the model. However, only transport reactions with metabolites already present in the metabolic network are included in the model. The draft model contains intracellular and extracellular reactions, originating a compartmentalized model.

This module also provides genes, proteins, metabolites, reactions and pathways views, and additional tools that aid in the manual curation process. These tools are the following:

- Unconnected reactions and metabolites finder tool;
- Unbalanced reaction finder tool;
- A tool that corrects reaction reversibility;
- The e-biomass tool, which helps in the formulation of the biomass composition;
- A drain/exchange reaction generator that automatically creates an exchange reaction for each extracellular metabolite;
- A tool that generates GPR rules based on KEGG Orthology information.

merlin is also able to export the model in a SBML format, in order to be used in other applications, such as OptFlux and COBRA toolbox.

2.2.2 *Simulation tools*

The simulation tools used in this work, OptFlux and COBRA toolbox 3.0, are described in the following sections.

2.2.2.1 *OptFlux*

OptFlux is an open-source software used for *in silico* metabolic engineering [56]. This software contains several tools and methods to analyze metabolic models. OptFlux features are presented in the following list:

- Phenotype simulation, such as wild-type, gene-deletions and reaction-knockouts predictions, using FBA, MOMA or ROOM;
- Methods for *Metabolic Flux Analysis (MFA)* [68];

- Elementary Modes Analysis by calculating *Elementary Flux Modes (EFM)* [69];
- Strain optimization methods, using *Evolutionary Algorithms (EA)* [70], *Simulated Annealing (SA)* [71] and the OptKnock algorithm [72];
- Metabolic network visualization tool.

2.2.2.2 *COBRA toolbox*

COBRA toolbox is a software suite of COBRA methods [57]. The COBRA toolbox was released as a package for MATLAB [73] and since, two major updates were released, each one with more available tools [74, 57]. Additionally, COBRA methods have been released in a Phyton package, COBRApy [75]. These methods have been used in a wide range of fields [76, 77, 78, 79, 80], being especially important for metabolic engineering [81, 15, 82].

COBRA provides many features to analyze biochemical networks, which makes COBRA an useful resource for metabolic reconstructions. This software contains several methods to aid in the following analysis of a metabolic network:

- FBA;
- Gap-Filling;
- Network Visualization;
- Strain Engineering;
- Dead-end metabolite identification;
- Data integration;
- Gene essentiality.

2.3 XYLELLA FASTIDIOSA

2.3.1 *Background*

Xylella fastidiosa, first proposed as a novel species and described in 1987 [4], is a gram-negative, nonmotile, aflagelate, strictly aerobic, nonfermentative, xylem-limited [3] phytopathogenic bacteria. These bacteria are nutritionally fastidious, as indicated by its name, which means it requires a specialized medium to grow [4]. Despite the fastidious nature of the organism, its genome indicates a very efficient energy metabolism, since it contains all genes required for the glycolytic pathway, tricarboxylic acid cycle and electron transport

chain. Additionally, different carbon sources, such as glucose, fructose, mannose and glycerol may be used in glycolysis. To complement the poor nutrient environment of xylem vessels, the bacterium is also able to hydrolyze cellulose into glucose, which suggests that cellulose may be used as a monosaccharide source while the organism survives within the host xylem cells [3]. The xylem is an essential specialized tissue responsible for the transport of water and minerals in vascular plants [83]. Xylem vessels, the tissue infected by *X. fastidiosa*, displays a low concentration of sugars, amino acids and organic acids, making it a nutrient-limited environment [84].

Currently, six different subspecies of *X. fastidiosa* have been proposed, namely: *pauca, fastidiosa, multiplex, sandyi, tashke* and *morus* [85, 86, 9, 87]. Unlike other phytopathogens with host specificity [88], *X. fastidiosa* has a wide range of plant hosts, being able to infect 563 plant species, as reported in the *Xylella* spp. host plant database [89]. However, certain subspecies of the pathogen display somewhat of a host specificity and are related to specific plant diseases [90, 91].

Various plant pathogens from the *Xanthomonas* genus are closely related to *X. fastidiosa* as these organisms also colonize xylem cells. However, a major difference separates *X. fastidiosa* from other organisms, which is the organism dispersion mechanism. *X. fastidiosa* is vector-transmitted among plants by xylem-fluid feeding insects, while *Xanthomonas* species spread by splash dispersal and windblown rain [6, 92]. Additionally, compared to Xanthomonads, *X. fastidiosa* is deprived of a Type III secretory system, responsible for suppressing the host plant defensive responses [6, 93].

X. fastidiosa is described as being able to live in 'two Worlds' [6], as it lives within plants and insects during its life cycle. The time span between the infection of the host plant and the appearance of symptoms varies, going from a few months to years, which depends on multiple factors, such as host species, *X. fastidiosa* genotype, growing conditions, among others [9]. However, many host plants remain symptomless despite infection, and may act as reservoirs for the phytopathogen [5].

The complex life cycle of *X. fastidiosa* is regulated by a *Diffusible Signalling Factor (DSF)*. In the early stage of infection, considered to be the 'exploratory phase', the pathogen displays a planktonic state, expresses genes related to twitching motility and extracellular enzymes, which enables movement and colonization of new vessels. As cells multiply, the DSF levels increase, leading to a phenotype switch to an adhesive state. In this stage, genes related to the synthesis of adhesins and exopolysaccharides are upregulated. Afterwards, cells attach to xylem walls and aggregate, forming a biofilm. Nevertheless, some cells remain dispersed from the biofilm, allowing further propagation in the host plant. In this 'sticky phase', the phytopathogen may be ingested by xylem-fluid feeding insects. Upon ingestion, the bacteria adhere to the vector foregut, where it can remain for several months [94]. Once

the insect feeds on a new plant, *X. fastidiosa* is injected to the xylem, restarting its life cycle. Figure 4 depicts the aforementioned life cycle in a simplistic manner.

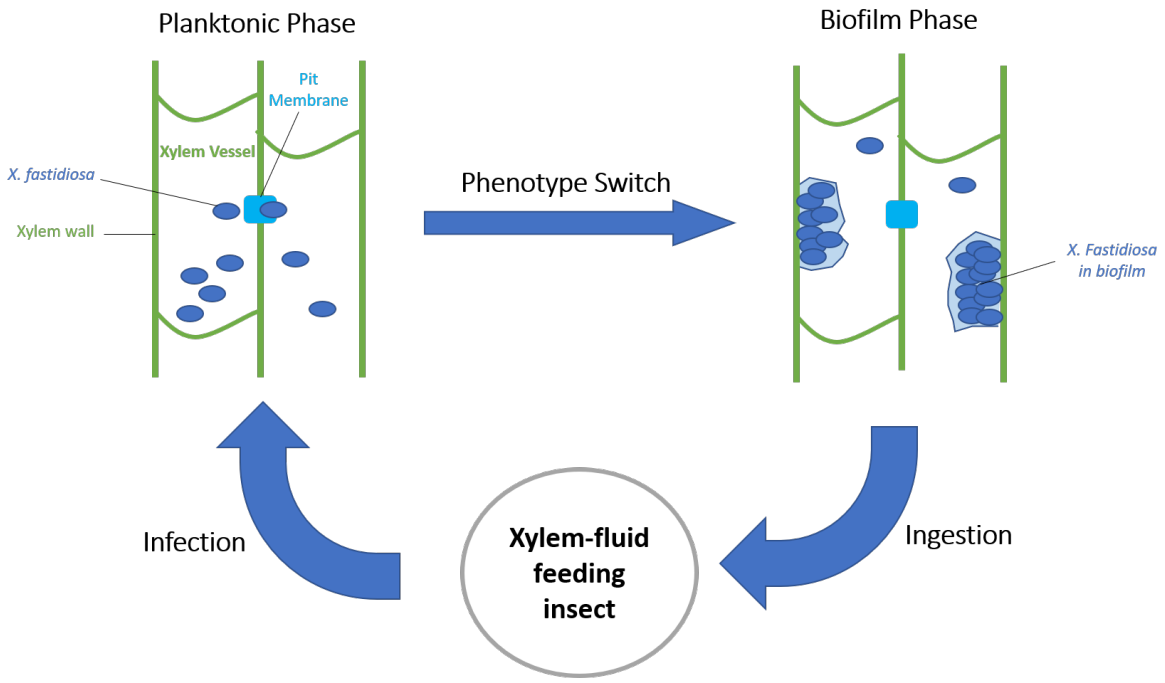


Figure 4.: Illustration of *Xylella fastidiosa* life cycle. After infection, *Xylella fastidiosa* displays a planktonic state within xylem vessels and is able to migrate between them by degrading pit membranes. Proliferation of cells in the xylem increases the diffusible signalling factor which leads to a phenotype switch. During this biofilm phase, the phytopathogen is attached to xylem walls and can be picked by xylem-fluid feeding insects. Finally, this insect is able to infect other plants and the infection cycle restarts.

X. fastidiosa is known to cause disease in a variety of plants with high economic value throughout the world, especially all over the American Continent [95, 96, 97, 85, 86] and Europe [8, 10]. The most known diseases caused by the phytopathogen are the Pierce's Disease of Grapevine [5], citrus variegated chlorosis [97] and olive quick decline syndrome [98], which affects grapes, citrus and olives, respectively. Usually, diseases caused by *X. fastidiosa* are leaf-scorch associated and the symptoms appear during summer days corresponding to when the plant is water stressed [99].

These bacteria are spreading at an alarming pace throughout European territory. In recent years, there have been reports of severe outbreaks of olive quick decline syndrome in southern Italy that lead to the death of thousands of olive trees [8]. In France, coffee plants contaminated by *X. fastidiosa* were detected [10], and, according to the European Commission, several plants were also infected in Spain. Additionally, at the start of 2019, *X. fastidiosa* was detected for the first time in Portugal [9], in lavender plants from Santo Inácio Zoo (Porto).

Despite the global dispersal of *X. fastidiosa*, there is no established cure for diseases caused by the bacteria. Currently, the only measure applied to control *X. fastidiosa* relies on the eradication of the infected plant and adoption of confinement measures in the infection zone. However, some innovative treatments have emerged as being able to control the phytopathogen to some extent. These treatments can be divided by the target organism: host plant, bacterium or insect vector.

At the level of the host plant, pruning [100], cold treatment [101] and the screening of resistant plants [102] can be used to control the pathogen. As far as targeting the bacterium, the efficacy of several compounds has been tested *in vitro* and in fields. Chemicals such as tetracyclines [103], polyphenols [104] and radicinin [105] have an inhibitory activity on the bacterium. Other strategies that target the phytopathogen are symbiotic control [106] and biocontrol using specific phages that can control *X. fastidiosa* growth [107]. Insect vector targeting is also able to control the spread of *X. fastidiosa*. The use of insecticides [108] and the introduction of natural enemies [109] are methods to effectively reduce the population of vector insects.

2.3.2 *In vivo growth*

Traditional media developed for bacterial growth cannot be used to culture *X. fastidiosa*, due to its fastidious nature. Therefore, several media that support the phytopathogen growth were created, namely, PW [110], PD2 [111], BCYE [112], SPW [113], PYE, GYE [114] and 3G10R [115], being the PW medium the most used to cultivate all strains of *X. fastidiosa* [116, 117]. However, growth under different media can lead to differential gene expression and subsequently distinct phenotypes.

PW is a nutrient rich-medium that contains various growth components. The axenic culture of *X. fastidiosa* in this medium displays elevated growth rates with typical bacterial growth curves [118]. This growth rate is associated with the upregulation of genes involved in metabolic pathways important for bacterial growth. A higher growth rate requires increased energy consumption, which can be maintained by a higher ATP production that may result from the overexpression of various genes involved in major steps of the central metabolic pathways. Reactions from the glycolytic pathway, Krebs cycle and respiratory electron transport chain are upregulated to increase ATP production which ultimately leads to increased production of cell components. As the oxidative phosphorylation pathway has increased activity, the O_2 consumption is higher in PW-grown cells. Growth rate in this medium is similar to less complex media such as BCYE, PYE and GYE, which might indicate an identical phenotypical response of the organism in both types of media.

Although these media support the growth of the phytopathogen, these do not represent accurately the nutrient-limited xylem environment. Leite *et al.* [115] developed a

xylem-based, chemical defined medium, called 3G10R. Cells grown in this medium exhibit characteristics usually found during the infection process *in planta*, such as the increased aggregation capacity and biofilm formation. Therefore, this medium is a useful tool when studying the metabolic profile of *X. fastidiosa*. 3G10R cultures display continuous growth, without a detectable bacterial growth curve, identical to cells grown in pure xylem sap [119]. The respiratory activity of 3G10R-grown cells is significantly less when compared with PW-grown cells.

As previously stated, certain genes, especially the ones involved in the infection process of plants, are upregulated when *X. fastidiosa* cells are grown in 3G10R medium. The absence of amino acids in the medium leads to an overexpression of genes associated with amino acid biosynthesis, instead of proteolytic enzymes, as observed in PW-growth. Additionally, while growing in 3G10R exopolysaccharides and adhesion factors synthesis, considered to be virulence factors in *X. fastidiosa* [120, 121], increase leading to intense cellular aggregation, biofilm formation, and higher infective capabilities [118]. Bacteriocin genes are also overexpressed which indicates a competing strategy against other bacteria within xylem vessels [122].

2.3.3 *Xylella fastidiosa subsp. pauca De Donno*

X. fastidiosa subsp. *pauca* causes two of the most known diseases induced by the pathogen, citrus variegated chlorosis [88] and olive quick decline syndrome [98]. Despite the fact that these plants are infected by the same subspecies, citrus variegated chlorosis leads to small, hardened fruits with zero commercial value and usually, infected plants remain alive [123]. However, in olive quick decline syndrome there is an extensive branch and twig dieback which ultimately causes tree death [124].

In 2013, the quarantine pathogen *X. fastidiosa* was detected in southern Italy in olive trees, were the infected trees showed symptoms of olive quick decline syndrome. It was the first outbreak of the phytopathogen in the European Union. The organism was isolated and by 2017, as part of the Pest Organisms Threatening Europe (POnTE) research project, the complete genome sequence of *X. fastidiosa* subsp. *pauca* De Donno was released. Its genome comprises a single chromosome and a circular plasmid, with around 2.5Mb and 30kb, respectively [125].

3

MATERIALS AND METHODS

Chapter focused on the Materials and Methods used in the GSM model reconstruction.

3.1 SOFTWARE

The GSM model was reconstructed using *merlin* platform [29], which provides a friendly interface for the major steps of the reconstruction process. Additionally, it contains several tools that allow a semi-automated reconstruction.

An in house created Python script, based on the COBRApy [75] package, was used to scout potential errors in the model and test the synthesis of biomass precursors. In addition, another script, dependent on the BioPython package [126], was created to identify potential drug targets for *X. fastidiosa*.

In order to validate the model, *in silico* simulations were made using the *Optflux* software [56].

3.2 GENOME FILES

RefSeq assembly genome files of *X. fastidiosa* subsp. *pauca* De Donno, accessible via the NCBI assembly accession number ASM211787v1 [125], were retrieved automatically by *merlin*, using the organism NCBI [26] taxonomy ID (698414) within the framework.

3.3 GENOME ANNOTATION

Upon retrieving the genome files within *merlin*, it was possible to perform the enzyme annotation. For each of the coding sequences found in the genome, a homology search using BLAST [37] was performed. Since *merlin* stores and provides a friendly interface to review annotations, it was important to gather high-quality homologies. Therefore, a maximum *e-Value* of 1E-30 was selected along with a maximum of 100 hits per sequence. Additionally, the substitution matrix was automatically selected based on the sequence length. With

these parameters, the homology search was run against two databases, UniProt *Translated EMBL Nucleotide Sequence Data Library (TrEMBL)* and UniProt Swiss-Prot [28].

As aforementioned in section 2.2.1.1, *merlin* automatically assigns a product name and an EC number (if available) based on a score calculated using product name/EC number frequency and the homologous gene organism taxonomy. However, to obtain a more reliable genome functional annotation, another strategy was employed. This strategy consists in annotating each candidate gene (gene that has at least one of the homologous genes associated to an EC number) automatically based on homologous hits revision status and source organism. In addition, a confidence label is assigned to each of these candidate genes.

The annotation was performed using *merlin*'s Automatic Workflow tool. This tool employs the previously stated strategy to annotate candidate genes, using a list of user-selected organisms, usually phylogenetically close organisms. The chosen organisms' order is crucial, since this tool initially searches, in the blast results, the first selected organism in the list. If a match is found, the gene is annotated based on the hit's information. Otherwise, the next organism is considered, creating in this way an iterative process.

During the annotation process, it was considered that any reviewed gene was more trustworthy than an unreviewed gene. Therefore, since *merlin* organizes BLAST results per database, the automatic workflow was executed twice, firstly for UniProt Swiss-Prot results and then for UniProt TrEMBL results.

In figure 5 is depicted the annotation workflow followed during this work. Candidate genes with hits against the UniProt Swiss-Prot database were annotated and assigned with a label based on a set of selected organisms. *X. fastidiosa* subsp. *pauca* De Donno, the strain used in this work, was selected as the most reliable source for annotations. However, this is not the species' most studied strain, therefore the genus *Xylella* was used to annotate candidate genes if no match was found previously. Additionally, *X. campestris* pv. campestris B100, the phylogenetically closest phytopathogenic organism to *X. fastidiosa*, and *E. coli* K12, the model organism for gram-negative bacteria, were considered in this annotation process. If none of these organisms was found in the hits list, the candidate gene was automatically annotated based on the information of the best scored hit.

Subsequently, the remaining unannotated candidate genes with matches against UniProt TrEMBL were annotated. However, since genes from this database are unreviewed, it was assumed that hits from well-studied organisms were more suitable to annotate candidate genes. Thus, the selected organism order was the opposite of the prior automatic workflow.

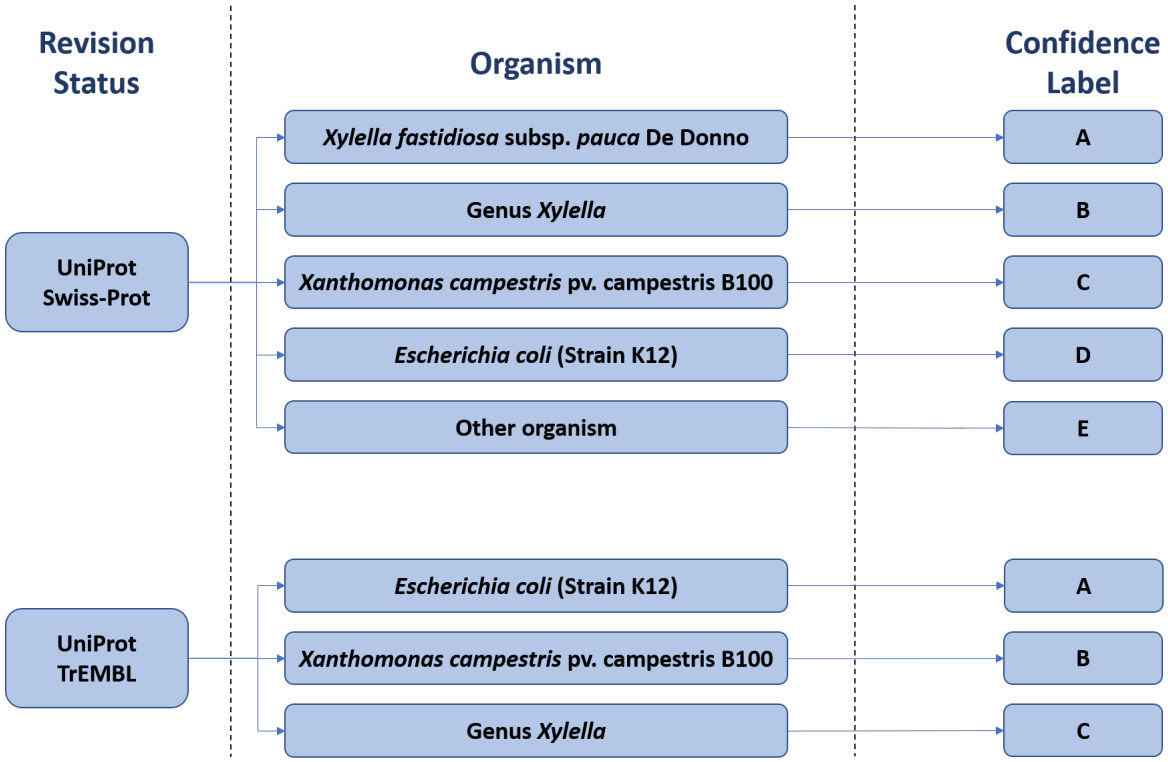


Figure 5.: Scheme of the annotation workflow followed in this work.

3.4 METABOLIC NETWORK RECONSTRUCTION

Methods used to assemble the metabolic network of the GSM model are presented in the following sections.

3.4.1 *Metabolic Data*

In order to assemble the metabolic network, it was necessary to collect metabolic data, which includes metabolites, enzymes and reactions. This data was retrieved automatically from KEGG [25] by *merlin*. As the metabolic network only contains a subset of the collected reactions which comprises spontaneous and enzyme catalyzed reactions, being the latter encoded in the organism's genome, by integrating the enzyme annotation results to the model, a draft metabolic network was obtained. However, during the integration step, reactions and genes related to DNA methylation, rRNA modification and signal transfer were ignored and not included in the model. Additionally, reactions that contained generic compounds as reactants were removed automatically from the model. The automatically removed metabolites can be seen in the supplementary file A.1.

3.4.2 *Transport Reactions, Compartments and Exchange Reactions*

A new tool that is being implemented in *merlin*, the *Transport Systems Tracker (TranSyT)*, was used to generate transport reactions. TranSyT is able to identify genes that encode carriers and the corresponding transported metabolites. Then, transport reactions are automatically created and associated to the encoding gene.

The reactions created by TranSyT were added to the model, which lead to the creation of two compartments within *merlin*. These compartments represent the extra and intracellular spaces and were named as outside and inside, respectively. For each of the existent metabolites in the outside compartment, an exchange reaction was automatically created. These exchange reactions represent the cell growth medium and their bounds can be edited to replicate chemically defined media.

3.4.3 *GPR Associations*

GPR associations are extremely important for an high-quality GSM model, since mutant phenotype predictions are severely affected by it. Usually, these associations are made manually, by searching information on literature and available databases. However, as in previous cases, an automatic tool implemented in *merlin* was used to generate GPR rules. This tool retrieves information from KEGG BRITE [127] regarding the protein structure and searches for the necessary subunits in the organisms genome. With this information, a GPR rule is implemented.

3.5 BIOMASS AND ENERGY REQUIREMENTS

The biomass reaction represents the cell composition of an organism and its high-quality formulation is required to avoid futures problems in the model validation step. Therefore, it is necessary to use organism-specific experimental data from literature, or, in its absence, information from the organism's genome or data from closely related organisms.

In this work, the reactants of the biomass equation are complex macromolecules, and their respective stoichiometric coefficients are represented in grams of macromolecule per gram of biomass. In other words, this can be considered as the relative abundance of each macromolecule in a cell. Since no studies that report the macromolecular composition of *X. fastidosa* were found, the models of *Escherichia coli* (iJO1366) [128] and *Xanthomonas campestris* [129] were used as templates.

A total of seven macromolecular entities was created to formulate the biomass reaction. By using *merlin*'s e-biomass tool [130], a draft biomass equation was created along with four other reactions, each one representing the synthesis of one of the following macromolecules:

DNA, *Ribonucleic Acid (RNA)*, protein and cofactor. Additionally, this tool determines the deoxynucleotide, nucleotide and amino acid contents for the DNA, RNA and protein synthesis reactions, respectively, using the previously downloaded genome files. The automatically generated cofactor reaction has a set of predefined precursors. However, according to literature, the template models and the metabolic capabilities of the assembled network, the necessary precursors to synthesize a cofactor macromolecule were changed.

The remaining macromolecules, namely peptidoglycan, lipopolysaccharide and lipid were inferred from the template models [128, 129]. In order to obtain the best possible results, the molecular composition of each one of these entities was retrieved from literature [4, 131, 132, 133, 134, 135].

The created lipid macromolecule accounts for different phospholipid molecules, which are all dependent on the synthesis of fatty acids of variable size. Hence, a metabolite that represents the average fatty acid and its corresponding synthesis reaction were generated based on data from multiple strains of *X. fastidiosa* reported by Wells *et al* [4].

Since the final biomass reaction is represented in grams of macromolecule per gram of cell dry weight, macromolecular entities' synthesis reactions should be adjusted accordingly. Therefore, in these reactions, molecular precursor stoichiometry must be represented in millimolar of precursor per gram of macromolecule. A schematic view on how to achieve this conversion is represented in figure 6.

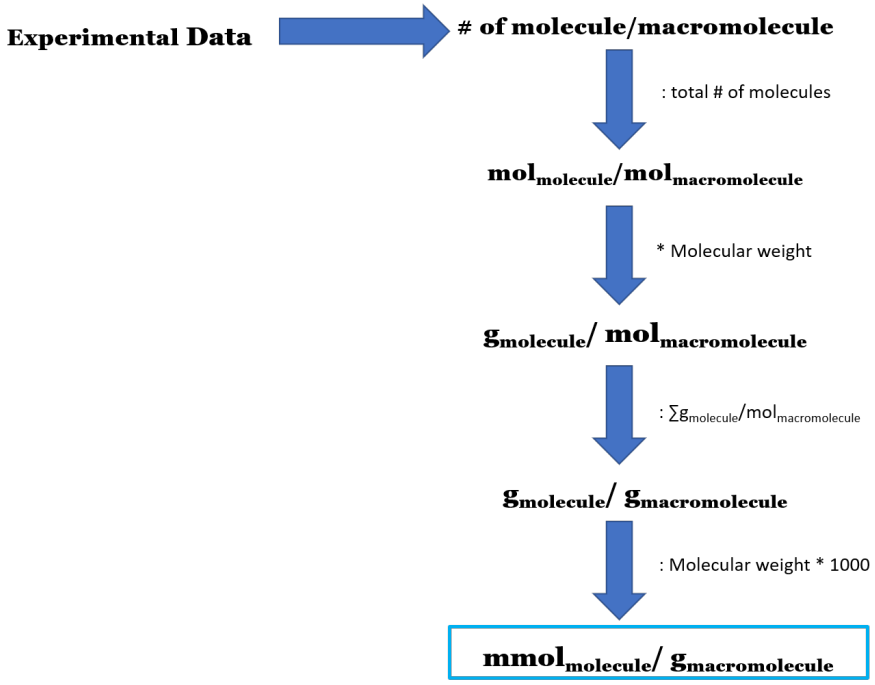


Figure 6.: Workflow to calculate precursors' stoichiometric coefficients in macromolecular synthesis reactions.

Due to the lack of information on energy requirements for *X. fastidiosa*, both growth associated and non-growth associated energy requirements were retrieved from the *E. coli* model [128]. Growth associated energy necessities were included in the biomass reaction with a value of 53.5, represented in ATP moles per gram of biomass. On the other hand, the maintenance associated energy was represented in a created ATP hydrolysis reaction, where both its bounds were restrained to a value of 3, in order to force flux through the reaction.

3.6 MANUAL CURATION

The following sections show strategies employed during the manual curation phase.

3.6.1 *Reversibility and Directionality*

All the reactions loaded from KEGG to assemble the draft network display a reversible status (default). Therefore, *merlin*'s 'Correct Reversibility' tool was used to automatically correct the reactions reversibility and directionality, by selecting Zeng database [136] as a reversibility source and the 'GramNegative' organism template in the tool's parameters.

However, in order to obtain a feasible model, multiple reactions were manually curated. A few databases such as BiGG, MetaCyc, BRENDA and ModelSEED [31], the template models of *E. coli* and *X. campestris* and the eQuilibrator [50] tool were used to infer the correct direction and reversibility of several reactions.

3.6.2 *Gap-Filling*

Since the synthesis of multiple biomass precursors was compromised, and therefore the model feasibility, a gap-filling analysis was performed. This process was supported by the *merlin*'s 'Unconnected Reactions' tool, which evaluates the metabolites connectivity within the network and identifies and highlights reactions where dead-end metabolites (metabolites that are only reactants or products) intervene. *merlin* also possesses a 'Draw in Browser' tool, which opens a selected KEGG pathway map on the web browser and colors all enzymes and reactions. Therefore, by using both tools, gaps are easily detected.

Figure 7 shows an example of a colored KEGG metabolic pathway. As depicted in the figure, EC numbers can be colored in different colors, each indicating a particular situation. The following list shows the meaning of each color:

- Green - Reaction is included in the metabolic network and its promoting enzyme belongs to the shown pathway;

- Dark blue (not included in figure 7) - Reaction is included in the metabolic network but the promoting enzyme is missing or included in another pathway;
- Cyan blue - Reaction is included in the metabolic network but is connected to a dead-end;
- Red - Reaction is included in the metabolic network, however it is a dead-end;
- No color - Reaction is not included in the metabolic network.

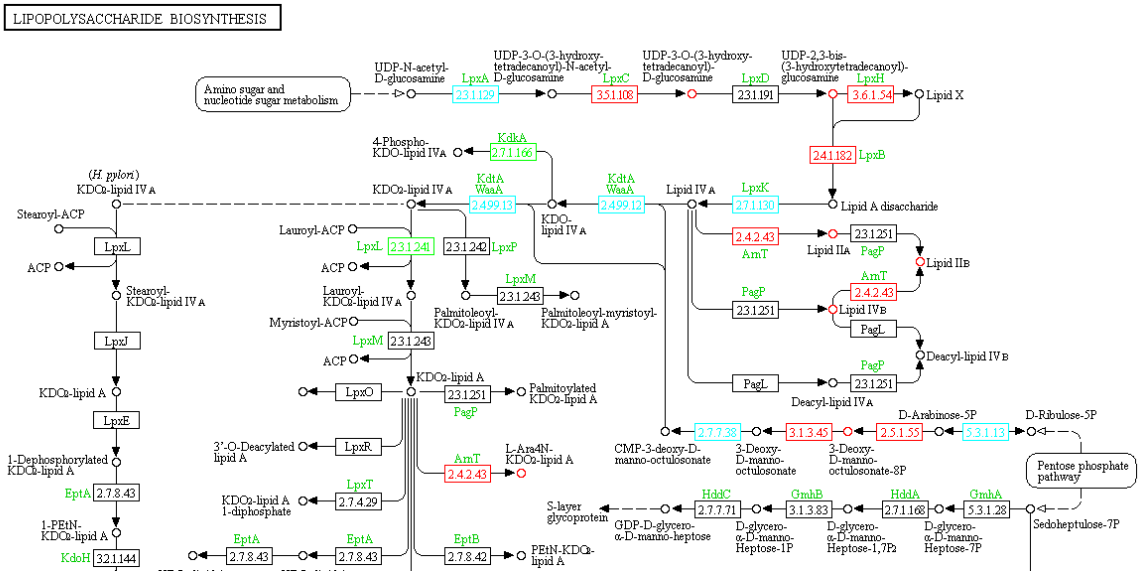


Figure 7.: Example of a KEGG pathway map section 'painted' by *merlin*'s 'Draw in browser' tool.

A collection of literature on *X. fastidiosa* metabolic potential, especially for pathways with blocked reactions, was analyzed beforehand. Additionally, KEGG's *X. fastidiosa* reference pathway was examined to find potential errors in a given pathway.

During the gap-filling process, multiple annotation issues were corrected, because whenever a gap was found, supporting evidence at the genome level was searched in *merlin*'s annotation table. If the gap reaction promoting protein was encoded in the genome, the reaction and its corresponding GPR association were added to the model. However, occasionally, reactions only supported by literature were added to the model, despite the lack of evidence in the genome annotation. Therefore, these reactions were not associated to an enzyme and a gene.

3.6.3 *Reaction's Balance*

Unbalanced reactions in a GSM model can largely jeopardize its predictions. Hence, stoichiometrically compromised reactions were rapidly corrected. Some of the errors found in unbalanced reactions were:

- Lack or excess of protons;
- Missing formulae on metabolites;
- Synthesis or breakdown of macromolecular metabolites (e.g. Cellulose).

As in previous steps of the manual curation stage, a *merlin* tool was used to aid in the reaction balance process. This tool highlights reactions where the sum of its elements differ between the reactants and products. Additionally, these reactions were exported to an excel file. Each one of the stoichiometric unbalanced reactions were sought in biological databases such as BRENDA and MetaCyc and updated accordingly in *merlin*.

3.7 MODEL VALIDATION

The model validation step started right after obtaining a feasible model. In this stage, simulations were performed using FBA and model predictions were compared to available literature, qualitatively and if possible, quantitatively. Whenever the results were not in accordance with the literature, further curation of the model was performed. The following aspects were evaluated in this stage:

- Growth under aerobic conditions as *X. fastidiosa* is an obligate aerobe;
- Metabolic flux pattern of glucose;
- Specific growth and exopolysaccharide production rates;
- Metabolic capabilities to grow under several carbon sources;
- Amino acid auxothrophies;
- Growth in several chemically defined media, created for *X. fastidiosa*.

In order to replicate these situations, several environmental conditions were created by changing exchange reactions bounds in the *Optflux* framework.

3.7.1 *Aerobic Metabolism*

Aeration conditions were defined in the simulations through means of an O_2 uptake reaction. Two different conditions were created for the simulations: anaerobic, restricting the lower bound to 0, and aerobic, leaving the reaction unbounded. All simulations were performed with glucose as a carbon source and all amino acids. These environmental conditions are detailed in Table 2.

Table 2.: Environmental conditions used to simulated aerated and anaerobic environments.

	Anaerobic			**Aerobic**	
Reaction Name	**Metabolite**	**Lower Bound**	**Reaction Name**	**Metabolite**	**Lower Bound**
R_EX_C00007_out	Oxygen	0	R_EX_C00007_out	Oxygen	-1000
R_EX_C00009_out	Orthophosphate	-1000	R_EX_C00009_out	Orthophosphate	-1000
R_EX_C14818_out	Fe^{2+}	-1000	R_EX_C14818_out	Fe^{2+}	-1000
R_EX_C00031_out	D-Glucose	-2.0	R_EX_C00031_out	D-Glucose	-2.0
R_EX_C00152_out	L-Asparagine	-0.1	R_EX_C00152_out	L-Asparagine	-0.1
R_EX_C00135_out	L-Histidine	-0.1	R_EX_C00135_out	L-Histidine	-0.1
R_EX_C00148_out	L-Proline	-0.1	R_EX_C00148_out	L-Proline	-0.1
R_EX_C00183_out	L-Valine	-0.1	R_EX_C00183_out	L-Valine	-0.1
R_EX_C00188_out	L-Threonine	-0.1	R_EX_C00188_out	L-Threonine	-0.1
R_EX_C00065_out	L-Serine	-0.1	R_EX_C00065_out	L-Serine	-0.1
R_EX_C00073_out	L-Methionine	-0.1	R_EX_C00073_out	L-Methionine	-0.1
R_EX_C00079_out	L-Phenylalanine	-0.1	R_EX_C00079_out	L-Phenylalanine	-0.1
R_EX_C00078_out	L-Tryptophan	-0.1	R_EX_C00078_out	L-Tryptophan	-0.1
R_EX_C00062_out	L-Arginine	-0.1	R_EX_C00062_out	L-Arginine	-0.1
R_EX_C00123_out	L-Leucine	-0.1	R_EX_C00123_out	L-Leucine	-0.1
R_EX_C00082_out	L-Tyrosine	-0.1	R_EX_C00082_out	L-Tyrosine	-0.1
R_EX_C00097_out	L-Cysteine	-0.1	R_EX_C00097_out	L-Cysteine	-0.1
R_EX_C00041_out	L-Alanine	-0.1	R_EX_C00041_out	L-Alanine	-0.1
R_EX_C00037_out	Glycine	-0.1	R_EX_C00037_out	Glycine	-0.1
R_EX_C00047_out	L-Lysine	-0.1	R_EX_C00047_out	L-Lysine	-0.1
R_EX_C00049_out	L-Aspartate	-0.1	R_EX_C00049_out	L-Aspartate	-0.1
R_EX_C00025_out	L-Glutamate	-0.1	R_EX_C00025_out	L-Glutamate	-0.1
R_EX_C00407_out	L-Isoleucine	-0.1	R_EX_C00407_out	L-Isoleucine	-0.1
R_EX_C00064_out	L-Glutamine	-0.1	R_EX_C00064_out	L-Glutamine	-0.1

3.7.2 *Metabolic Flux Pattern*

During this validation step, the carbohydrate metabolism was scrutinized. Firstly, several reactions associated with different conformations of glucose (e.g. alpha-Glucose, beta-

Glucose) and fructose (including the metabolites' derivatives) were promptly removed from the model. Instead of directly using these metabolites, the conformationless compounds (e.g. D-glucose, D-Fructose) were adopted for the network. However, in some occasions, reactions had to be edited in order to maintain a connected network, by replacing the metabolite with conformation for the generic compound.

It has been proposed that *X. fastidiosa* uses the Entner-Doudoroff pathway to convert glucose into pyruvate, instead of the typical glycolytic pathway [137]. While no studies directly prove this behavior in *X. fastidiosa*, the Entner-Doudoroff pathway has been demonstrated to be the main catabolic route of glucose in *X. campestris* [138].

Due to the metabolic similarities between these organisms, a metabolic flux through the Entner-Doudoroff pathway was considered for *X. fastidiosa*.

The flux of glucose through the network was evaluated using a medium that would allow a direct comparison to *X. campestris* results. Therefore, glucose and ammonia were used as carbon and nitrogen sources, respectively, as shown in table 3. In addition, a flux restriction through the phosphofructokinase reaction (conversion of fructose 6-phosphate into fructose 1,6-biphosphate) was imposed. Similarly to other xanthomonads, a pyrophosphate-dependent phosphofructokinase was identified in *X. fastidiosa*, which displays a low processivity in comparison to ATP-dependent phosphofructokinase [138, 139]. Consequently, flux through this reaction was limited to a minimal value, that is, only 1% of the glucose uptake rate.

Table 3.: Environmental conditions applied to *X. fastidiosa* model during validation of the glucose metabolic flux pattern.

Exchange Reaction Name	**Metabolite**	**Lower Bound**
R_EX_C00007_out	Oxygen	-10000
R_EX_C00031_out	D-Glucose	-2
R_EX_C00014_out	Ammonia	-10000
R_EX_C00009_out	Orthophosphate	-10000
R_EX_C00283_out	Hydrogen Sulfide	-10000
R_EX_C14818_out	Fe^{2+}	-10000

3.7.3 *Specific Growth and Exopolysaccharide production rates*

The reconstructed model was quantitatively evaluated, in terms of specific growth and exopolysaccharide production rates. Due to the lack of information for these cases in *X. fastidiosa*, the model was properly restrained in order to be directly comparable to data from *X. campestris*, reported by *Letisse et al.* [140]

While the *X. campestris* model displays an uptake rate of sucrose of 1.8 mmol/gDW/h, the model developed in this work does not contain sucrose in its network (see section 4.5.4). Hence, in order to maintain the same uptake of carbon, glucose flux was constrained to a lower bound of 3.6 mmol/gDW/h. The remaining constraints are displayed in table 4.

Table 4.: Environmental conditions applied to *X. fastidiosa* model during validation of specific growth and exopolysaccharide production rates.

Exchange Reaction Name	**Metabolite**	**Lower Bound**
R_EX_C00007_out	Oxygen	-10000
R_EX_C00031_out	D-Glucose	-3.6
R_EX_C00014_out	Ammonia	-10000
R_EX_C00009_out	Orthophosphate	-10000
R_EX_C00283_out	Hydrogen Sulfide	-10000
R_EX_C14818_out	Fe^{2+}	-10000

3.7.4 *Metabolism under different carbon sources*

The model's predictive capabilities were tested using different carbon sources. Thus, environmental conditions that would simulate a minimal medium with distinct carbohydrates were created. These minimal conditions are all based on the environmental conditions presented in section 3.7.2. In order to maintain the same number of carbon atoms within the network, the influx of carbon was properly constrained (e.g. while glucose was constrained to a value of 2mmol/gDW/h, malate had an uptake rate of 3mmol/gDW/h). However, as no information on growth rates under different carbon sources uptake was found for *X. fastidiosa*, data from *X. campestris* [129], once again, was used to validate the model.

3.7.5 *Amino acid auxotrophies*

No amino acid auxotrophies have been reported for *X. fastidiosa*. Additionally, the organism is able to grow in media with glutamine as the sole nitrogen source [141]. Validation procedures for this situation started with the formulation of an environmental condition with availability for all amino acids, as presented in table 5. From there on, single omissions were performed to test if auxotrophies were detected within the network.

Table 5.: Environmental conditions applied during the amino acid auxotrophies validation phase.

Reaction Name	Metabolite	Lower Bound
R_EX_C00007_out	Oxygen	-1000
R_EX_C00009_out	Orthophosphate	-1000
R_EX_C14818_out	Fe^{2+}	-1000
R_EX_C00283_out	Hydrogen Sulfide	-1000
R_EX_C00031_out	D-Glucose	-2.0
R_EX_C00152_out	L-Asparagine	-0.1
R_EX_C00135_out	L-Histidine	-0.1
R_EX_C00148_out	L-Proline	-0.1
R_EX_C00183_out	L-Valine	-0.1
R_EX_C00188_out	L-Threonine	-0.1
R_EX_C00065_out	L-Serine	-0.1
R_EX_C00073_out	L-Methionine	-0.1
R_EX_C00079_out	L-Phenylalanine	-0.1
R_EX_C00078_out	L-Tryptophan	-0.1
R_EX_C00062_out	L-Arginine	-0.1
R_EX_C00123_out	L-Leucine	-0.1
R_EX_C00082_out	L-Tyrosine	-0.1
R_EX_C00097_out	L-Cysteine	-0.1
R_EX_C00041_out	L-Alanine	-0.1
R_EX_C00037_out	Glycine	-0.1
R_EX_C00047_out	L-Lysine	-0.1
R_EX_C00049_out	L-Aspartate	-0.1
R_EX_C00025_out	L-Glutamate	-0.1
R_EX_C00407_out	L-Isoleucine	-0.1
R_EX_C00064_out	L-Glutamine	-0.1

3.7.6 *Growth in different chemically defined media*

Several chemically defined media have been created throughout the years to support *X. fastidiosa* growth. CHARD2, 3G10-R [115], XDM1-XDM5 [141], XfD1-XfD9 [142] and XF-26 [143] are examples of chemically defined media based on xylem fluid chemistry and genomic information developed for the phytopathogenic bacterium. In order to qualitatively validate the model, each of these media were simulated by restraining model's exchange reactions properly. The metabolites used to simulate these media are shown in table 6. *In silico* predictions were subsequently compared to the reported literature data.

Table 6.: Metabolites used to simulate different chemically defined media. These metabolites are divided in three columns: Carbon Source, Amino Acids and Others.

Medium	Carbon Source	Amino Acids	Others	Reference
CHARD2	-	Ala, Arg, Asp, Cys, Gln	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[115]
3G10-R	Glucose	Gln	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[115]
XDM1	Glucose	Gln, Asp, Ala, Gly, Ser Thr	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[141]
XDM2	Glucose	Gln, Met, Asn, Ser	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[141]
XDM3	Glucose	Gln, Met, Asp, Ser	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[141]
XDM4	Glucose	Gln	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[141]
XDM5	Glucose	Gln, Met	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[141]
XF-26	Citrate, Succinate	Ala, Asn, Asp, Cys, Gly, Gln, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Val	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[143]
XfD1	Citrate, Succinate	Gln, Asp, Cys	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD2	Citrate, Succinate	Gln, Asp, Cys	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD3	Citrate	Gln, Asp, Cys	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD4	Citrate, Succinate	Gln	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD5	Citrate	Gln	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD6	Citrate	Gln	Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD7	Citrate	Gln	Heme, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD8	Citrate	Gln	Orthophosphate, Hydrogen Sulfide, Oxygen	[142]
XfD9	Citrate	Gln	Fe^{2+}, Orthophosphate, Hydrogen Sulfide, Oxygen	[142]

3.8 POTENTIAL DRUG TARGETS

One of the major purposes of the development of a *X. fastidiosa* GSM model is the identification of potential drug targets. In order to pinpoint these targets, essential genes determined by the 'Critical genes' tool of *Optflux* [56] were filtered using a Python script. For each of these genes, an homology search against NCBI available *Olea europaea* genomes was performed. Genes with no homologies or having all hits with e-values above 0.001 (arbitrary selected value), were considered to be potential drug targets for the bacterium. However, the initial set of genes determined by *Optflux* tool is highly dependent on the given environmental conditions.

The medium based on xylem fluid chemistry CHARD2 [115] was used to simulate environmental conditions since it would replicate the natural environment of the phytopathogen. Additionally, the 'Critical genes' tool was run for both planktonic and biofilm growth. Finally, the filtered genes were reviewed manually, evaluated on their essentiality and drugs known to inhibit their activity were sought on DrugBank [144].

4

RESULTS AND DISCUSSION

4.1 GENOME ANNOTATION

The genome annotation performed in this work identified out of the 2066 genes encoded in *X. fastidiosa* subsp. *pauca* De Donno genome, a total of 825 candidate genes. From these, 724 were annotated based on reviewed genes, whereas the remaining 101 were annotated with unreviewed information. Results from the automatic workflow are displayed in table 7, where candidate genes were accounted for, based on their assigned label. *X. fastidiosa* subsp. *pauca* De Donno's genome has not been reviewed yet, thus all genes were compared with reviewed genes from *Xylella* spp. and assigned with label B. As shown in table 7, this is the most represented label, which reflects an overall high-quality genome annotation.

Table 7.: Genome annotation automatic workflow results. Gene count is presented for each label for both UniProt Swiss-Prot and UniProt TrEMBL databases.

Assigned Label	Gene Count	
UniProt Swiss-Prot		
A	0	
B	334	
C	7	
D	279	
E	104	
		724
UniProt TrEMBL		
A	0	
B	56	
C	45	
		101

A total of 735 metabolic genes, genes associated to EC numbers, were annotated from the 825 candidate genes. Thus, from the gathered results, we assume that about 35,6% of the organism's genome encode genes associated with metabolic functions. As shown in table 8, Transferases and hydrolases are the most represented classes from the annotated genes, while the recently added translocases class displays the lowest percentage among the possible enzyme classes. Genes associated with multiple functions, from different classes, were included in the class of the first EC number annotated, as this is usually the main function of the enzyme.

Table 8.: Distribution of metabolic genes among EC classes and their respective percentage.

EC Class	**Gene Count**	**Percentage of annotated genes**
Oxidoreductases	107	15%
Transferases	240	33%
Hydrolases	190	26%
Lyases	61	8%
Isomerases	46	6%
Ligases	62	8%
Translocases	29	4%

Although about 16% of the annotated genes (117) were assigned with incomplete EC numbers, most of these genes were not considered to the network reconstruction. However, when relevant KEGG reactions were associated to incomplete EC numbers, the respective gene was included in the model. Gene B9J09_RS08675 is an example of this situation, as it encodes the enzyme associated to the EC number 2.7.8.-. This enzyme catalyzes reaction R07390, required for the synthesis of cardiolipin. The final model had a considerably lower number of integrated genes when compared to 735 identified metabolic genes, as genes with undefined function (incomplete EC number) or associated with irrelevant metabolic processes, modelling wise, were not included.

4.2 TRANSPORT REACTIONS

TranSyT generated 306 transport reactions, which are divided among different classes, such as antiport, symport, uniport and *ATP-Binding Cassette (ABC)* transporters. Table 9 shows the distribution of the transport reactions among the different classes as well as the number of included reactions in the model.

Symport reactions for all amino acids were predicted by TranSyT. In addition, several transport reactions for carbon sources, such as glucose, galactose, fructose, mannose, lactose

and glycerol were included in the model. Organic acids, compounds commonly found in xylem sap [84], also contain carriers within the model.

Only a few manually created carriers had to be included in order to obtain a feasible model, which reflects the outstanding performance of the TranSyT tool. Diffusion of water, carbon dioxide, ammonia and oxygen were some of the created transport reactions. Additionally, as reported in literature, *X. fastidiosa* contains a citrate symporter [145]. Since no reaction reflecting this transport was automatically generated by TranSyT, a manually created citrate symporter was included in the model.

The transport reactions for formate and xanthomonadin, infered by-products of *X. fastidiosa* metabolism, were also included in the model. The reasoning behind these additions will be discussed in the model validation section.

Table 9.: Number of generated and included in the model transport reactions organized by class.

Transporter Type	Number of reactions	
	Generated	In model
Antiport	36	23
Symport	78	78
Uniport	92	76
ABC	96	72
Other	4	0
Manual	7	7
	313	256

4.3 BIOMASS AND ENERGY REQUIREMENTS

As described before, the biomass reaction of this model required seven macromolecular entities. All of these entities, and their respective fraction (represented in grams of macromolecule per gram of cell dry weight), were inferred from *E. coli* [128] and *X. campestris* [129] models and are compiled in table 10.

Both template models account for glycogen in the formulation of the biomass, although it is not directly included in the biomass reaction of *X. campestris*. However, this is a distinct case for *X. fastidiosa*. As reported by Henrissat *et al.* [146], the glycogen metabolism seems to be a lost trait among several symbiotic and parasitic bacteria. No glycogen biosynthetic and degrative enzymes were found in *X. fastidiosa* genome, thus this multibranched polysaccharide was not considered while formulating the biomass reaction for the model. Nonetheless, the glycogen fraction was unevenly distributed between two macromolecules

Table 10.: Biomass composition of *E. coli* [128], *X. campestris* [129] and *X. fastidiosa* models.

Macromolecule	***E. coli***	***X. campestris***	***X. fastidiosa***
		g/gDW	
DNA	0.031	0.04	0.04
RNA	0.205	0.16	0.16
Protein	0.55	0.49	0.49
Lipid	0.091	0.13	0.13
Glycogen	0.025	-	0
Peptidoglycan	0.025	-	0.0588
Cofactor + Polyamine	0.039	-	0.0706
Lipopolysaccharide	0.034	0.033	0.0505

that contain carbohydrates in their constitution. Petidoglycan was assigned with 30% of the glycogen fraction and lipopolysaccharides with 70%.

The *X. campestris* model was used to retrieve both DNA and RNA macromolecular content, namely the nucleotides mass per gram of biomass, 0.04 and 0.16, respectively. These fractions represent the weight (in grams) of DNA and RNA per 1 gram of cell biomass. By using *merlin*'s e-Biomass tool [130], two metabolites, e-DNA and e-RNA, and their respective synthesis reactions, which are summarized in tables 11 and 12, were created. The stoichiometric coefficients used in these reactions are represented in millimoles of molecule necessary to produce 1 gram of macromolecule, which were determined by using *X. fastidiosa*'s genomic information. Additionally, the precursor metabolites are in their triphosphate forms. However, it is known that during DNA and RNA polymerization reactions diphosphate is released. Therefore, this was taken to account when assembling the reactions.

As in the aforementioned macromolecules, the average protein content in the overall biomass was infered from *X. campestris*. The protein synthesis reaction, presented in table 13, accounts for all amino acids. By using proteomic information and the e-Biomass tool [130], the stoichiometric coefficient of each amino acid was calculated and incorporated in the reaction as e-Protein precursors. Along with this manually created metabolite, water was included as a product of the reaction, since during the formation of peptide bounds water is produced.

The e-Lipid metabolite encompasses a set of very important molecules in the cell membrane, phospholipids. However, no information on the lipid composition of *X. fastidiosa* was found. Therefore, data reported by Dianese and Schaad for the *X. campestris* B-24 strain [131] was adapted. The precursors required to form the lipid macromolecule, which

Table 11.: DNA synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric Coefficient (mmol/gDNA)
Reactants		
dATP	C00131	0.849
dCTP	C00458	0.781
dGTP	C00286	0.822
dTTP	C00459	0.781
Products		
e-DNA	-	1
Diphosphate	C00013	3.233

Table 12.: RNA synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric Coefficient (mmol/gRNA)
Reactants		
ATP	C00002	0.623
CTP	C00063	0.850
GTP	C00044	0.994
UTP	C00075	0.628
Products		
e-RNA	-	1
Diphosphate	C00013	3.095

represents a 0.13 portion of the cell dry biomass, as in the *X. campestris model*, are shown in table 14.

Lipid molecules contain a variable number of fatty acid chains, making them dependent on the synthesis of fatty acids. In addition, fatty acid chains are highly variable in size. Therefore, for modelling purposes, the average fatty acid composition of *X. fastidiosa*, presented in table 15, was calculated using data of multiple strains [4]. Due to the lack of information regarding odd number fatty acids in KEGG reactions, only fatty acids with even number of carbon atoms were considered to formulate the average fatty acid synthesis reaction. It is also noteworthy to mention that esters associated to an acyl carrier protein (e.g. Octanoyl-[acp]) were used in the synthesis reaction instead of the typical fatty acid molecules (e.g. Octanoic acid), as no thioester hydrolases were found encoded in the

Table 13.: Protein synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric Coefficient (mmol/gProtein)
Reactants		
Glutamyl-tRNA	C02987	0.381
Glycyl-tRNA(Gly)	C02412	0.690
L-Alanyl-tRNA(Ala)	C00886	0.933
L-Arginyl-tRNA(Arg)	C02163	0.611
L-Asparaginyl-tRNA(Asn)	C03402	0.298
L-Aspartyl-tRNA(Asp)	C02984	0.483
L-Cysteinyl-tRNA(Cys)	C03125	0.100
L-Glutaminyl-tRNA(Gln)	C02282	0.475
L-Histidyl-tRNA(His)	C02988	0.242
L-Isoleucyl-tRNA(Ile)	C03127	0.489
L-Leucyl-tRNA(Leu)	C02047	0.991
L-Lysyl-tRNA(Lys)	C01931	0.323
L-Methionyl-tRNA(Met)	C02430	0.213
L-Phenylalanyl-tRNA(Phe)	C03511	0.318
L-Prolyl-tRNA(Pro)	C02702	0.445
L-Seryl-tRNA(Ser)	C02553	0.532
L-Threonyl-tRNA(Thr)	C02992	0.507
L-Tryptophanyl-tRNA(Trp)	C03512	0.132
L-Tyrosyl-tRNA(Tyr)	C02839	0.239
L-Valyl-tRNA(Val)	C02554	0.682
Products		
e-Protein	-	1
H_2O	C00001	9.084

genome. Therefore, to maintain a steady-state, *Acyl-Carrier Protein (ACP)* had to be included as a product in the average fatty acid synthesis reaction (R_Fatty_Acid_Entity).

Lipopolysaccharides (LPS) are constituents of the outer membranes, specifically the outer leaflet, in a majority of Gram-negative bacteria. A LPS molecule is composed of three parts: lipid A, a core oligosaccharide and an O-antigen [132]. Usually, both lipid A and core oligosaccharide units are conserved among different organism, whereas the O-antigen moiety, an important virulence determinant, is highly variable [133].

Table 14.: Lipid synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric coefficient (mmol/gLipid)
Reactants		
Cardiolipin	C05980	0.126
Phosphatidylethanolamine	C00350	0.126
Phosphatidylglycerol	C00344	0.318
Phosphatidylserine	C02737	0.507
Products		
e-Lipid	-	1

Table 15.: Fatty acid synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric coefficient
Reactants		
Octanoyl-[acp]	C05752	0.012
Decanoyl-[acp]	C05755	0.038
Dodecanoyl-[acp]	C05223	0.010
Hexadecenoyl-[acp]	C05763	0.404
Hexadecanoyl-[acp]	C05764	0.454
Linoleoyl-[acp]	-	0.056
Oleoyl-[acp]	C01203	0.026
Products		
Fatty Acid	-	1
ACP	C00229	1

The LPS fraction in the overall biomass was retrieved from the *X. campestris* model but was incremented afterwards with a percentage of the glycogen portion, as stated before. For the LPS synthesis reaction it was assumed that the macromolecule was composed of two major parts, a KDO2-Lipid A and glycosyl residues, where the former represents the lipid A and an inner part of the core oligosaccharide and the latter comprises the outer section of the core and the O-antigen. The distribution of glycosyl residues was inferred based on the data reported for *X. fastidiosa* by Rapicavoli *et al* [134].

A manually created reaction for the production of peptidoglycan was assembled, while the existing peptidoglycan synthesis KEGG pathway was ignored. This macromolecule accounts for 0.0588 of the overall cell biomass, which was obtained by adapting data from the *E. coli* model. As no description on *X. fastidiosa* peptidoglycan was found, information from

Table 16.: Lipopolysaccharide synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric coefficient (mmol/gLPS)
Reactants		
KDO2-Lipid A	C06026	0.419
dTDP-L-rhamnose	C03319	0.343
UDP-D-Xylose	C00190	0.0502
GDP-mannose	C00096	0.00836
UDP-glucose	C00029	0.00836
UDP-glucuronate	C00167	0.00836
UDP-N-acetyl-alpha-D-glucosamine	C00043	0.00125
Products		
e-LPS	-	1
UDP	C00015	0.06817
GDP	C00035	0.00836
dTDP	C00363	0.343

E. coli was used to formulate the reaction [135]. According to the gathered data, peptidoglycan is composed by two monosaccharide units and five amino acids, being one of them D-glutamate. However, no pathway for the synthesis of this metabolite was found within the network, thus it was replaced by the enantiomer, L-glutamate. Table 17 summarizes the formulated reaction for peptidoglycan synthesis.

The remaining fraction of the biomass was assigned to Cofactors, which represent a group of metabolites essential for enzyme functionality. By using the e-Biomass tool, a reaction with a predefined set of metabolites was created. However, while the majority of them was kept in the reaction, some had to be removed since no genomic or literature based evidence was found for their synthesis in *X. fastidiosa*. Thiamine, pyridoxal and biotin are examples of cofactors not included in the reaction. The stoichiometric coefficient of each cofactor, shown in table 18, was calculated based on the assumption that each one should have the same weight (in grams) in the overall biomass.

To finalize the biomass formulation, energy requirements were added to the model. Both growth and non-growth associated energy requirements, represented in the form of ATP moles per gram of biomass, were retrieved from the *E. coli* model [128]. A value of moles of 53.5 ATP per gram of biomass per hour was assigned to the biomass equation, which represents the growth associated energy, whilst a value of 3 moles of ATP per gram of biomass per hour was allocated for maintenance energy.

Table 17.: Peptidoglycan synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric Coefficient (mmol/gPeptidoglycan)
Reactants		
UDP-N-acetyl-alpha-D-glucosamine	C00043	1.054
UDP-N-acetyl-muramate	C01050	1.054
L-Lysine	C00047	1.054
L-Alanine	C00041	1.054
L-Glutamate	C00025	1.054
D-Alanine	C00133	2.108
Products		
e-Peptidoglycan	-	1
UDP	C00015	2.108
H_2O	C00001	5.270

Table 18.: Cofactor synthesis reaction for *X. fastidiosa* subsp. *pauca* De Donno model.

Metabolite name	KEGG ID	Stoichiometric coefficient (mmol/gCofactor)
Reactants		
CoA	C00010	0.109
Glutathione	C00051	0.271
Tetrahydrofolate	C00101	0.187
NAD^+	C00003	0.125
$NADP^+$	C00006	0.112
Riboflavin	C00255	0.221
FAD	C00016	0.106
Putrescine	C00134	0.945
Spermidine	C00315	0.574
S-adenosyl-L-methionine	C00019	0.209
Heme	C00032	0.135
FMN	C00061	0.183
Products		
e-Cofactor	-	1

The maximization of biomass production is the typical objective of simulations with GSM models. However, compounds that compete for the same carbon pool, known to be produced by the organism and not directly linked to the biomass reaction, have a zero production rate during biomass maximization simulations. Exopolysaccharides, a known product of *X. fastidiosa* and an important virulence factor during its life cycle [147], were not included in the biomass reaction of this model. Instead, a pathway for the synthesis of fastidian gum, the *X. fastidiosa* exopolysaccharide, was created based on information retrieved from available literature [148]. Without exopolysaccharides in the biomass reaction, maximization of the biomass reaction would replicate the planktonic phase of *X. fastidiosa* life cycle. On the other hand, to replicate the biofilm phase of the organism, a manually created reaction that ensures both the production of biomass and exopolysaccharides, entitled as 'R_Coupled_Biomass', was added to the model. By maximizing this reaction during simulations, synthesis of biomass and exopolysaccharides is forced at specific predefined output ratios. Therefore, the developed *X. fastidiosa* model is able to replicate both stages of the organisms' life cycle. This strategy was based on the *X. campestris* model [129] and the output ratios were adapted from published data [149]. Table 19 displays the established ratios in both models.

Table 19.: Output ratios of competing entities in both *X. campestris* and *X. fastidiosa* models. The sum of ratios are equal to 1.

Entity	***X. campestris***	***X. fastidiosa***
	Output Ratio	
Biomass	0.14	0.203
Glycogen	0.31	0
Exopolysaccharides	0.55	0.797

4.4 MODEL CURATION

The current section displays all the changes made to the model during the model curation phase.

4.4.1 *Reversibility and Directionality*

The assembled reactions in the network were firstly corrected, automatically, using *merlin*'s tool. However, since the production of several biomass precursors remained compromised and futile ATP generating cycles were found, manual curation efforts were made. A total of 39 reactions were manually curated in order to address this issues. Table 20 shows a

compiled list of reactions grouped by correction. Additionally, reactions added in following steps of the manual curation were reviewed separately.

Table 20.: Summary of manually curated reactions, in terms of reversibility and directionality.

Correction	Reaction List
Directionality change	R02777, R03443, R00200, R01073, R00226, R01072, R00430, R02017, R02019, R01138, R01858, R02024, R02018, R04292, R03254, R00177, R03191, R03192, R01213, R02016, R03458, R00937, R00940, R02241, R02864, R01132, R00572, R00659, R00724
To Reversible	R01512, R02291, R01652, R00936, R00939, R02300
To Irreversible	R10948, R00848, R01257, R07641, R00351, R02093, R02094

4.4.2 *Gap-Filling*

Several metabolic gaps were found within the network, especially in amino acids, fatty acid, phospholipid and cofactors pathways. To address this situation a total of 28 reactions, represented in table 21, were added. Reactions R08161 and R03370 from the fatty acid biosynthesis pathway were the only reactions included in the model without any evidence at the genome and literature level.

According to published literature, *X. fastidiosa* is capable of *de novo* folate synthesis [150]. Reaction R05553, involved in folate biosynthesis, is promoted by aminodeoxychorismate lyase (EC 4.1.3.38), which was not found to be encoded in the genome, since this enzyme is hard to distinguish from the large branched-chain amino acid aminotransferase family (EC 2.6.1.42) [150]. By analysing KEGG reference pathways of several *X. fastidiosa* strains, it was found that the enzyme 3.1.3.1 (catalyzes reaction R04620) is encoded in multiple strains of the organism. However, this enzyme is not encoded in the genome of *X. fastiosa* subsp. *pauca* De Donno. This suggests that this strain might be defective in folate synthesis or a different mechanism/enzyme is involved in this process. Despite the lack of genomic evidence, the organism has been shown to grow in media that lacks folate in its composition [115]. Hence, both reactions were included in the model to close the metabolic gaps of the folate pathway.

Regardless of the absence of genomic evidence, two additional reactions, R07280 and R02472, which are associated to riboflavin metabolism and pantothenate and *Coenzyme A (CoA)* biosynthesis pathways, respectively, were included in the model. However, *X. fastidiosa* expresses genes of these pathways, which suggests metabolic capabilities to synthesize these metabolites [118].

Table 21.: Reactions added to the model during gap-filling analysis.

Reaction ID	EC number	Encoded in Genome	Reference
R07390	2.7.8.-	x	-
R02029	3.1.3.27	x	-
R08161	1.14.19.2	-	-
R03370	1.14.19.2	-	-
R00351	2.3.3.16	x	-
R02164	1.3.5.1	x	-
R07764	4.2.1.59	x	-
R00259	2.3.1.1	x	-
R02291	1.2.1.11	x	-
R04475	2.6.1.17	x	-
R00891	4.2.1.22	x	-
R01290	4.2.1.22	x	-
R00582	3.1.3.3	x	-
R01776	2.3.1.31	x	-
R10597	4.1.3.45	x	[151]
R01993	3.5.2.3	x	-
R01868	1.3.5.2	x	-
R00485	3.5.1.1	x	-
R00355	2.6.1.1	x	-
R03314	Non-enzymatic	-	-
R04550	2.3.1.191	x	-
R05075	2.3.1.243	x	-
R05146	2.3.1.241	x	-
R04620	3.1.3.1	-	[150]
R05553	4.1.3.38	-	[150]
R02472	1.1.1.169	-	[118]
R07280	3.1.3.104	-	[118]
R03222	1.3.3.4	x	[152]

During the gap-filling process, misannotations were found and promptly corrected (table 22). Generally, after reviewing blast results for each of the misannotated genes, a solution was found. However, the metabolic function of two genes, B9J09_RS10630 and B9J09_RS00885, was inferred from literature [151, 152]. B9J09_RS10630 displays a high sim-

ilarity to the XanB2 gene of *X. campestris*, which is a bifunctional chorismatase related to both ubiquinone and xanthomonadin synthesis [151]. Therefore, this gene was assigned with EC numbers 4.1.3.45 and 4.1.3.40. Finally, gene B9J09_RS00885 was found to be linked to porphyrin metabolism. This gene is homologous to the COG1981 gene of *Acinobacter baylyi*, which encodes a novel protoporphyrinogen oxidase, an enzyme essential for the biosynthesis of heme [152].

Table 22.: Annotation corrections made during gap-filling analysis.

Locus Tag	Previous EC number	New EC number
B9J09_RS09595	2.7.2.8	2.7.2.8, 2.3.1.1
B9J09_RS07910	-	1.2.1.11
B9J09_RS07665	2.6.1.11	2.6.1.11, 2.6.1.17
B9J09_RS09295	1.4.3.16	1.3.5.1
B9J09_RS09305	-	1.3.5.1
B9J09_RS02385	4.5.1.5, 2.5.1.47	4.2.1.22
B9J09_RS00940	-	3.1.3.3
B9J09_RS02720	2.3.1.-	2.3.1.31
B9J09_RS10630	3.3.2.13	4.1.3.45, 4.1.3.40
B9J09_RS09655	3.5.2.5	3.5.2.3
B9J09_RS09205	-	3.5.1.1
B9J09_RS00150	2.6.1.107, 2.6.1.57	2.6.1.1
B9J09_RS09430	2.3.1.-	2.3.1.191
B9J09_RS00430	2.3.1.241	2.3.1.243
B9J09_RS08000	2.3.1.-	2.3.1.241
B9J09_RS00885	-	1.3.3.4
B9J09_RS10990	3.1.3.-	3.1.3.27

4.4.3 *Reaction's Balance*

Relying on *merlin*'s tool, 102 reactions were found to be unbalanced. These reactions were reviewed one by one and corrected, if needed. Table 23 summarizes the corrections made, by grouping the number of reactions by correction. Multiple reactions related to the biomass pathway or containing formula-less participating metabolites were incorrectly considered as unbalanced. As a result, the missing metabolite formulae was added, when possible, or the reaction was ignored. Additionally, spontaneous and unconnected unbal-

anced reactions were promptly removed from the model. The most common mistake out of the 51 corrected reactions lies in the incorrect number of protons.

Table 23.: Summary of corrections made during reaction balance manual curation.

Correction	Number of Reactions
Add proton (H^+) to products	8
Add proton (H^+) to reactants	17
Remove proton (H^+) from reactants	2
Metabolite breakdown	3
Add H_2O to products	1
Added Missing Formulae	20
Remove Reaction	21
None	30

A specific case of corrections is related to metabolite breakdown reactions. Usually, these reactions display the same metabolite as reactant and product, which leads to flux issues during *in silico* simulations. Reactions of this nature found within the network, displayed in table 24, were corrected.

Table 24.: Reactions associated to metabolite breakdown errors and their respective correction.

Reaction	Before Correction	After Correction
R01105	Galactan + H_2O =>D-Galactose + Galactan	Galactan + H_2O =>2 D-Galactose
R01982	Pectate + H_2O =>D-Galacturonate + Pectate	Pectate + H_2O =>2 D-Galacturonate
R11307	Cellulose + H_2O <=>Cellulose + Cellodextrin	Cellulose + H_2O =>2 Cellodextrin

4.5 MODEL VALIDATION

The current section shows the results obtained during the *X. fastidiosa* model validation phase.

4.5.1 *Aerobic Metabolism*

The aerobic metabolism, an essential part of *X. fastidiosa* survival, was the first aspect evaluated of the reconstructed model. However, before initiating this process, the oxidative

phosphorylation pathway was manually created and inserted in the model, since these reactions are not automatically integrated to the model. Enzymatic complexes of this pathway, such as NADH dehydrogenase, succinate dehydrogenase, ubiquinol oxidase and ATP synthase, were identified in the organism's genome. A manual created compartment, entitled as 'pmf', was created to simulate the periplasmatic space and the generated electrochemical proton gradient generated by most of these enzymatic complexes. In addition, since no information was found for *X. fastidiosa* P/O ratio, the included pathway was adapted from the oxidative phosphorylation of the *X. campestris* model, which displays a P/O ratio of 1.

An aerated environment resulted in the prediction of growth, while an anaerobic environment (oxygen uptake rate restrained to zero) led to no growth prediction. In fact, the model was infeasible in an oxygen restricted environment, meaning that not even the ATP maintenance requirements could be fulfilled. Therefore, a viable energy synthesis metabolism can only be obtained in an aerated environment, which reflects the *in vivo* situation of the pathogen.

4.5.2 *Metabolic Flux Pattern*

In a cellular growth only (without synthesis of exopolysaccharides) maximization condition, 90% and 8% of the carbohydrate flux is predicted to be directed through the Entner-Doudoroff and Pentose phosphate pathways, respectively. These predictions fall within range of experimentally determined carbon fluxes for several xanthomonads [153], including *X. campestris* [138], which range between 81-93% and 7-19% for the Entner-Doudoroff and pentose phosphate pathways. A visualization of the glucose degradation route is displayed in figure 8.

The availability and activity of three different glucose catabolism pathways is a highly unexpected feature. Considering that the glycolysis is the most energetically efficient pathway, the question on why the Entner-Doudoroff pathway is the prevalent catabolic route arises. As previously stated, *X. fastidiosa* contains a pyrophosphate-dependent phosphofructokinase that displays low processivity. This enzyme also represents a minimal fraction of the total proteome in related organisms [129]. Therefore, flux through glycolysis might be limited by this enzyme. Additionally, it is noteworthy to mention that the lack of activity of aldolase, the enzyme responsible for the reaction after phosphofructokinase, has been reported for *X. fastidiosa* [137]. In the absence of this enzyme's activity, glycolysis is impaired.

The Entner-Doudoroff pathway may offer an advantageous property for *X. fastidiosa*. This pathway generates NADPH, which is involved both in biosynthetic reactions and the detoxification of reactive oxygen species. These reactive species are often produced by plants upon detection of invasive pathogens [154] and can be used as a defense response mechanism

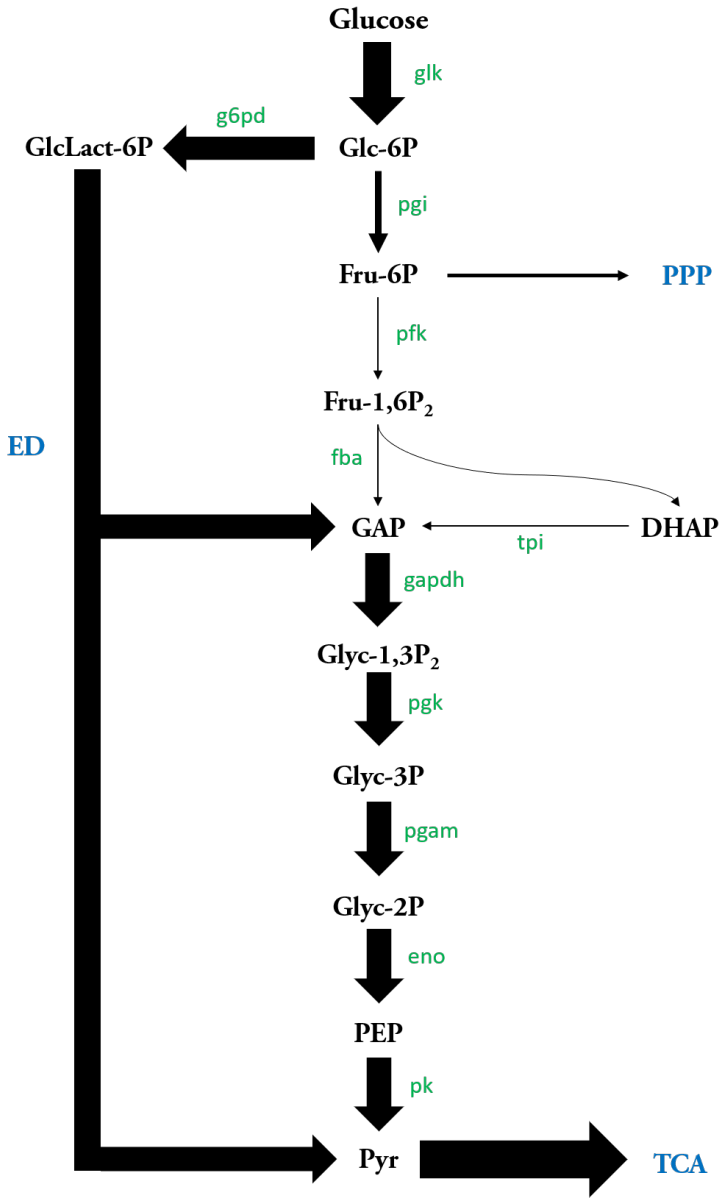


Figure 8.: Schematic view of the carbon flux distribution predicted for *X. fastidiosa*. Arrow width indicates the flux magnitude (not proportional to predicted fluxes) through reactions/-pathways. Terms in black, green and blue represent metabolites, enzymes and pathways, respectively. Glc-6P, glucose 6-phosphate; GlcLact-6P, D-glucono-1,5-lactone 6-phosphate; Fru-6P, fructose 6-phosphate; Fru-1,6P_2, fructose 1,6-biphosphate; GAP, D-glyceraldehyde 3-phosphate; DHAP, glycerone phosphate; Glyc-1,3P_2, 3-Phospho-D-glyceroyl phosphate; Glyc-3P, 3-Phospho-D-glycerate; Glyc-2P, 2-Phospho-D-glycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; glk, glucokinase; pgi, glucose 6-phosphate isomerase; pfk, phosphofructokinase; fba, aldolase; tpi, triosephosphate isomerase; gapdh, glyceraldehyde 3-phosphate dehydrogenase; pgk, phosphoglycerate kinase; pgam, phosphoglycerate mutase; eno, enolase; pk, pyruvate kinase; g6pd, glucose 6-phosphate 1-dehydrogenase; ED, Entner-Doudoroff; PPP, Pentose phosphate pathway; TCA, Citrate Cycle.

against these organisms [155]. *Pseudomonas putida*, a bacterium known to metabolize glucose via Entner-Doudoroff pathway, as been shown to be highly tolerant to oxidative stress. However, a transgenic *P. putida* with an activated glycolysis, was proven to be sensitive to reactive oxygen species [156]. Thus, an inactivated Entner-Doudoroff pathway results in an insufficient redox potential (NADPH), necessary to resist against oxidative stress. In conclusion, this can be a possible explanation for the use of Entner-Doudoroff pathway both as a catabolic route of glucose and as a source of redox potencial (fundamental as a counteracting measure against plant defense mechanisms) for *X. fastidiosa*.

4.5.3 *Specific Growth and Exopolysaccharide production rates*

In silico simulations predicted a growth rate of 0.1125 h^{-1} and a specific production rate of 0.44 mmol/gDW/h of fastidian gum for the *X. fastidiosa* model, as shown in table 25. These results do not deviate significantly from the experimental data and corroborate with the fact that *X. fastidiosa* displays a slower growth rate and exopolysaccharide production rate than *X. campestris* [157, 158].

Table 25.: Predicted specific growth and exopolysaccharide production rates for *X. fastidiosa* and *X. campestris* models, as well as experimental data of *X. campestris* retrieved from *Letisse et al.*[140]

Source	Growth Rate (h^{-1})	Exopolysaccharide production (mmol/gDW/h)
X. fastidiosa Model	0.11	0.44
X. campestris Model	0.12	0.47
X. campestris experimental data	0.13	0.52

An assessment on other end-products of the *X. fastidiosa* metabolism was performed through a *in silico* simulation, using the previously given environmental conditions. The obtained products and the corresponding synthesis rate are compiled and presented in table 26. Both water (H_2O) and carbon dioxide (CO_2) are natural by-products of an aerobic metabolism, which results in a higher production rate for these metabolites than the remaining predicted end-products. On the other hand, predictions indicate a small production of xanthomonadin and formate while glycolaldehyde, protons and acetate are synthesized in residual amounts.

Before achieving the previously presented results, model debugging procedures had to be performed. Both the synthesis of xanthomonadin and formate were inferred by studying the organisms' metabolic network, which resulted in the addition of carriers for these metabolites, as previously stated in section 4.2.

Table 26.: Production rate of metabolites predicted by a *in silico* simulation.

Metabolite	Production Rate (mmol/gDW/h)
H_2O	11.605
CO_2	4.888
Xanthomonadin	0.0815
Formate	0.0182
Glycolaldehyde	0.00149
H^+	0.00026
Acetate	0.00016

While analysing the metabolic network, it was found that the transaldolase enzyme (EC 2.2.1.2) was not encoded in *X. fastidiosa*'s genome. As a result, D-Erythrose-4-Phosphate produced in the pentose phosphate pathway could not be maintained at a steady-state during model simulations. Since this metabolite could only be metabolized through the shikimate pathway, dead-end metabolites in the vicinity of this pathway were analysed. Through research, it was discovered that chorismate, the final metabolite of the shikimate pathway, is linked to the synthesis of ubiquinone and xanthomonadins through a bifunctional chorismatase [151], encoded in the genome of *X. fastidiosa*.

Xanthomonadins are yellow pigments with photobiological damage protection properties synthesized by *Xanthomonas* spp.[159] Essential genes involved in this metabolite production have been reported for *X. campestris* [160]. To ensure that *X. fastidiosa* was able to effectively synthesize xanthomonadins, homologues of the aforementioned essential genes were sought in the organism's genome. As a result, it was found that *X. fastidiosa* contains all the necessary genes to produce xanthomonadins, which lead to the inclusion of this metabolite in the model, resolving the steady-state issue of D-Erythrose-4-phosphate.

According to the model's predictions, formate is produced as a by-product in cofactors synthesis pathways. However, no enzymes that could metabolize the produced formate, such as formate dehydrogenase, which converts formate into carbon dioxide [161], were identified in the reconstructed network. Therefore, the formate transport reaction was introduced in the model to obtain a feasible model.

4.5.4 *Metabolism under different carbon sources*

Considering an uptake rate of 2mmol/gDW/h of glucose, the model reported in this work predicts a growth rate of $0.06h^{-1}$, in opposition to a value of $0.05h^{-1}$ for *X. campestris*. In addition, the same results are observable for growth under fructose. With this carbon

intake, it seems that *X. fastidiosa* displays a higher rate growth rate than *X. campestris*, which opposes the gathered data stating a lower growth rate for *X. fastidiosa*[157]. Hence, it is hypothesized that *X. fastidiosa* should have a slower uptake rate than *X. campestris* for carbohydrates in a *in vivo* situation. Note that, as presented in figure 9, no growth was predicted both under sucrose or mannose for *X. fastidiosa*, while the organic acid malate supports the growth of the organism.

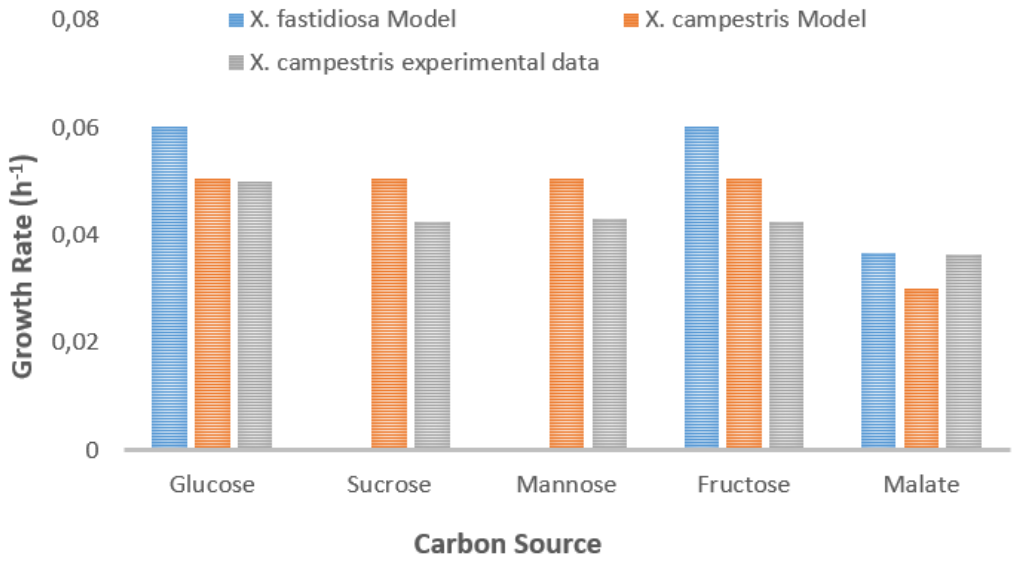


Figure 9.: Growth rates predictions under different carbon sources by *X. fastidiosa* and *X. campestris* models. Experimental data from *X. campestris* is also presented.

Growth under sucrose seems to be a highly unlikely feature for *X. fastidiosa*. Firstly, no invertase was found to be encoded in the organism's genome during the annotation procedures. Additionally, *X. fastidiosa* possesses a significantly lower number of TonB-Dependent Receptor proteins than the related xanthomonad *X. campestris* [162]. These proteins have been reported to be related to the transport of sucrose in *X. campestris*. Therefore, a homology search was performed to check if *X. fastidiosa* contained this transport mechanism. However, no homologues were found in the organism's genome, which leads to the conclusion that the phytopathogenic bacterium might be a defective sucrose scavenger.

X. fastidiosa also seems unable to use mannose as carbon source for its growth. The lack of a mannokinase enzyme in the genome of the phytopathogen, responsible for the phosphorylation of D-Mannose to D-Mannose 6-phosphate, creates a metabolic gap between D-Mannose and the remaining catabolic network. Therefore, the only possible way to use mannose as a carbon source would be through a sugar *Phosphotransferase System (PTS)*, which phosphorylates mannose upon its transport across the membrane. Auxiliary enzymes of the PTS have been found in *X. fastidiosa* but only one subunit of the mannose

permease is encoded in its genome [145]. It still remains unclear if *X. fastidiosa* is able to use a PTS to import carbohydrates. Here, it is proposed that the phytopathogen might be deficient in this sugar transport mechanism.

Growth predictions were also performed for other carbon sources, such as lactose, galactose, glycerol, citrate and succinate. *In silico* simulations predict growth under lactose, glycerol, citrate and succinate, while galactose was unable to support *X. fastidiosa*'s growth.

The organism seems to be capable to degrade lactose through a beta-galactosidase, which breaks the metabolite into glucose and galactose. This enzyme displays a strong homology to eukaryotic beta-galactosidases and it has been proposed that it might have evolved from a cellulase or was acquired by horizontal gene transfer from a plant host [163]. However, according to data from *X. campestris*, the enzyme displays a low level of substrate affinity [164].

Organic acids, such as citrate, succinate and malate, are components of the xylem sap [84], and have been reported to support the growth of *X. fastidiosa* [143]. As previously stated, the reconstructed model can accurately predict growth under these conditions. According to our predictions, to synthesize several biomass precursors using organic acids as a carbon source, it is necessary to perform gluconeogenesis. Previous studies indicated that the gluconeogenesis pathway was incomplete due to the missing fructose-1,6-biphosphatase enzyme [3]. Unlike ATP-dependent phosphofructokinase, the pyrophosphate-dependent phosphofructokinase (detected in *X. fastidiosa*) catalyzed reaction is reversible, indicating that it can replace the fructose-1,6-biphosphatase in the gluconeogenesis pathway [139]. Therefore we assume that *X. fastidiosa* can use the gluconeogenesis pathway, unlike previously thought.

4.5.5 *Amino acid auxotrophies*

In silico simulations of the model predicted growth in any singular amino acid omission, which indicates that *X. fastidiosa* does not display auxotrophy for amino acids, as shown in table 27. The fact that all the enzymes necessary for the synthesis of amino acids are encoded in *X. fastidiosa*'s genome shows the extensive metabolic biosynthetic capabilities of the organism. This might be an essential feature of the phytopathogenic bacterium, since the organism survives in a poor and limited nutrient environment. As stated in section 3.7.5, no auxotrophies are reported for *X. fastidiosa*. Moreover, as the organism is able to grow in an environment where glutamine functions as the sole nitrogen source [141], it is assumed that the reconstructed model correctly predicts the behavior of the organism.

Another aspect investigated during this stage was availability of amino acid catabolic routes for *X. fastidiosa*. According to our predicitons, glycine, alanine, asparagine, aspartate, glutamate, glutamine and serine can be used as both carbon and nitrogen sources, since

Table 27.: Results of amino acids auxotrophies obtained for *X. fastidiosa* subsp. *pauca* De Donno.

Amino Acid Omission	Predicted Growth	Auxotrophy
Glycine	Yes	No
L-Alanine	Yes	No
L-Arginine	Yes	No
L-Asparagine	Yes	No
L-Aspartate	Yes	No
L-Cysteine	Yes	No
L-Glutamate	Yes	No
L-Glutamine	Yes	No
L-Histidine	Yes	No
L-Isoleucine	Yes	No
L-Leucine	Yes	No
L-Lysine	Yes	No
L-Methionine	Yes	No
L-Phenylalanine	Yes	No
L-Proline	Yes	No
L-Serine	Yes	No
L-Threonine	Yes	No
L-Tryptophan	Yes	No
L-Tyrosine	Yes	No
L-Valine	Yes	No

enzymes related to the breakdown of these amino acids are encoded in the organisms' genome.

4.5.6 *Growth in different chemically defined media*

Qualitative analysis for growth predictions by *X. fastidiosa* model in several different chemically defined media was performed. The comparison between the results obtained by *in silico* simulations and experimental data for each of the tested media is displayed in table 28. Only XfD6, XfD8 and XfD9 do not display growth of *X. fastidiosa* experimentally, which can be caused by the lack of an iron source. Note that only one discrepancy is observed between the growth predictions results and experimental growth, which is for the XfD9 medium. Designed media by Almeida *et al.* [142], XfD1-XfD9, contain as an iron source a porphyrin (heme) or iron pyrophosphate. According to their experiments, *X. fastidiosa*

growth is not supported in media lacking porphyrin in its composition [142]. However, this opposes data from all the other developed media for *X. fastidiosa* growth, such as CHARD2, 3G10-R [115], XDM1-XDM5 [141] and XF-26 [143]. Therefore, it is assumed that the phytopathogenic bacterium is able to grow in a medium without a protoporphyrin in its composition, and that the reconstructed model accurately predicts this situation.

Table 28.: Qualitative evaluation of model predictions simulating several chemically defined media developed for *X. fastidiosa*.

Medium	Model Predicted Growth	Experimental Growth
CHARD2	Yes	Yes
3G10-R	Yes	Yes
XDM1	Yes	Yes
XDM2	Yes	Yes
XDM3	Yes	Yes
XDM4	Yes	Yes
XDM5	Yes	Yes
XF-26	Yes	Yes
XfD1	Yes	Yes
XfD2	Yes	Yes
XfD3	Yes	Yes
XfD4	Yes	Yes
XfD5	Yes	Yes
XfD6	No	No
XfD7	Yes	Yes
XfD8	No	No
XfD9	Yes	No

A common trait among all the tested media is the presence of glutamine in its composition. In fact, several media use glutamine as a sole nitrogen source (see section 3.7.6 for media composition) and support the growth of *X. fastidiosa*, which indicates a lack of auxotrophies for the organism, as stated and predicted in the previous section. The growth of *X. fastidiosa* in the CHARD2 medium also displays an interesting property of metabolic capabilities of the organism. Since this medium does not contain a carbon source in its composition, we assume that *X. fastidiosa* possesses catabolic routes for some amino acids, as indicated previously.

4.6 TOPOLOGICAL ANALYSIS OF THE MODEL

The final model of *X. fastidiosa* subsp. *pauca* De Donno is available in SBML level 3 file at *https://bit.ly/33jotWk*. Out of the 2066 genes encoded in this strain genome only 524 were actually used in the reconstruction process. This representation is relatively low when compared to the total number of genes, but nevertheless reasonable, as GSM models usually only encompass metabolic genes.

Currently, the model contains a set of 1280 metabolic reactions, whereas 881 comprise the internal metabolic network, 256 are transport reactions and the remaining 143 represent exchange reactions (Drains). A total of 246 reaction of the model are not associated to genes. These reactions encompass both drains and spontaneous reactions as well as some reactions related to the biomass pathway and inserted manually without genomic evidence. Additionally, only a small part of the reaction set is not associated to any KEGG metabolic pathway, as shown in table 29.

Table 29.: Summary of the reconstructed network of *X. fastidiosa*'s model.

	Number of Reactions
Reactions	1280
Transport Reactions	256
Drains	143
Reactions Non-associated to genes	246
Spontaneous Reactions	68
Biomass Reactions	12
Reactions Non-Associated to Pathways	97

Despite the manual curation efforts made throughout this work, several reactions may remain unconnected from the remaining metabolic network. These reactions might be associated to metabolic properties not studied during the GSM model reconstruction. However, the unconnected reactions should not cause any impairment during *in silico* simulations.

Model's reactions are distributed among various metabolic pathways. Pathways associated to carbon metabolism, amino acids and nucleic acids biosynthesis are complete, whereas pathways associated to the synthesis of antibiotics such as novobiocin and streptomicyn are heavily incomplete. Additionally, several cofactor biosynthetic pathways are present in the model. The set of pathways also include 3 manually created pathways: oxidative phosphorylation, exopolysaccharide biosynthesis and xanthomonadin biosynthesis. A detailed description of reaction distribution by pathway can be accessed through the supplementary file A.2.

4.7 DRUG TARGETS

Optflux's 'Critical genes' tool predicted a total of 223 and 214 essential genes for biofilm and planktonic growth simulations, respectively, in a CHARD2 replicated environment. However, only a small set of these critical genes had no homologies with sequences from *O. europaea*. These genes' locus tags are represented in Supplementary file A.3, were it accounts for 53 genes in the biofilm prediction and 44 genes in the planktonic prediction.

The difference between these two predictions rely on a total of 9 genes, all of them associated with the synthesis of exopolysaccharides (represented in bold in Supplementary file A.3). However, as the organism is able to grow without producing exopolysaccharides (planktonic state) and studies have shown that mutations in genes associated with exopolysaccharide production do not impair growth *in vitro* [165]. Hence, these 9 genes were not considered essential for *X. fastidiosa*. Additionally a study has shown that when DSF production is blocked, resulting in no production of exopolysaccharides, *X. fastidiosa* displays an hypervirulent phenotype [166]. Therefore, targeting genes related to the synthesis of exopolysaccharides could be counterproductive against the phytopathogen.

Genes associated to xanthomonadin synthesis reaction (B9J09_RS06665, B9J09_RS10610, B9J09_RS10585, B9J09_RS10605 and B9J09_RS10620) were predicted to be essential in both simulations. According to *Poplawsky et al.* [159], the survival rate of a *X. campestris* xanthomonadin mutant strain is tremendously lower than wild type strains, as this compound protects the organism against photobiological damage. However, despite being considered essential genes, as the xanthomonadin biosynthetic pathway is poorly described, it was not possible find any drugs that could target enzymes encoded by these genes. Nevertheless, it should be interesting to test experimentally xanthomonadin essentiality in *X. fastidiosa*.

The remaining essential genes are associated to multiple biosynthetic pathways, as shown Table 30. Note that the most prominent metabolic pathways include the synthesis of LPS, amino acid, cofactors, CoA and the shikimate pathway.

Table 30.: Pathways or metabolic functions associated to essential genes.

Gene	Biosynthetic Pathway/Function
B9J09_RS07735	LPS synthesis
B9J09_RS08985	Encodes ATP Synthase subunit
B9J09_RS00100	Shikimate Pathway
B9J09_RS00145	Pyrimidine/Purine synthesis
B9J09_RS00430	LPS synthesis
B9J09_RS00480	Amino Acid synthesis (Lysine)

B9J09_RS00870	Phospholipid synthesis
B9J09_RS01165	Nucleotide Sugar synthesis
B9J09_RS03245	Pyrimidine/Purine synthesis
B9J09_RS03250	Cofactor synthesis
B9J09_RS03315	Amino Acid synthesis (Tyrosine)
B9J09_RS03845	Amino Acid synthesis (Histidine)
B9J09_RS04015	Cofactor synthesis
B9J09_RS04800	Pyrimidine/Purine synthesis
B9J09_RS04890	Cofactor synthesis
B9J09_RS06005	Phospholipid synthesis
B9J09_RS06185	Cofactor synthesis
B9J09_RS06205	CoA synthesis
B9J09_RS08675	Phospholipid synthesis
B9J09_RS09170	Amino Acid synthesis (Lysine)
B9J09_RS09480	Phospholipid synthesis
B9J09_RS09830	Cofactor synthesis
B9J09_RS10190	Amino Acid synthesis (Methionine)
B9J09_RS10245	D-Alanine synthesis
B9J09_RS10630	Shikimate Pathway
B9J09_RS10990	Phospholipid synthesis
B9J09_RS11045	LPS synthesis
B9J09_RS01155	LPS synthesis
B9J09_RS00200	Shikimate Pathway
B9J09_RS00940	Amino Acid synthesis (Serine)
B9J09_RS08000	LPS synthesis
B9J09_RS07150	Cofactor synthesis
B9J09_RS00885	Cofactor synthesis
B9J09_RS01020	CoA synthesis
B9J09_RS01930	Phospholipid synthesis

B9J09_RS03580	Transport of carbon based molecules
B9J09_RS09020	Nucleotide Sugar synthesis
B9J09_RS09690	CoA synthesis
B9J09_RS02865	Pyrimidine/Purine synthesis

Only experimental drugs with unknown action mechanisms were found in DrugBank [144], for the enzymes encoded by the predicted essential genes. Therefore, it was not possible to find any inhibiting drugs for essential genes of *X. fastidiosa*. Nevertheless, the number of possible drug targets from a much larger set of metabolic genes was narrowed down.

It was determined that genes B9J09_RS10630 and B9J09_RS00885 are relevant as potential drug targets of *X. fastidiosa*. Gene B9J09_RS10630 encodes a bifunctional chorismatase, required for both the synthesis of ubiquinone and xanthomonadin [151]. This gene deletion mutant should be impaired in ubiquinone and xanthomonadin synthesis, which could reduce greatly the organism defense against oxidative stress, as both metabolites are known to provide resistance to oxidative stress [167, 159]. Other potential drug target, gene B9J09_RS00885, encodes a novel protoporphyrinogen oxidase. By disrupting this gene the organism should be unable of *de novo* synthesis of heme, which is essential for respiratory processes in aerobic organisms [168]. Additionally, it is assumed that targeting amino acids biosynthetic pathways could severely affect the organism survival, as its *in vivo* environment is very nutrient-limited.

5

CONCLUSION AND FUTURE WORK

In this work, a GSM model of *X. fastidiosa* subsp. *pauca* De Donno (iMS524) is presented. To the best of our knowledge, it is the first metabolic model of the phytopathogenic bacterium. This manually curated GSM model is capable of simulating the phenotypic behavior of the organism, which can serve as a starting point to unveil various metabolic properties of *X. fastidiosa* as well as to identify new therapeutic targets to fight off this agricultural-destructive phytopathogen.

High-quality genome annotation and extensive manual curation efforts, based on literature, biological databases, organism-specific and related-organisms information, were made in order to elevate the model's predictive capabilities and reliability. Therefore, this model can simulate two stages of *X. fastidiosa* life cycle *in planta*, the planktonic and biofilm phases, as well as several environmental conditions with trustworthy qualitative and quantitative results.

Model's predictions indicate the organism's extensive biosynthetic metabolic capabilites. For instance, all amino acids synthesis pathways are available in *X. fastidiosa*'s genome, and its functionality was shown through the reconstructed metabolic network, proving the lack of amino acid auxotrophies in *X. fastidiosa*. In addition, the phytopathogen's carbohydrate utilization was thoroughly explored, in which specific metabolic properties were found. The usage of Entner-Doudoroff pathway as means to degrade carbon might confer an advantage during growth *in planta*, since it generates redox potential necessary to defend the organism against oxidative stress.

Coupling the organism's extensive metabolic potential with fact that it infects and survives within xylem vessels, a nutrient scarce environment, the model accurately predicts *X. fastidiosa*'s growth under these nutrient-limited conditions.

One of the major goals of this work was to identify potential drug targets of the phytopathogenic bacterium. Despite the lack of information on inhibiting drugs for the identified essential genes, we determined an interesting set of genes/enzymes that should be subjected to further research. Additionally, other environmental conditions could be applied to determine potential drug targets for the organism.

The overall biomass composition adopted for this model was based on data retrieved essentially from literature of closely-related organisms and other reconstructed GSM models. Although this does not pose as a problem for model's *in silico* simulations, organism-specific information on biomass composition would enhance the model's reliability even further. In addition, the lack of experimental data for *X. fastidiosa* limits the extent of applications for the reconstructed model. Therefore, as long as new experimental data regarding the phytopathigen is published, further updates and improvements to the GSM model can be performed at any time, if needed.

The periplasmatic space, a delimited space between the inner and outer membranes in gram-negative bacteria, is not included in the reconstructed model. Including this compartment, could provide additional insight on metabolite movement across both membranes of the organism.

An important aspect of *X. fastidiosa*'s life cycle is its transmission through an insect vector. While this was not explored throughout this work, the model could possibly be used to study metabolic properties while *X. fastidiosa* resides inside its vector. The utilization of chitin by the bacterium as carbon source seems to be a required trait for the organism host colonization [169, 170] and should also be accessed in future work.

The present model could serve as a base to develop other *X. fastidiosa* strains' GSM models. By comparing the metabolic capabilities among different subspecies of *X. fastidiosa* it could be possible to check if the metabolic potential of these subspecies is correlated to their respective host specificity range. In addition, identifying potential drug targets in other strains and subspecies of *X. fastidiosa* could pinpoint a common target among subspecies.

BIBLIOGRAPHY

[1] Isabel Rocha, Jochen Förster, and Jens Nielsen. Design and Application of Genome-Scale Reconstructed Metabolic Models. *Methods in Molecular Biology*, 416:409–431, 2008.

[2] Ines Thiele and Bernhard Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121, 2010.

[3] A. J.G. Simpson, F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M.C. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bové, M. R.S. Briones, M. R.P. Bueno, A. A. Camargo, L. E.A. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C.R. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J.S. Ferreira, V. C.A. Ferreira, J. A. Ferro, J. S. Fraga, S. C. França, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H.S. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C.C. Leite, E. G.M. Lemos, M. V.F. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M.B.N. Madeira, H. M.F. Madeira, C. L. Marino, M. V. Marques, E. A.L. Martins, E. M.F. Martins, A. Y. Matsukuma, C. F.M. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteiro-Vitorello, D. H. Moon, M. A. Nagai, A. L.T.O. Nascimento, L. E.S. Netto, A. Nhani, F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A.G. Pereira, H. A. Pereira, J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J.M. Rosa, V. E. de Rosa, R. G. de Sá, R. V. Santelli, H. E. Sawasaki, A. C.R. da Silva, A. M. da Silva, F. R. da Silva, W. A. Silva, J. F. da Silveira, M. L.Z. Silvestri, W. J. Siqueira, A. A. de Souza, A. P. de Souza, M. F. Terenzi, D. Truffi, S. M. Tsai, M. H. Tsuhako, H. Vallada, M. A. Van Sluys, S. Verjovski-Almeida, A. L. Vettore, M. A. Zago, M. Zatz, J. Meidanis, and J. C. Setubal. The genome sequence of the plant pathogen Xylella fastidiosa. *Nature*, 406(6792):151–157, 2000.

[4] John M. Wells, Boligala C. Raju, Hsueh-Yun Hung, William G. Weisburg, Linda Mandelco-Paul, and Don J. Brenner. Xylella fastidiosa gen. nov. , sp. nov: Gram-Negative, Xylem- Limited, Fastidious Plant Bacteria Related to Xanthomonas spp. *International Journal of Systematic Bacteriology*, 37(2):136–143, 1987.

[5] D. L. Hopkins and A. H. Purcell. Xylella fastidiosa: Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. *Plant disease*, 86(10):1056–1066, 2002.

[6] Subhadeep Chatterjee, Rodrigo P. P Almeida, and Steven Lindow. Living in two Worlds: The Plant and Insect Lifestyles of Xylella fastidiosa. *Annual Review of Phytopathology*, 46(1):243–271, 2008.

[7] Richard A. Redak, Alexander H. Purcell, João R.S. Lopes, Matthew J. Blua, Russell F. Mizell III, and Peter C. Andersen. The Biology of Xylem FluidFeeding Insect Vectors of Xylella fastidiosa and their relation to Disease Epidemiology. *Annual Review of Entomology*, 49(1):243–270, 2004.

[8] Stefania M. Mang, Salvatore Frisullo, Hazem S. Elshafie, and Ippolito Camele. Diversity evaluation of Xylella fastidiosa from infected olive trees in Apulia (Southern Italy). *Plant Pathology Journal*, 32(2):102–111, 2016.

[9] Sybren Vos, Melanie Camilleri, Makrina Diakaki, Elena Lázaro, Stephen Parnell, Gritta Schrader, and Antonio Vicent. Pest survey card on Xylella fastidiosa. *EFSA Supporting Publications*, 2019.

[10] N. Denancé, B. Legendre, M. Briand, V. Olivier, C. de Boisseson, F. Poliakoff, and M. A. Jacques. Several subspecies and sequence types are associated with the emergence of Xylella fastidiosa in natural settings in France. *Plant Pathology*, 66(7):1054–1064, 2017.

[11] Hiroaki Kitano. Systems Biology: A Brief Overview. *Science*, 295(March):1662–1664, 2002.

[12] Bernhard Ø. Palsson. *Systems Biology - Properties of Reconstructed Networks*. 2006.

[13] Adam M. Feist, Markus J. Herrgård, Ines Thiele, Jennie L. Reed, and Bernhard Ø. Palsson. Reconstruction of Biochemical Networks in Microbial Organisms. *Nature Reviews Microbiology*, 7(2):129–143, 2009.

[14] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Ø. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, 2007.

[15] Adam M. Feist and Bernhard Ø. Palsson. The Growing Scope of Applications of Genome-scale Metabolic Reconstructions: the case of E. coli. *Nature Biotechnology*, 26(6):659–667, 2008.

[16] Jochen Forster, Iman Famili, Patrick Fu, Bernhard Ø. Palsson, and Jens Nielsen. Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome Research*, 13:244–253, 2003.

[17] Natalie C. Duarte, Markus J. Herrgård, and Bernhard Ø. Palsson. Reconstruction and Validation of Saccharomyces cerevisiae iND750 , a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Research*, 14:1298–1309, 2004.

[18] Markus J. Herrgård, Neil Swainston, Paul Dobson, Warwick B. Dunn, K. Yalçin Arga, Mikko Arvas, Nils Blüthgen, Simon Borger, Roeland Costenoble, Matthias Heinemann, Michael Hucka, Nocalas Le Novère, Peter Li, Wolfram Libermeister, Monica L. Mo, Ana Paula Oliveira, Dina Petranovic, Stephen Pettifer, Evagelos Simeonidis, Kieran Smallbone, Irena Spasić, Dieter Weichart, Roger Brent, David S. Broomhead, Hans V. Westerhoff, Betül Kirdar, Merja Penttilä, Edda Klipp, Bernhard Ø. Palsson, Uwe Sauer, Stephen G. Oliver, Pedro Mendes, Jens Nielsen, and Douglas B. Kell. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26(10):1155–1160, 2008.

[19] Oscar Dias and Isabel Rocha. Systems Biology in Fungi. In *Mucormycosis. Food and Water Borne Mycotoxigenic and Mycotic Fungi*, chapter 6, pages 69–88. 2015.

[20] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[21] Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard Ø. Palsson, and Nathan E. Lewis. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):515–522, 2016.

[22] Ron Caspi, Richard Billington, Luciana Ferrer, Hartmut Foerster, Carol A. Fulcher, Ingrid M. Keseler, Anamika Kothari, Markus Krummenacker, Mario Latendresse, Lukas A. Mueller, Quang Ong, Suzanne Paley, Pallavi Subhraveti, Daniel S. Weaver, and Peter D. Karp. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 44(D1):D471–D480, 2016.

[23] Lisa Jeske, Sandra Placzek, Ida Schomburg, Antje Chang, and Dietmar Schomburg. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research*, pages 1–8, 2019.

[24] Panu Artimo, Manohar Jonnalagedda, Konstantin Arnold, Delphine Baratin, Gabor Csardi, Edouard De Castro, Séverine Duvaud, Volker Flegel, Arnaud Fortier, Elisabeth Gasteiger, Aurélien Grosdidier, Céline Hernandez, Vassilios Ioannidis, Dmitry Kuznetsov, Robin Liechti, Sébastien Moretti, Khaled Mostaguir, Nicole Redaschi, Grégoire Rossier, Ioannis Xenarios, and Heinz Stockinger. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1):597–603, 2012.

[25] Minoru Kanehisa, Yoko Sato, Miho Furumichi, Kanae Morishima, and Mao Tanabe. New approach for understanding genome variations in KEGG. *Nucleic Acids Research*, pages 1–6, 2018.

[26] Eric W. Sayers, Tanya Barrett, Dennis A. Benson, Evan Bolton, Stephen H. Bryant, Kathi Canese, Vyacheslav Chetvernin, Deanna M. Church, Michael Dicuccio, Scott Federhen, Michael Feolo, Lewis Y. Geer, Wolfgang Helmberg, Yuri Kapustin, David Landsman, David J. Lipman, Zhiyong Lu, Thomas L. Madden, Tom Madej, Donna R. Maglott, Aron Marchler-Bauer, Vadim Miller, Ilene Mizrachi, James Ostell, Anna Panchenko, Kim D. Pruitt, Gregory D. Schuler, Edwin Sequeira, Stephen T. Sherry, Martin Shumway, Karl Sirotkin, Douglas Slotta, Alexandre Souvorov, Grigory Starchenko, Tatiana A. Tatusova, Lukas Wagner, Yanli Wang, W. John Wilbur, Eugene Yaschenko, and Jian Ye. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 38(D1):5–16, 2009.

[27] Milton H. Saier, Vamsee S. Reddy, Brian V. Tsu, Muhammad Saad Ahmed, Chun Li, and Gabriel Moreno-Hagelsieb. The Transporter Classification Database (TCDB): Recent advances. *Nucleic Acids Research*, 44(D1):D372–D379, 2016.

[28] Alex Bateman, Maria Jesus Martin, Claire O'Donovan, Michele Magrane, Emanuele Alpi, Ricardo Antunes, Benoit Bely, Mark Bingley, Carlos Bonilla, Ramona Britto, Borisas Bursteinas, Hema Bye-AJee, Andrew Cowley, Alan Da Silva, Maurizio De Giorgi, Tunca Dogan, Francesco Fazzini, Leyla Garcia Castro, Luis Figueira, Penelope Garmiri, George Georghiou, Daniel Gonzalez, Emma Hatton-Ellis, Weizhong Li, Wudong Liu, Rodrigo Lopez, Jie Luo, Yvonne Lussi, Alistair MacDougall, Andrew Nightingale, Barbara Palka, Klemens Pichler, Diego Poggioli, Sangya Pundir, Luis Pureza, Guoying Qi, Steven Rosanoff, Rabie Saidi, Tony Sawford, Aleksandra Shypitsyna, Elena Speretta, Edward Turner, Nidhi Tyagi, Vladimir Volynkin, Tony Wardell, Kate Warner, Xavier Watkins, Rossana Zaru, Hermann Zellner, Ioannis Xenarios, Lydie Bougueleret, Alan Bridge, Sylvain Poux, Nicole Redaschi, Lucila Aimo, Ghislaine ArgoudPuy, Andrea Auchincloss, Kristian Axelsen, Parit Bansal, Delphine Baratin, Marie Claude Blatter, Brigitte Boeckmann, Jerven Bolleman, Emmanuel Boutet, Lionel Breuza, Cristina Casal-Casas, Edouard De Castro, Elisabeth Coudert, Beatrice Cuche,

Mikael Doche, Dolnide Dornevil, Severine Duvaud, Anne Estreicher, Livia Famiglietti, Marc Feuermann, Elisabeth Gasteiger, Sebastien Gehant, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz-Gumowski, Ursula Hinz, Chantal Hulo, Florence Jungo, Guillaume Keller, Vicente Lara, Philippe Lemercier, Damien Lieberherr, Thierry Lombardot, Xavier Martin, Patrick Masson, Anne Morgat, Teresa Neto, Nevila Nouspikel, Salvo Paesano, Ivo Pedruzzi, Sandrine Pilbout, Monica Pozzato, Manuela Pruess, Catherine Rivoire, Bernd Roechert, Michel Schneider, Christian Sigrist, Karin Sonesson, Sylvie Staehli, Andre Stutz, Shyamala Sundaram, Michael Tognolli, Laure Verbregue, Anne Lise Veuthey, Cathy H. Wu, Cecilia N. Arighi, Leslie Arminski, Chuming Chen, Yongxing Chen, John S. Garavelli, Hongzhan Huang, Kati Laiho, Peter McGarvey, Darren A. Natale, Karen Ross, C. R. Vinayaka, Qinghua Wang, Yuqi Wang, Lai Su Yeh, and Jian Zhang. UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017.

[29] Oscar Dias, Miguel Rocha, Eugénio C. Ferreira, and Isabel Rocha. Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Research*, 43(8):3899–3910, 2015.

[30] Adam P. Arkin, Robert W. Cottingham, Christopher S. Henry, Nomi L. Harris, Rick L. Stevens, Sergei Maslov, Paramvir Dehal, Doreen Ware, Fernando Perez, Shane Canon, Michael W. Sneddon, Matthew L. Henderson, Wiliam J. Riehl, Dan Murphy-Olson, Stephen Y. Chan, Roy T. Kamimura, Sunita Kumari, Meghan M. Drake, Thomas S. Brettin, Elizabeth M. Glass, Dylan Chivian, Dan Gunter, David J. Weston, Benjamin H. Allen, Jason Baumohl, Aaron A. Best, Ben Bowen, Steven E. Brenner, Christopher C. Bun, John-Marc Chandonia, Jer-Ming Chia, Ric Colasanti, Neal Conrad, James J. Davis, Brian H. Davison, Matthew DeJongh, Scott Devoid, Emily Dietrich, Inna Dubchak, Janaka N. Edirisinghe, Gang Fang, José P. Faria, Paul M. Frybarger, Wolfgang Gerlach, Mark Gerstein, Annette Greiner, James Gurtowski, Holly L. Haun, Fei He, Rashmi Jain, Marcin P. Joachimiak, Kevin P. Keegan, Shinnosuke Kondo, Vivek Kumar, Miriam L. Land, Folker Meyer, Marissa Mills, Pavel S. Novichkov, Taeyun Oh, Gary J. Olsen, Robert Olsen, Bruce Parrello, Shiran Pasternak, Erik Pearson, Sarah S. Poon, Gavin A. Price, Srividya Ramakrishnan, Priya Ranjan, Pamela C. Ronald, Michael C. Shatz, Samuel M. D. Seaver, Maulik Shukla, Roman A. Sutormin, Mustafa H. Syed, James Thomason, Nathan L. Tintle, Daifeng Wang, Fangfang Xia, Hyunseung Yoo, Shinjae Yoo, and Dantong Yu. KBase : The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7):566–569, 2018.

[31] Christopher S. Henry, Matthew Dejongh, Aaron A. Best, Paul M. Frybarger, and Ben Linsay. High-throughput Generation , Optimization and Analysis of Genome-scale

Metabolic Models. *Nature Biotechnology*, 28(9):977–U22, 2010.

[32] Daniel Machado, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 2018.

[33] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13:2498–2504, 2003.

[34] Xueyang Feng, You Xu, Yixin Chen, and Yinjie J Tang. MicrobesFlux : a web platform for drafting metabolic models from the KEGG database. *BMC Systems Biology*, 6(94):1–9, 2012.

[35] Peter D. Karp, Mario Latendresse, Suzanne M. Paley, Markus Krummenacker Quang Ong, Richard Billington, Anamika Kothari, Daniel Weaver, Tom Lee, Pallavi Subhraveti, Aaron Spaulding, Carol Fulcher, Ingrid M. Keseler, and Ron Caspi. Pathway Tools version 19 . 0 : Integrated Software for Pathway / Genome Informatics and Systems Biology. 2015.

[36] Akira Funahashi, Yukiko Matsuoka, Akiya Jouraku, Mineo Morohashi, Norihiro Kikuchi, and Hiroaki Kitano. CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE*, 96(8):1254–1265, 2008.

[37] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–10, 1990.

[38] Sean R. Eddy. Multiple alignment using hidden Markov model. *Intelligent Systems for Molecular Biology*, 95:114–120, 1995.

[39] International Union of Biochemistry, Molecular Biology. Nomenclature Committee, and E.C. Webb. *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. Published for the International Union of Biochemistry and Molecular Biology by Academic Press, 1992.

[40] Sinead Boyce and Keith Tipton. Enzyme Classification and Nomenclature. *Encyclopedia of Life Sciences*, pages 1–11, 2001.

[41] Nancy Y. Yu, James R. Wagner, Matthew R. Laird, Gabor Melli, Sébastien Rey, Raymond Lo, Phuong Dao, S. Cenk Sahinalp, Martin Ester, Leonard J. Foster, and

Fiona S.L. Brinkman. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13):1608–1615, 2010.

[42] Paul Horton, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C. J. Adams-Collier, and Kenta Nakai. WoLF PSORT: Protein localization predictor. *Nucleic Acids Research*, 35:585–587, 2007.

[43] Olof Emanuelsson, Henrik Nielsen, Søren Brunak, and Gunnar Von Heijne. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology*, 300(4):1005–1016, 2000.

[44] Tatyana Goldberg, Maximilian Hecht, Tobias Hamp, Timothy Karl, Guy Yachdav, Nadeem Ahmed, Uwe Altermann, Philipp Angerer, Sonja Ansorge, Kinga Balasz, Michael Bernhofer, Alexander Betz, Laura Cizmadija, Kieu Trinh Do, Julia Gerke, Robert Greil, Vadim Joerdens, Maximilian Hastreiter, Katharina Hembach, Max Herzog, Maria Kalemanov, Michael Kluge, Alice Meier, Hassan Nasir, Ulrich Neumaier, Verena Prade, Jonas Reeb, Aleksandr Sorokoumov, Ilira Troshani, Susann Vorberg, Sonja Waldraff, Jonas Zierer, Henrik Nielsen, and Burkhard Rost. LocTree3 prediction of localization. *Nucleic Acids Research*, 42(W1):350–355, 2014.

[45] Z. Lu, D. Szafron, R. Greiner, P. Lu, D. S. Wishart, B. Poulin, J. Anvik, C. Macdonell, and R. Eisner. Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics*, 20(4):547–556, 2004.

[46] Christof Francke, Roland J. Siezen, and Bas Teusink. Reconstructing the metabolic network of a bacterium from its genome. *Trends in Microbiology*, 13(11):550–558, 2005.

[47] Richard A. Notebaart, Frank H. J. van Enckevort, Christof Francke, Roland J. Siezen, and Bas Teusink. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics*, 7(296):1–10, 2006.

[48] Matthew D. Jankowski, Christopher S. Henry, Linda J. Broadbelt, and Vassily Hatzimanikatis. Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. *Biophysical Journal*, 95:1487–1499, 2008.

[49] R.M.T. Fleming, I. Thiele, and H.P. Nasheuer. Quantitative assignment of reaction directionality in constraint- based models of metabolism: Application to Escherichia coli. *Biophysical Chemistry*, 145(2-3):47–56, 2009.

[50] Avi Flamholz, Elad Noor, Arren Bar-Even, and Ron Milo. eQuilibrator the biochemical thermodynamics calculator. *Nucleic Acids Researche*, 40(D1):770–775, 2012.

[51] Jacques Izard and Ronald J. Limberger. Rapid screening method for quantitation of bacterial cell lipids from whole cells. *Journal of Microbiological Methods*, 55:411–418, 2003.

[52] S. Benthin, J. Nielsen, and J. Villadsen. A simple and reliable method for the determination of cellular RNA content. *Biotechnology Techniques*, 5(1):39–42, 1991.

[53] D. Herbert, P.J. Phipps, and R.E. Strange. Chemical Analysis of Microbial Cells. In *Methods in Microbiology*, volume 5, chapter 3, pages 209–344. 1971.

[54] Adam M. Feist, Christopher S. Henry, Jennifer L. Reed, Markus Krummenacker, Andrew R. Joyce, Peter D. Karp, Linda J. Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 1990(121):1–18, 2007.

[55] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010.

[56] Isabel Rocha, Paulo Maia, Pedro Evangelista, Paulo Vilaça, Simão Soares, José P. Pinto, Jens Nielsen, Kiran R. Patil, Eugénio C. Ferreira, and Miguel Rocha. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4(45):1–12, 2010.

[57] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, Jacek Wachowiak, Sarah M. Keating, Vanja Vlasov, Stefania Magnusdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H.J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A.P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Inna Kuperstein, Andrei Zinovyev, H. Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard Ø. Palsson, Ines Thiele, and Ronan M.T. Fleming. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nature Protocols*, 2(3):727–738, 2018.

[58] José Manuel Otero and Jens Nielsen. Industrial Systems Biology. *Biotechnology and Bioengineering*, 105(3):439–460, 2010.

[59] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, 3(119):1–15, 2007.

[60] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *PNAS*, 99(23):15112–15117, 2002.

[61] Tomer Shlomi, Omer Berkman, and Eytan Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *PNAS*, 102(21):7695–7700, 2005.

[62] Oscar Dias, Daniel Gomes, Paulo Vilaça, João Cardoso, Miguel Rocha, Eugénio C. Ferreira, and Isabel Rocha. Genome-wide Semi-automated Annotation of Transporter Systems. *Transactions on Computational Biology and Bioinformatics*, pages 1–15, 2016.

[63] Anders Krogh, Björn Larsson, Gunnar Von Heijne, and Erik L.L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, 2001.

[64] Moller Steffen, Michael D. R. Croning, and Rolf Apweiler. Evaluation ofmethods for the prediction of membrane spanning regions. *Bioinformatics*, 17(7):646–653, 2001.

[65] Lukas Käll, Anders Krogh, and Erik L. L. Sonnhammer. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5):1027–1036, 2004.

[66] Lukas Käll, Anders Krogh, and Erik L. L. Sonnhammer. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Research*, 35:429–432, 2007.

[67] T.F. Smith and M.S. Waterman. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147:195–197, 1981.

[68] G. Stephanopoulos, A.A. Aristidou, and J. Nielsen. *Metabolic Engineering: Principles and Methodologies*. Elsevier Science, 1998.

[69] Stefan Schuster, David A. Fell, and Thomas Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(March):326–332, 2000.

[70] Kiran Raosaheb Patil, Isabel Rocha, Jochen Förster, and Jens Nielsen. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 6(308):1–12, 2005.

[71] Miguel Rocha, Paulo Maia, Rui Mendes, José P. Pinto, Eugénio C. Ferreira, Jens Nielsen, Kiran Raosaheb Patil, and Isabel Rocha. Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC Bioinformatics*, 9(499):1–16, 2008.

[72] Anthony P. Burgard, Priti Pharkya, and Costas D. Maranas. OptKnock : A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnology and Bioengineering*, 85(7):1–11, 2003.

[73] Scott A. Becker, Adam M. Feist, Monica L. Mo, Gregory Hannum, Bernhard Ø. Palsson, and Markus J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models : the COBRA Toolbox. *Nature Protocols*, 2(3):727–738, 2007.

[74] Jan Schellenberger, Richard Que, Ronan M. T. Fleming, Ines Thiele, Jeffrey D. Orth, Adam M. Feist, Daniel C. Zielinski, Aarash Bordbar, Nathan E. Lewis, Sorena Rahmanian, Joseph Kang, Daniel R. Hyduke, and Bernhard Ø. Palsson. Quantitative prediction of cellular metabolism with constraint- based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–1307, 2012.

[75] Ali Ebrahim, Joshua A. Lerman, Bernhard Ø. Palsson, and Daniel R. Hyduke. COBRApy : COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(74):1–6, 2013.

[76] Markus W. Covert, Eric M. Knight, Jennifer L. Reed, Markus J. Herragard, and Bernhard Ø. Palsson. Integrating high-throughput and computational data. *Nature*, 429(May):2–6, 2004.

[77] Erwin P. Gianchandani, Andrew R. Joyce, Bernhard Ø. Palsson, and Jason A. Papin. Functional States of the Genome-Scale Escherichia Coli Transcriptional Regulatory System. *PLoS Computational Biology*, 5(6):1–18, 2009.

[78] Jason A. Papin, Tony Hunter, Bernhard Ø. Palsson, and Shankar Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, 6:99–111, 2005.

[79] Fan Li, Ines Thiele, Neema Jamshidi, and Bernhard Ø. Palsson. Identification of Potential Pathway Mediation Targets in Toll-like Receptor Signaling. *PLoS Computational Biology*, 5(2), 2009.

[80] Karthik Raman, Rohit Vashisht, and Nagasuma Chandra. Strategies for efficient disruption of metabolism in Mycobacterium tuberculosis from network analysis. *Molecular BioSystems*, 5:1740–1751, 2009.

[81] Adam M. Feist, Daniel C. Zielinski, Jeffrey D. Orth, Jan Schellenberger, Markus J. Herrgard, and Bernhard Ø. Palsson. Model-driven evaluation of the production potential for growth- coupled products of Escherichia coli. *Metabolic Engineering*, 12(3):173–186, 2010.

[82] Matthew A. Oberhardt, Bernhard Ø. Palsson, and Jason A. Papin. Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, 5(320):1–15, 2009.

[83] Alexander A. Myburg, Simcha Lev-Yadun, and Ronald R. Sederoff. Xylem Structure and Function. *Encyclopedia of Life Sciences*, pages 1–19, 2013.

[84] Nabil Killiny and Faraj Hijaz. Chemical composition of xylem sap of citrus sinensis L . Osbeck (sweet orange). *Proceedings of the Florida State Horticultural Society*, 128(1):114–118, 2015.

[85] Jennifer J. Randall, Natalie P. Goldberg, John D. Kemp, Maxim Radionenko, Jason M. French, Mary W. Olsen, and Stephen F. Hanson. Genetic analysis of a novel Xylella fastidiosa subspecies found in the Southwestern United States. *Applied and Environmental Microbiology*, 75(17):5631–5638, 2009.

[86] Xiaoli Yuan, Lisa Morano, Robin Bromley, Senanu Spring-Pearson, Richard Stouthamer, and Leonard Nunney. Multilocus Sequence Typing of Xylella fastidiosa Causing Pierce's Disease and Oleander Leaf Scorch in the United States. *Phytopathology*, 100(6):601–611, 2010.

[87] Michael Jeger, David Caf, Thierry Candresse, Elisavet Chatzivassiliou, Josep Anton, Jaques Miret, Katharina Dehnen-schmutz, Gianni Gilioli, Jean-claude Gr, Stephen Parnell, Roel Potting, Trond Rafoss, Alan Macleod, Maria Navajas Navarro, Vittorio Rossi, Gregor Urek, Ariena Van Bruggen, Wopke Van Der Werf, Jonathan West, Blanca Landa, Stephan Winter, Rodrigo Almeida, Domenico Bosco, Alexander Purcell, Maria Saponari, Ewelina Czwienczek, Alice Delbianco, Giuseppe Stancanelli, and Claude Bragard. Updated pest categorisation of Xylella fastidiosa. *EFSA Journal*, 16, 2018.

[88] Norman W. Schaad, Elena Postnikova, George Lacy, M'Barek Fatmi, and Chung Jan Chang. Xylella fastidiosa subspecies: X. fastidiosa subsp piercei, subsp. nov., X. fastidiosa subsp. multiplex subsp. nov., and X. fastidiosa subsp. pauca subsp. nov. *Systematic and Applied Microbiology*, 27(3):290–300, 2004.

[89] European Food Safety Authority (EFSA). Update of the Xylella spp . host plant database. Technical Report July, 2018.

[90] Leonard Nunney, Danel B. Vickerman, Robin E. Bromley, Stephanie A. Russell, John R. Hartman, Lisa D. Morano, and Richard Stouthamer. Recent Evolutionary Radiation

and Host Plant Specialization in the Xylella fastidiosa Subspecies Native to the United States. *Applied and Environmental Microbiology*, 79(7):2189–2200, 2013.

[91] Paolo Baldi and Nicola La Porta. Xylella fastidiosa : Host Range and Advance in Molecular Identification Techniques. *Frontiers in Plant Science*, 8:1–22, 2017.

[92] James H. Graham, Tim R. Gottwald, Jaime Cubero, and Diann S. Achor. Xanthomonas axonopodis pv . citri: factors affecting successful eradication of citrus canker. *Molecular Plant Pathology*, 5(1):1–15, 2004.

[93] Caroline Roper, Claudia Castro, and Brian Ingel. Xylella fastidiosa : bacterial parasitism with hallmarks of commensalism. *Current Opinion in Plant Biology*, 50:140–147, 2019.

[94] Enrico M. Bucci. Xylella fastidiosa, a new plant pathogen that threatens global farming: Ecology, molecular biology, search for remedies. *Biochemical and Biophysical Research Communications*, 502(2):173–182, 2018.

[95] Erin L. Schuenzel, Mark Scally, Richard Stouthamer, and Leonard Nunney. A multigene phylogenetic study of clonal diversity and divergence in North American strains of the plant pathogen Xylella fastidiosa. *Applied and Environmental Microbiology*, 71(7):3832–3839, 2005.

[96] M. A. Van Sluys, M. C. de Oliveira, C. B. Monteiro-Vitorello, C. Y. Miyaki, L. R. Furlan, L. E. A. Camargo, A. C.R. da Silva, D. H. Moon, M. A. Takita, E. G. M. Lemos, M. A. Machado, M. I. T. Ferro, F. R. da Silva, M. H. S. Goldman, G. H. Goldman, M. V. F. Lemos, H. El-Dorry, S. M. Tsai, H. Carrer, D. M. Carraro, R. C. de Oliveira, L. R. Nunes, W. J. Siqueira, L. L. Coutinho, E. T. Kimura, E. S. Ferro, R. Harakava, E. E. Kuramae, C. L. Marino, E. Giglioti, I. L. Abreu, L. M. C. Alves, A. M. do Amaral, G. S. Baia, S. R. Blanco, M. S. Brito, F. S. Cannavan, A. V. Celestino, A. F. da Cunha, R. C. Fenille, J. A. Ferro, E. F. Formighieri, L. T. Kishi, S. G. Leoni, A. R. Oliveira, V. E. Rosa Jr, F. T. Sassaki, J. A. D. Sena, A. A. de Souza, D. Truffi, F. Tsukumo, G. M. Yanai, L. G. Zaros, E. L. Civerolo, A. J. G. Simpson, N. F. Almeida Jr, J. C. Setubal, and J. P. Kitajima. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of Xylella fastidiosa. *Journal of Bacteriology*, 185(3):1018–1026, 2003.

[97] Welington L. Araújo, Joelma Marcon, Walter Maccheroni Jr, Jan Dirk van Elsas, Jim W. L. van Vuurde, and João Lúcio Azevedo. Diversity of Endophytic Bacterial Populations and Their Interaction with Xylella fastidiosa in Citrus Plants. *Applied and Environmental Microbiology*, 68(10):4906–4914, 2002.

[98] Toufic Elbeaino, Franco Valentini, Raied Aboud Kubaa, Peter Moubarak, Thaer Yaseen, and Michele Digiaro. Multilocus sequence typing of Xylella fastidiosa isolated from olive affected by olive quick decline syndrome in Italy. *Phytopathologia Mediterranea*, 53(3):533–542, 2014.

[99] Andrew J. McElrone, James L. Sherald, and Irwin N. Forseth. Effects of Water Stress on Symptomatology and Growth of Parthenocissus quinquefolia Infected by Xylella fastidiosa. *Plant Disease*, 85(11):1160–1164, 2001.

[100] G. P. Martelli, D. Boscia, F. Porcelli, and M. Saponari. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *European Journal of Plant Pathology*, 2015.

[101] J. H. Lieth, M. M. Meyer, K. H. Yeo, and B. C. Kirkpatrick. Modeling Cold Curing of Pierce's Disease in Vitis vinifera Pinot Noir' and Cabernet Sauvignon' Grapevines in California. *Phytopathology*, 101(12):1492–1500, 2011.

[102] Andrea Baù, Alice Delbianco, Giuseppe Stancanelli, and Sara Tramontini. Susceptibility of Olea europaea L. varieties to Xylella fastidiosa subsp. pauca ST53: systematic literature search up to 24 March 2017. *EFSA Journal*, 15(4):1–18, 2017.

[103] Naser Amanifar, Mohsen Taghavi, and Mohammad Salehi. Xylella fastidiosa from almond in Iran: overwinter recovery and effects of antibiotics. *Phytopathologia Mediterranea*, 55(3):337–345, 2016.

[104] Christina E. Maddox, Lisa M. Laur, and Li Tian. Antibacterial activity of phenolic compounds against the phytopathogen Xylella fastidiosa. *Current Microbiology*, 60(1):53–58, 2010.

[105] Thomas J. Aldrich, Philippe E. Rolshausen, M. Caroline Roper, Jordan M. Reader, Matthew J. Steinhaus, Jeannette Rapicavoli, David A. Vosburg, and Katherine N. Maloney. Radicinin from Cochliobolus sp. inhibits Xylella fastidiosa, the causal agent of Pierce's Disease of grapevine. *Phytochemistry*, 116:130–137, 2015.

[106] M. Morelli, G. D'Attoma, M. Saponari, and P. Saldarelli. Evaluation of an Endophytic Symbiont as a putative Biocontrol Agent of the CoDiRO strain of Xylella fastidiosa. *Journal of Plant Pathology*, 99, 2017.

[107] Mayukh Das, Tushar Suvra Bhowmick, Stephen J. Ahern, Ry Young, and Carlos F. Gonzalez. Control of Pierce's Disease by Phage. *PLoS ONE*, 10(6):1–15, 2015.

[108] Gerane Celly Dias Bezerra-Silva, Márcio Alves Silva, Marcelo Pedreira de Miranda, and João Roberto Spotti Lopes. Effect of Contact and Systemic Insecticides on

the Sharpshooter Bucephalogonia xanthophis (Hemiptera: Cicadellidae), a Vector of Xylella fastidiosa in Citrus. *Florida Entomologist*, 95(4):854–861, 2012.

[109] D. J. W. Morgan, S. V. Triapitsyn, R. A. Redak, L. G. Bezark, and M. S. Hoddle. Biological Control of the Glassy-Winged Sharpshooter: Current status and future potential. *California Conference on Biological Control II*, 2000.

[110] Michael J. Davis, William J. French, and Norman W. Schaad. Axenic culture of the bacteria associated with phony disease of peach and plum leaf scald. *Current Microbiology*, 6(5):309–314, 1981.

[111] M. J. Davis, A. H. Purcell, and S. V. Thomson. Isolation Media for the Pierce's Disease Bacterium. *Phytopathology*, 70(5):425–429, 1980.

[112] Alexander H. Purcell and Donald L. Hopkins. Fastidious Xylem-Limited Bacterial Plant Pathogens. *Annual Review of Phytopathology*, 34:131–151, 1996.

[113] John S. Hartung, Julia Beretta, Ronald H. Brlansky, Joan Spisso, and Richard F. Lee. Citrus Variegated Chlorosis Bacterium: Axenic Culture, Pathogenicity, and Serological Relationships with Other Strains of Xylella fastidiosa. *Phytopathology*, 84(6):591–597, 1994.

[114] João Carlos Campanharo, Manoel Victor Franco Lemos, and Eliana Gertrudes De Macedo Lemos. Growth optimization procedures for the phytopathogen Xylella fastidiosa. *Current Microbiology*, 46(2):99–102, 2003.

[115] Breno Leite, Peter Craig Andersen, and Maria Lucia Ishida. Colony aggregation and biofilm formation in xylem chemistry-based media for Xylella fastidiosa. *FEMS Microbiology Letters*, 230(2):283–290, 2004.

[116] S. J. Kostka, T. A. Tattar, J. L. Sherald, and S. S. Hurtt. Mulberry Leaf Scorch, New Disease Caused by a Fastidious, Xylem-Inhabiting Bacterium. *Plant Disease*, 70(7):690–693, 1986.

[117] S. M. Fry, R. D. Milholland, and Pi-Yu Huang. Isolation and Growth of Strains of Xylella fastidiosa from Infected Grapevines on Nutrient Agar Media. *Plant Disease*, 74(7):522–524, 1990.

[118] Maristela Boaceff Ciraulo, Daiene Souza Santos, Ana Claudia De Freitas Oliveira Rodrigues, Marcus Vinícius De Oliveira, Tiago Rodrigues, Regina Costa De Oliveira, and Luiz R. Nunes. Transcriptome analysis of the phytobacterium xylella fastidiosa growing under xylem-based chemical conditions. *Journal of Biomedicine and Biotechnology*, pages 1–18, 2010.

[119] Paulo A. Zaini, Leonardo De La Fuente, Harvey C. Hoch, and Thomas J. Burr. Grapevine xylem sap enhances biofilm development by Xylella fastidiosa. *FEMS Microbiology Letters*, 295:129–134, 2009.

[120] Alessandra A. de Souza, Marco A. Takita, Helvécio D. Coletta-Filho, Camila Caldana, Giane M. Yanai, Nair H. Muto, Regina C. de Oliveira, Luiz R. Nunes, and Marcos A. Machado. Gene expression profile of the plant pathogen Xylella fastidiosa during biofilm formation in vitro. *FEMS Microbiology Letters*, 237:341–353, 2004.

[121] Shane E. Cotter, Hye-Jeong Yeo, Twyla Juehne, and Joseph W. St. Geme III. Architecture and Adhesive Activity of the Haemophilus influenzae Hsf Adhesin. *Journal of Bacteriology*, 187(13):4656–4664, 2005.

[122] Jesús Munõz-Rojas, Luis E. Fuentes-Ramírez, and Jesús Caballero-Mellado. Antagonism among Gluconacetobacter diazotrophicus strains in culture media and in endophytic association. *FEMS Microbiology Ecology*, 54:57–66, 2005.

[123] J. D. Janse and A. Obradovic. Xylella fastidiosa: Its Biology, Diagnosis, Control and Risks. *Journal of Plant Pathology*, 92(1):35–48, 2010.

[124] M. Saponari, D. Boscia, F. Nigro, and G. P. Martelli. Identification of DNA Sequences Related to Xylella fastidiosa in Oleander , Almond and Olive Trees Exhibiting Leaf Scorch Symptoms in Apulia (southern Italy). *Journal of Plant Pathology*, 95(3), 2013.

[125] Annalisa Giampetruzzi, Maria Saponari, Rodrigo P. P. Almeida, Salwa Essakhi, Donato Boscia, Giuliana Loconsole, and Pasquale Saldarelli. Complete Genome Sequence of the Olive-Infecting Strain Xylella fastidiosa subsp. pauca De Donno. *Genome Announcements*, 5(27):5–6, 2017.

[126] Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, and Michiel J. L. de Hoon. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.

[127] Mao Tanabe and Minoru Kanehisa. Using the KEGG Database Resource. *Current Protocols in Bioinformatics*, 38:1.12.1–1.12.43, 2012.

[128] Jeffrey D. Orth, Tom M. Conrad, Jessica Na, Joshua A. Lerman, Hojung Nam, Adam M. Feist, and Bernhard Ø. Palsson. A comprehensive genome-scale reconstruction of Escherichia coli metabolism 2011. *Molecular Systems Biology*, 7(535):1–9, 2011.

[129] Sarah Schatschneider, Marcus Persicke, Steven Alexander Watt, Gerd Hublik, Alfred Pühler, Karsten Niehaus, and Frank Jörg Vorhölter. Establishment, in silico analysis,

and experimental verification of a large-scale metabolic network of the xanthan producing Xanthomonas campestris pv. campestris strain B100. *Journal of Biotechnology*, 167(2):123–134, 2013.

[130] Sophia Santos and Isabel Rocha. Estimation of biomass composition from genomic and transcriptomic information. *Journal of Integrative Bioinformatics*, 13(2):1–14, 2016.

[131] J. C. Dianese and N. W. Schaad. Isolation and Characterization of Inner and Outer Membranes of Xanthomonas campestris pv. campestris. *Phytopathology*, 72(10):1284–1289, 1982.

[132] Xiaoyuan Wang and Peter J. Quinn. Lipopolysaccharide: Biosynthetic pathway and structure modification. *Progress in Lipid Research*, 49:97–107, 2010.

[133] Chris Whitfield. Biosynthesis of lipopolysaccharide O antigens. *Trends in Microbiology*, 3(5):178–185, 1995.

[134] Jeannette N. Rapicavoli, Barbara Blanco-Ulate, Artur Muszynski, Rosa Figueroa-Balderas, Abraham Morales-Cruz, Parastoo Azadi, Justyna M. Dobruchowska, Claudia Castro, Dario Cantu, and M. Caroline Roper. Lipopolysaccharide O-antigen delays plant innate immune recognition of Xylella fastidiosa. *Nature Communications*, 9(390):1–12, 2018.

[135] Waldemar Vollmer and Joachim-Volker Holtje. The Architecture of the Murein (Peptidoglycan) in Gram-Negative Bacteria : Vertical Scaffold or Horizontal Layer (s)? *Journal of Bacteriology*, 186(18):5978–5987, 2004.

[136] Hongwu Ma and An-Ping Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2):270–277, 2003.

[137] Agda Paula Facincani, Jesus Aparecido Ferro, João Martins Pizauro Jr., Haroldo Alves Pereira Jr., Eliana Gertrudes de Macedo Lemos, Alessandro Luis do Prado, and Maria Inês T. Ferro. Carbohydrate metabolism of Xylella fastidiosa: Detection of glycolytic and pentose phosphate pathway enzymes and cloning and expression of the enolase gene. *Genetics and Molecular Biology*, 26(2):203–211, 2003.

[138] Sarah Schatschneider, Claudia Huber, Heiko Neuweger, Tony Francis Watt, Alfred Pühler, Wolfgang Eisenreich, Christoph Wittmann, Karsten Niehaus, and Frank-Jörg Vorhölter. Metabolic flux pattern of glucose utilization by Xanthomonas campestris pv . campestris : prevalent role of the Entner Doudoroff pathway and minor fluxes through the pentose phosphate pathway and glycolysis. *Molecular BioSystems*, 10:2663–2676, 2014.

[139] Marcel Frese, Sarah Schatschneider, Julia Voss, Frank-Jörg Vorhölter, and Karsten Niehaus. Characterization of the pyrophosphate-dependent 6-phosphofructokinase from Xanthomonas campestris pv. campestris. *Archives of Biochemistry and Biophysics*, 546:53–63, 2014.

[140] Fabien Letisse, Paule Chevallereau, Jean-Luc Simon, and Nic D. Lindley. Kinetic analysis of growth and xanthan gum production with Xanthomonas campestris on sucrose , using sequentially consumed nitrogen sources. *Applied Microbiology and Biotechnology*, 55:417–422, 2001.

[141] Eliana Gertrudes de Macedo Lemos, Lúcia Maria Carareto Alves, and João Carlos Campanharo. Genomics-based design of defined growth media for the plant pathogen Xylella fastidiosa. *FEMS Microbiology Letters*, 219:39–45, 2003.

[142] Rodrigo P. P. Almeida, Renee Mann, and Alexander H. Purcell. Xylella fastidiosa Cultivation on a Minimal Solid Defined Medium Xylella fastidiosa Cultivation on a Minimal Solid Defined Medium. *Current Microbiology*, 48:368–372, 2004.

[143] C. J. Chang and R. C. Donaldson. Xylella fastidiosa : Cultivation in Chemically Defined Medium. *Phytopathology*, 83(2):192–194, 1993.

[144] David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(November 2017):1074–1082, 2018.

[145] Joao Meidanis, Marilia D. V. Braga, and Sergio Verjovski-Almeida. Whole-Genome Analysis of Transporters in the Plant Pathogen Xylella fastidiosa. *Microbiology and Molecular Biology Reviews*, 66(2):272–299, 2002.

[146] Bernard Henrissat, Emeline Deleury, and Pedro M. Coutinho. Glycogen metabolism loss : a common marker of parasitic behaviour in bacteria? *TRENDS in Genetics*, 18(9):437–440, 2002.

[147] N. Killiny, R. Hernandez Martinez, C. Korsi Dumenyo, D. A. Cooksey, and R. P. P. Almeida. The Exopolysaccharide of Xylella fastidiosa Is Essential for Biofilm Formation , Plant Virulence , and Vector Transmission. *Molecular Plant-Microbe Interactions*, 26(9):1044–1053, 2013.

[148] Felipe Rodrigues da Silva, Andre Luiz Vettore, Edson Luis Kemper, Adilson Leite, and Paulo Arruda. Fastidian gum: the Xylella fastidiosa exopolysaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiology Letters*, 203(2):165–171, 2001.

[149] P. Pielken, K. L. Schimz, L. Eggeling, and H. Sahm. Glucose metabolism in Xanthomonas campestris and influence of methionine on the carbon flow. *Canadian Journal of Microbiology*, 34:1333–1337, 1988.

[150] Valérie De Crécy-Lagard, Basma El Yacoubi, Rocío Díaz de la Garza, Alexandre Noiriel, and Andrew D. Hanson. Comparative genomics of bacterial and plant folate synthesis and salvage : predictions and validations. *BMC Genomics*, 8(245):1–15, 2007.

[151] Lian Zhou, Jia-Yuan Wang, Jianhe Wang, Alan Poplawsky, Shuangjun Lin, Bangshang Zhu, Changqing Chang, Tielin Zhou, Lian-Hui Zhang, and Ya-Wen He. The diffusible factor synthase XanB2 is a bifunctional chorismatase that links the shikimate pathway to ubiquinone and xanthomonadins biosynthetic pathways. *Molecular Microbiology*, 87(1):80–93, 2013.

[152] Tye O. Boynton, Svetlana Gerdes, Sarah H. Craven, Ellen L. Neidle, John D. Phillips, and Harry A. Dailey. Discovery of a Gene Involved in a Third Bacterial Protoporphyrinogen Oxidase Activity through Comparative Genomic Analysis and Functional Complementation. *Applied and Environmental Microbiology*, 77(14):4795–4801, 2011.

[153] A. C. Zagallo and C. H. Wang. Comparative Glucose Catabolism of Xanthomonns Species. *Journal of Bacteriology*, 93(3):970–975, 1967.

[154] Julia Bailey-Serres and Ron Mittler. The Roles of Reactive Oxygen Species in Plant Cells. *Plant Physiology*, 141:311, 2006.

[155] Jose A. O'Brien, Arsalan Daudi, Vernon S. Butt, and G. Paul Bolwell. Reactive oxygen species and their role in plant defence and cell wall metabolism. *Planta*, 236:765–779, 2012.

[156] Max Chavarría, Pablo I. Nikel, Danilo Pérez-pantoja, Víctor De Lorenzo, San José, and Costa Rica. The Entner Doudoroff pathway empowers Pseudomonas putida KT2440 with a high tolerance to oxidative stress. *Environmental Microbiology*, 15(6):1772–1785, 2013.

[157] Helene Feil and Alexander H. Purcell. Temperature-Dependent Growth and Survival of Xylella fastidiosa in Vitro and in Potted Grapevines. *Plant Disease*, 85(12):1230–1234, 2001.

[158] M. Caroline Roper, L. Carl Greve, John M. Labavitch, and Bruce C. Kirkpatrick. Detection and Visualization of an Exopolysaccharide Produced by Xylella fastidiosa In Vitro and In Planta. *Applied and Environmental Microbiology*, 73(22):7252–7258, 2007.

[159] A. R. Poplawsky, S. C. Urban, and W. Chun. Biological Role of Xanthomonadin Pigments in Xanthomonas campestris pv. Campestris. *Applied and Environmental Microbiology*, 66(12):5123–5127, 2000.

[160] Xue-Qiang Cao, Jia-Yuan Wang, Lian Zhou, Bo Chen, Yu Jin, and Ya-Wen He. Biosynthesis of the yellow xanthomonadin pigments involves an ATP-dependent 3-hydroxybenzoic acid : acyl carrier protein ligase and an unusual type II polyketide synthase pathway. *Molecular Microbiology*, 110(1):16–32, 2018.

[161] James G. Ferry. Formate dehydrogenase. *FEMS Microbiology Reviews*, 87(1):377–382, 1990.

[162] Servane Blanvillain, Damien Meyer, Alice Boulanger, Martine Lautier, Catherine Guynet, Nicolas Denancé, Jacques Vasse, Emmanuelle Lauber, and Matthieu Arlat. Plant Carbohydrate Scavenging through TonB-Dependent Receptors: A Feature Shared by Phytopathogenic and Aquatic Bacteria. *PLoS ONE*, 2(2):1–21, 2007.

[163] Christopher H. Taron, Jack S. Benner, Linda J. Hornstra, and Ellen P. Guthrie. A novel β-galactosidase gene isolated from the bacterium Xanthomonas manihotis exhibits strong homology to several eukaryotic β-galactosidases. *Glycobiology*, 5(6):603–610, 1995.

[164] Joseph F. Frank and George A. Somkuti. General Properties of Beta-Galactosidase of Xanthomonas campestris. *Applied and Environmental Microbiology*, 38(3):554–556, 1979.

[165] Leonardo C. A. Souza, Nelson A. Wulff, Patrice Gaurivaud, Anelise G. Mariano, Andréa C.D. Virgílio, João L. Azevedo, and Patrícia B. Monteiro. Disruption of Xylella fastidiosa CVC gumB and gumF genes affects biofilm formation without a detectable influence on exopolysaccharide production. *FEMS Microbiology Letters*, 257(1):236–242, 2006.

[166] Ellen D. Beaulieu, Michael Ionescu, Subhadeep Chatterjee, Kenji Yokota, Dirk Trauner, and Steven Lindow. Characterization of a diffusible signaling factor from Xylella fastidiosa. *mBio*, 4(1):1–9, 2013.

[167] Britta Soballe and Robert K. Poole. Microbial ubiquinones: multiple roles in respiration, gene regulation and oxidative stress management. *Microbiology*, 145(1):1817–1830, 1999.

[168] Jeffery A. Mayfield, Carolyn A. Dehner, and Jennifer L. DuBoius. Recent advances in bacterial heme protein biochemistry. *Current Opinion in Chemical Biology*, 15(2):260–266, 2011.

[169] Nabil Killiny, Simone S. Prado, and Rodrigo P. P. Almeida. Chitin Utilization by the Insect-Transmitted Bacterium Xylella fastidiosa. *Applied and Environmental Microbiology*, 76(18):6134–6140, 2010.

[170] Fabien Labroussaa, Michael Ionescu, Adam R. Zeilinger, Steven E. Lindow, and Rodrigo P. P. Almeida. A chitinase is required for Xylella fastidiosa colonization of its insect and plant hosts. *Microbiology*, 163:502–509, 2017.

A

SUPPORT MATERIAL

A.1 GENERIC METABOLITES

Table 31.: Generic metabolites automatically removed from the metabolic network. Reactions associated with these compounds were also removed.

Metabolite Name	**KEGG ID**
Peptide	C00012
Protein	C00017
Acceptor	C00028
Reduced acceptor	C00030
DNA	C00039
Amino acid	C00045
RNA	C00046
Carboxylate	C00060
L-Amino acid	C00151
2-Oxo acid	C00161
NTP	C00201
Primary alcohol	C00226
rRNA	C00240
NDP	C00454
Protein tyrosine	C00585
Protein histidine	C00615
Deoxynucleoside triphosphate	C00677

Base	C00701
DNA adenine	C00821
DNA cytosine	C00856
Ribonucleoside	C00911
Sugar phosphate	C00934
Aromatic oxo acid	C00972
Aromatic amino acid	C01021
Protein tyrosine phosphate	C01167
5′-Phospho-DNA	C02128
Deoxynucleoside	C02269
Protein Dithiol	C02315
'Activated' tRNA	C02342
L-Leucyl-protein	C02429
Protein disulfide	C02582
Protein glutamine	C02583
[Protein]-L-cysteine	C02743
DNA 5-methylcytosine	C02967
Peptide-L-methionine	C03023
(5-L-Glutamyl)-peptide	C03193
Protein L-isoaspartate	C03306
5-L-Glutamyl amino acid	C03363
DNA 6-methylaminopurine	C03391
Peptidylproline (omega=0)	C03633
Ribonucleoside diphosphate	C03723
(5-L-Glutamyl)-L-amino acid	C03740
Peptidylproline (omega=180)	C03798
Protein S-methyl-L-cysteine	C03800
Ribonucleoside triphosphate	C03802
Substituted beta-amino acid	C03806

Peptide-L-methionine (S)-S-oxide	C03895
rRNA containing N2-methylguanine	C04153
tRNA containing a thionucleotide	C04161
DNA containing 6-O-methylguanine	C04250
Protein N(pi)-phospho-L-histidine	C04261
Protein N(tau)-phospho-L-histidine	C04262
Protein L-isoaspartate methyl ester	C04311
Membrane-derived-oligosaccharide D-glucose	C04566
Membrane-derived-oligosaccharide 6-(glycerophospho)-D-glucose	C04809
alpha-Amino acid	C05167
DNA containing guanine	C11475
Sugar	C11477
(S)-2-Hydroxyacid	C15565
Quinone	C15602
Hydroquinone	C15603
Peptide-L-methionine (R)-S-oxide	C15653
Protein N6-(lipoyl)lysine	C16237
Protein N6-(dihydrolipoyl)lysine	C16832
Adenine in rRNA	C20648
N6-Dimethyladenine in rRNA	C20796
Protein N5-methyl-L-glutamine	C20858
5′-Triphospho-[mRNA]	C20864
DNA 5-hydroxymethylcytosine	C21031

A.2 REACTION DISTRIBUTION

Table 32.: Reaction distribution among the different pathways included in the model.

Pathway	Number of Reactions
2-Oxocarboxylic acid metabolism	36
Acarbose and validamycin biosynthesis	2
Aflatoxin biosynthesis	1
Alanine, aspartate and glutamate metabolism	23
alpha-Linolenic acid metabolism	4
Amino sugar and nucleotide sugar metabolism	20
Aminoacyl-tRNA biosynthesis	20
Aminobenzoate degradation	2
Arachidonic acid metabolism	3
Arginine and proline metabolism	15
Arginine biosynthesis	15
Ascorbate and aldarate metabolism	7
Benzoate degradation	1
Benzoxazinoid biosynthesis	1
beta-Alanine metabolism	4
Betalain biosynthesis	12
Biomass Pathway	12
Biosynthesis of unsaturated fatty acids	1
Biosynthesis of vancomycin group antibiotics	1
Biotin metabolism	13
Butanoate metabolism	2
C5-Branched dibasic acid metabolism	3
Caprolactam degradation	4
Carbapenem biosynthesis	3
Carbon fixation in photosynthetic organisms	13

Carbon fixation pathways in prokaryotes	16
Chloroalkane and chloroalkene degradation	3
Chlorocyclohexane and chlorobenzene degradation	6
Citrate cycle (TCA cycle)	17
Cyanoamino acid metabolism	10
Cysteine and methionine metabolism	37
D-Alanine metabolism	2
D-Glutamine and D-glutamate metabolism	6
Degradation of aromatic compounds	2
Drains pathway	143
Drug metabolism - cytochrome P450	3
Drug metabolism - other enzymes	16
Exopolysaccharide Biosynthesis	5
Fatty acid biosynthesis	40
Fatty acid degradation	8
Fatty acid elongation	7
Fatty acid metabolism	44
Folate biosynthesis	27
Fructose and mannose metabolism	12
Galactose metabolism	4
Geraniol degradation	1
Glucosinolate biosynthesis	3
Glutathione metabolism	15
Glycerolipid metabolism	6
Glycerophospholipid metabolism	13
Glycine, serine and threonine metabolism	22
Glycolysis / Gluconeogenesis	17
Glyoxylate and dicarboxylate metabolism	12
Histidine metabolism	10

Indole alkaloid biosynthesis	4
Inositol phosphate metabolism	4
Isoquinoline alkaloid biosynthesis	1
Limonene and pinene degradation	4
Lipoic acid metabolism	1
Lipopolysaccharide biosynthesis	19
Lysine biosynthesis	14
Lysine degradation	2
Metabolism of xenobiotics by cytochrome P450	23
Methane metabolism	13
Monobactam biosynthesis	5
Neomycin, kanamycin and gentamicin biosynthesis	1
Nicotinate and nicotinamide metabolism	14
Nitrogen metabolism	6
Novobiocin biosynthesis	2
One carbon pool by folate	19
Oxidative Phosphorylation	4
Pantothenate and CoA biosynthesis	16
Pentose and glucuronate interconversions	5
Pentose phosphate pathway	15
Phenazine biosynthesis	9
Phenylalanine metabolism	4
Phenylalanine, tyrosine and tryptophan biosynthesis	22
Phenylpropanoid biosynthesis	1
Polyketide sugar unit biosynthesis	4
Porphyrin and chlorophyll metabolism	17
Propanoate metabolism	8
Purine metabolism	56
Pyrimidine metabolism	36

Pyruvate metabolism	14
Riboflavin metabolism	10
Selenocompound metabolism	9
Sphingolipid metabolism	1
Spontaneous	68
Starch and sucrose metabolism	10
Streptomycin biosynthesis	7
Sulfur metabolism	7
Taurine and hypotaurine metabolism	1
Terpenoid backbone biosynthesis	11
Thiamine metabolism	10
Transporters pathway	256
Tropane, piperidine and pyridine alkaloid biosynthesis	2
Tryptophan metabolism	3
Tyrosine metabolism	4
Ubiquinone and other terpenoid-quinone biosynthesis	18
Valine, leucine and isoleucine biosynthesis	18
Valine, leucine and isoleucine degradation	6
Vitamin B6 metabolism	9
Xanthomonadin Biosynthesis	1
Non-Associated to Pathway	97

A.3 CRITICAL GENES

Table 33.: Locus Tag of critical genes of the reconstructed model without similarities with *O. europaea*. Locus Tags represented in bold are exclusive to the biofilm predictions.

Critical Genes			
Biofilm		**Planktonic**	
B9J09_RS06665	B9J09_RS06185	B9J09_RS06665	B9J09_RS10585
B9J09_RS07735	B9J09_RS06205	B9J09_RS07735	B9J09_RS10630
B9J09_RS08985	B9J09_RS08675	B9J09_RS08985	B9J09_RS10990
B9J09_RS10610	B9J09_RS09170	B9J09_RS10610	B9J09_RS11045
B9J09_RS00100	B9J09_RS09480	B9J09_RS00100	B9J09_RS01155
B9J09_RS00145	B9J09_RS09830	B9J09_RS00145	B9J09_RS00200
B9J09_RS00430	B9J09_RS10190	B9J09_RS00430	B9J09_RS00940
B9J09_RS00480	B9J09_RS10245	B9J09_RS00480	B9J09_RS10605
B9J09_RS00870	B9J09_RS10585	B9J09_RS00870	B9J09_RS08000
B9J09_RS01165	B9J09_RS10630	B9J09_RS01165	B9J09_RS07150
B9J09_RS03175	B9J09_RS10990	B9J09_RS03245	B9J09_RS00885
B9J09_RS03180	B9J09_RS11045	B9J09_RS03250	B9J09_RS01020
B9J09_RS03185	B9J09_RS01155	B9J09_RS03315	B9J09_RS01930
B9J09_RS03190	B9J09_RS00200	B9J09_RS03845	B9J09_RS03580
B9J09_RS03195	B9J09_RS00940	B9J09_RS04015	B9J09_RS09020
B9J09_RS03200	B9J09_RS10605	B9J09_RS04800	B9J09_RS09690
B9J09_RS03215	B9J09_RS08000	B9J09_RS04890	B9J09_RS10620
B9J09_RS03220	B9J09_RS07150	B9J09_RS06005	
B9J09_RS03245	B9J09_RS00885	B9J09_RS06185	
B9J09_RS03250	B9J09_RS01020	B9J09_RS06205	
B9J09_RS03315	B9J09_RS01930	B9J09_RS08675	
B9J09_RS03845	**B9J09_RS03205**	B9J09_RS09170	
B9J09_RS04015	B9J09_RS03580	B9J09_RS09480	
B9J09_RS04800	B9J09_RS09020	B9J09_RS09830	
B9J09_RS04890	B9J09_RS09690	B9J09_RS10190	
B9J09_RS06005	B9J09_RS10620	B9J09_RS10245	
B9J09_RS02865		B9J09_RS02865	