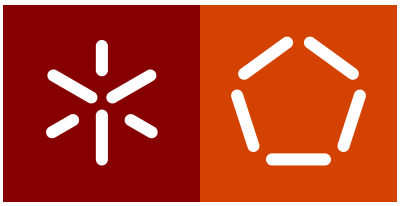
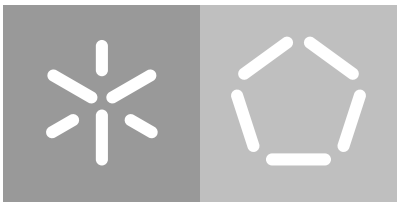


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Ana Isabel Gomes Fernandes

Unveiling the production of metabolites with flavor enhancement through genome--scale metabolic modelling of a Lambic beer microbial community

July 2021



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Ana Isabel Gomes Fernandes

Unveiling the production of metabolites with flavor enhancement through genome--scale metabolic modelling of a Lambic beer microbial community

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Oscar Manuel Lima Dias

July 2021

ACKNOWLEDGMENTS

Este ano foi sem dúvida dos anos mais desafiantes para fim. Ao longo deste trabalho ultrapassei vários desafios que fez com que eu evoluísse tanto a nível pessoal como a nível profissional, pois todos os meus limites foram testados, adquirindo novas competências, conhecimentos e experiência. Desde já agradeço a todas as pessoas que fizeram parte deste meu percurso académico, porque sem vocês não seria possível!

Primeiramente, gostaria de agradecer ao meu orientador Professor Doutor Óscar Dias pela confiança e por todo o conhecimento transmitido na área de Biologia de Sistemas ao longo deste difícil percurso. Foi extremamente gratificante ultrapassar todos os desafios propostos.

À Sophia e ao Emanuel por estarem sempre disponíveis para tirarem qualquer dúvida e me ajudarem em tudo. Agradeço por toda a paciência que tiveram comigo!

Aos meus colegas de mestrado, em especial aos meus companheiros da sala 0.09, obrigada por todas as conversas nas horas de descanso bem como a ajuda e força transmitida nas horas de trabalho.

Ao Francês, ao Emanuel, ao Miguel e ao Gava que sempre me receberam tao bem na nossa Covilhã. Quero agradecer todas as palavras de conforto, todos os jantares em família e principalmente pela amizade que a distância nunca quebrou. Uma vez Ubiana para sempre Ubiana! Gosto muito de vocês!

Ao Bando dos Quatro e às minhas Anas que estão comigo já há uns bons anos. Obrigada por todas as boas memórias que temos e que teremos, por estarem sempre do meu lado, tanto nos maus como nos bons momentos! Gosto tudo de vocês!

A ti, Natanael, não há palavras para descrever o que foste e és para mim. Obrigada por nunca teres soltado a minha mão, por teres sempre acreditado em mim, por me fazeres sempre sorrir e principalmente por todo o amor demonstrado. Enches-me de orgulho a cada conquista, inspiras-me por seres um exemplo de perseverança e resiliência, que sempre me motiva a nunca desistir dos meus sonhos. São anos de partilha que se tornam vida!

À minha família que está sempre comigo, que contribui desde sempre para a minha formação pessoal e profissional. À minha irmã por todo a paciência, força e sobretudo carinho. És o meu primeiro amor, fizeste-me crescer e mostraste-me o que é amar mais que a nós mesmos. Serás para sempre a minha bebé de olhos azuis e flor na bochecha. Aos meus pais, que sem eles nada seria possível. Sou eternamente grata por cada palavra de conforto, de força e principalmente por cada ensinamento. Por cada gesto de carinho e por todos os dias acreditarem em mim. A vossa força, o vosso respeito e claro, o vosso amor são um exemplo de vida. O meu sonho é um dia ser um terço daquilo que vocês são! Com o maior orgulho do Mundo, serei para sempre a vossa menina!

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

Systems biology studies biological processes on a global scale, involving different omics. It uses bioinformatic approaches, such as the reconstruction of genome-scale metabolic models to understand the biological system of a cell, organism, or microbial community. Genome-scale metabolic models are metabolic models based on the well-known stoichiometry of biochemical reactions. It offers a whole system view, predicting the metabolic phenotype, based on the genome and biochemical information. These models haves several applications in different areas such as biotechnological and pharmaceutical.

Lambic beers are commercial beers from Belgium that still use old brewing styles. This type of beer is gaining interest worldwide due to its unique flavour profile obtained by a mixed yeast-bacteria culture fermentation. Therefore, in this thesis, the lactic acid bacterium *Pediococcus damnosus* and *Brettanomyces bruxellensis* yeast, which play an important role in the acidification and maturation phase of the lambic beer fermentation, will be studied. *Pediococcus damnosus* is a gram-positive bacterium belonging to the lactic acid bacteria group commonly found in brewery environments. *Pediococcus damnosus* produces only lactate by the sugar metabolism, which confers an acidic and tart flavour to the beer. In turn, *Brettanomyces bruxellensis* is a facultatively anaerobic yeast, also responsible for the typical aroma of lambic beer. It uses several carbon sources and produces several volatile phenolic compounds not desired in common fermentations crucial in this type of beer.

Usually, in nature, the microorganisms appear in communities. Thus, the study of microbial communities is essential to understand their development, interaction and evolution. The main aim of this thesis is to unveil the production of metabolites with flavour enhancement in the acid lambic beer through the reconstruction and simulation of genome-scale metabolic models for each microorganism and therefore for the microbial community composed by them, to understand the interactions between the species and how these affects the lambic beer flavour.

Two genome-scale metabolic models were reconstructed: the model of the bacterium *Pediococcus damnosus* and the model of the yeast *Brettanomyces bruxellensis*. The tool used for model reconstruction was *merlin*, which automates several reconstruction processes and having a user-friendly interface. The *Pediococcus damnosus* genome-scale metabolic model consists of 809 reactions and 589 metabolites. In turn, the *Brettanomyces bruxellensis* genome-scale metabolic model has 2095 reactions and 1249 metabolites. In the simulations performances, the genome-scale metabolic models showed the ability to grow in the minimal medium provided, as described in the literature. Furthermore, simulations predicted the production of certain compounds, such as butanediol in the bacterium *Pediococcus damnosus* and 4-ethylphenol in the yeast *B. bruxellensis*, which may influence the Lambic beer flavour. Interactions between the genome-scale metabolic models, especially amino acid exchanges, were predicted. The model of

the *Pediococcus damnosus-Brettanomyces bruxellensis* community was assembled using ReFramed. The community model's simulation results show that the interaction of these microorganisms results in the production of compounds that may flavour and thus be responsible for the unique flavour profile of Lambic beer.

Keywords: Genome-Scale Metabolic Model, Acid lambic beer, Microbial communities, *Pediococcus damnosus*, *Brettanomyces bruxellensis*

RESUMO

A biologia de sistemas estuda os processos biológicos numa escala global, envolvendo diferentes ómicas. Utiliza abordagens bioinformáticas, como a construção de modelos metabólicos à escala genómica, de modo a perceber o sistema biológico de uma célula, organismo ou comunidade microbiana. Os modelos metabólicos à escala genómica são baseados na estequiometria bem conhecida das reações bioquímicas de um dado organismo. Oferece uma perspetiva do sistema como um todo, sendo capaz de prever o fenótipo do metabolismo, baseando-se no genoma e em informações bioquímicas. Estes modelos tem várias aplicações em diferentes áreas como a indústria biotecnológcia e farmacêutica.

As cervejas *lambic* são cervejas comerciais típicas da Bélgica que ainda utilizam processos de produção de cerveja antigos. Esta cerveja tem vindo a ganhar interesse a nível mundial devido ao seu perfil aromático único que é obtido através da fermentação de uma cultura de bactérias e leveduras. Nesta tese serão estudadas a bactéria ácida láctica *Pediococcus damnosus* e a levedura *Brettanomyces bruxellensis*, que possuem um papel importante nas fases de acidificação e maturação da fermentação desta cerveja.

Pediococcus damnosus é uma bactéria gram-positiva que pretence ao grupo das bactérias ácidas lácticas e é geralmente encontrada em ambientes de fermentação de cerveja. A bactéria *Pediococcus damnosus* produz apenas lactato pelo metabolismo dos açucares, conferindo um sabor ácido e azedo à cerveja. Por sua vez, a levedura *Brettanomyces bruxellensis* é uma levedura anaeróbica facultativa, também responsável pelo aroma típico da cerveja *lambic*. Utiliza inúmeras fontes de carbono e produz muitos compostos fenólicos voláteis que não são desejados em fermentações comuns, mas são cruciais neste tipo de cerveja. Geralmente, os microrganismos aparecem na natureza, em comunidades. O estudo de comunidades microbianas é importante para perceber o seu desenvolvimento, interação e evolução. O objetivo desta dissertação de mestrado é encontrar metabolitos que conferem o aroma característico da cerveja ácida *lambic*, de modo a melhorar a sua produção, usando para isso construção e simulação de modelos metabólicos à escala genómica para cada microrganismo e para a comunidade microbiana, de forma a perceber as interações entre espécies e como estas influenciam o aroma da cerveja *lambic*.

Assim foram construidos dois modelos metabólicos à escala genomA ferramenta utilizada para a construção destes modelos metabólicos à escala genómica foi o *merlin*, uma vez que automatiza varios processo de construçao e possui um interface intuitiva. O modelo metabólico à escala genómica da bactéria *Pediococcus damnosus* é constituído por 809 reações e 589 metabolitos. Por sua vez, o modelo metabólico à escala genómica da levedura *Brettanomyces bruxellensis* possui 2095 reações e 1249 metabolitos. Nas simulações executadas, os modelos metabólicos à escala genómica mostraram capacidade de crecer no meio minimo fornecido, como descrito

na literatura. Além disso, as simulações preveram a produção de certos compostos, como o butanodiol na bactéria *Pediococcus damnosus* e o 4-etilfenol na levedura *B. bruxellensis*, que podem influenciar o sabor da cerveja Lambic. Foram previstas interações entre os modelos metabólicos à escala genómica, sobretudo trocas de aminoácidos. O modelo da comunidade *Pediococcus damnosus-Brettanomyces bruxellensis* foi construido usando o ReFramed. Analisando os resultados da simulação do modelo da comunidade, pode-se concluir que a interação dos dois microorganismos resulta na produção de compostos que tem a capcidade de conferir sabor e assim serem responsáveis pelo aroma tão unico da cerveja *lambic*.

Palavras-Chave: Modelo Metabólico à Escala Genómica, Cerveja ácida *lambic*, Comunidades microbiana, *Pediococcus damnosus*, *Brettanomyces bruxellensis*

CONTENTS

ACRONYMS

A

AAB Acetic acid bacteria.

ATP Adenosine triphosphate.

B

BLAST Basic local alignment search tool.

BRENDA BRaunschweig ENzyme DAtabase.

C

CDM Chemically Defined Medium.

CDS Coding sequence.

COBRA COnstraint-Based Reconstruction and Analysis.

D

DNA Deoxyribonucleic acid.

E

E.C. Enzyme Commission.

EPS Exopolysaccharide.

F

FBA Flux Balance Analysis.

FVA Flux Variability Analysis.

G

GC Guanine-cytosine.

GOLD Genomes OnLine Database.

GPR Gene-Protein-Reaction.

GSMM Genome-scale metabolic model.

I

IBM Individual-based model.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LAB Lactic acid bacteria.

LP Linear Programming.

LTA Lipoteichoic acid.

M

MFA Metabolic Flux Analysis.

MILP Mixed-Integer Linear Programming.

MIRIAM Minimum information requested in the annotation of biochemical models.

MOMA Minimization of metabolic adjustment.

MRS Man-Rogosa-Sharpe.

N

NCBI National Center for Biotechnology Information.

P

PG Peptidoglycan.

PLM Population-level model.

R

RAVEN Reconstruction, Analysis and Visualization of Metabolic Networks.

RNA Ribonucleic Acid.

ROOM Regulatory on–off minimization.

S

SBML Systems Biology Markup Language.

SMETANA Species Metabolic Interaction Analysis.

T

T.C. Transporter Classification.

TA Teichoic acid.

TCDB Transporter Classification Database.

V

VRBG Violet red bile glucose.

W

WTA Wall teichoic acid.

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 Context and Motivation

Over the years, new technological approaches have been developed in order to replace traditional methods, such as experimental trials. Systems biology contributes to the development of *in silico* approaches [1], using high performance technologies that bring together computational techniques and biological big data [2]. One of the systems biology's tools of interest is the reconstruction, simulation, and optimization of Genome-scale metabolic models. Studies in all omics fields have increased significantly with new next-generation sequencing techniques, contributing to the GSMMs' development [3].

GSMMs are based on the well-known stoichiometry of biochemical reactions and they are used for the analysis of metabolism, simulating *in silico*, the phenotypic behavior of a microorganism in a range of environmental and genetic conditions in order to achieve a certain goal, such as the identification of potential target sites or over/under-production of compounds with biotechnological interest [3, 4]. GSMMs can be developed using bioinformatic tools, such as Metabolic Models Reconstruction Using Genome-Scale Information (*merlin*) [5] and CarveMe [6], which automate the process [5]. This process involves five main steps: the genome annotation, the assembly of the reactions network, the conversion of the network to a mathematical model, and the validation of the model with biological data from previously published or specifically designed experiments, following by a fifth stage for a prospective use [3, 7]. Usually, GSMMs are reconstructed for single organisms, however the GSMMs reconstruction of microbial communities has been promoted, due to the increase of interest on microbial community composition and interactions [8]. An example is the study of the community composition involved in the acid lambic beer fermentation.

Acidic lambic beers are among the oldest brewing types of beers still brewed. These beers are gaining interest worldwide, not only due to their unique flavour profile (acid and fruity) but also because its flavour is not due to pure cultures but relies on a dynamic mixed yeast-bacteria microbial community. The increasing popularity of lambic beer has led to greater interest in studying its fermentation in order to find new ways to improve the flavour profile of the beer. For this, we suggest the study of microorganisms involved in the fermentation of this type of beer, such as their metabolism, the compounds produced and their interactions. *Pediococcus damnosus* (*P. damnosus*) and *Brettanomyces bruxellensis* (*B. bruxellensis*) are two of the

microorganisms which dominate the lambic beer fermentation processes after six months [9, 10].

P. damnosus is a lactic acid bacterium belonging to the order *Lactobacillales*, commonly found in beer spoilage. It is a gram-positive bacterium, strictly fermentative, but facultatively anaerobic. This bacterium is homofermentative, producing only lactate which confers an acidic and tart flavour to the beer. *P. damnosus*, one of the microorganisms responsible for the acidification of lambic beers, is present in the acidification and maturation phase of lambic beer production process [10]. In the maturation stage is also present *B. bruxellensis*, a yeast which belongs to the *Pichiaceae* family. *Dekkera* is the sexual (teleomorphic) form of the asexual (anamorphic) genus *Brettanomyces*. This yeast is facultatively anaerobic and it can be distinguished from common yeast species due to its cycloheximide resistance [10]. *B. bruxellensis* produce ethanol by alcoholic fermentation and under aerobic conditions, being classified as Crabtree-positive yeast [11]. *B. bruxellensis* plays an important role during the maturation stage of lambic beer production because it is essential for the flavour composition [10].

The increase of the interest in bioinformatics and the development of tools led us to use it in the study of *P. damnosus* and *B. bruxellensis*, instead of the use of traditional and laborious wet lab methods. Thus, the reconstruction for GSMMs of each one of these organisms followed by the reconstruction of the community model, will provide further knowledge not only about both organisms metabolism but also the possible interactions between them, in order to achieve the intended objective.

1.2 Objective

The main aim of this work is to build GSMMs for the bacterium *P. damnosus*, for the yeast *B. bruxellensis*, and a community model consisting of both organisms. From this work, we expect to add knowledge to the microorganisms, metabolism and community interactions with a final goal of finding metabolites and/or proteins responsible for the lambic beer flavour profile. In order to achieve this objective, the following goals need to be fulfilled:

- Obtain a high-quality functional annotation of each organism's genome;
- Generate a draft metabolic network for each microorganism;
- Perform refinement and manual curation of the network using information gathered from literature;
- Convert the metabolic networks to stoichiometric models;
- Validate the reconstructed models using experimental data;
- Generate the metabolic model for the microbial community composed by the two organisms;
- Analyze the organisms' and community metabolic networks to find metabolites and/or proteins responsible for lambic beer flavour profile.

1.3 Structure of the document

This document is structured as follows:

2 - STATE OF THE ART

- Description of GSMM reconstruction of single organisms and communities;
- Computational tools used to reconstruct and simulate models;
- Overview of *P. damnosus* and *B. bruxellensis* metabolism;
- Importance of the study of microbial communities.

3 - METHODOLOGY

- Description of tools and methods used in the work;
- Description of genome annotation;
- Methods used to assemble the initial draft GSMM of each organisms;
- Methodology used to determinate biomass and energy requirements.
- Methods to manually curate the model;
- Enumeration of strategies used to validate the model;
- Methods for community model assembling;
- Enumeration of strategies used in community model simulation.

4 - RESULTS AND DISCUSSION

- Result of genome annotation;
- Changes in the draft GSMM through manual curation;
- Formulation of biomass and energy requirements;
- Model validation by comparing *in silico* simulations with experimental data;
- Community model assembling details;
- Results of community model simulations.

5 - CONCLUSION

- Summary of the developed models;
- Possible applications of community model;
- Future perspectives.

2

STATE OF THE ART

The following chapter provides a state of the art regarding the analyses of a microbial community present in the lambic beer fermentation. Therefore, the process of reconstruction of GSMMs, as well as the computational tools used are described. In addition, a description about acid lambic beer and the studied microorganisms is provided .

2.1 Background

Systems biology has gained interest over the years and has become an important way to produce high quality scientific studies, being one of the disciplines of biological sciences [12]. The approaches developed in this area, specifically the reconstruction of GSMMs, are a great method to study the mixed yeast-bacteria microbial community involved in the spontaneous fermentation of acid lambic beer [8, 10]. Using bioinformatic tools, it will be possible to unveil the production of metabolites with flavor enhancement [3], by studying the lactic acid bacterium *Pediococcus damnosus* (*P. damnosus*) and the yeast *Brettanomyces bruxellensis* (*B. bruxellensis*) found in lambic beer fermentation [10].

2.2 Systems biology

Bioinformatics is a discipline that combines different areas such as mathematics and biology, responsible for the analysis and interpretation of biological data [13]. Therefore, bioinformatics can provide valuable information for systems biology [14].

Systems biology studies biological processes at cellular, organism and more recently at a community level on a global scale, in terms of their molecular constituents and functional interactions [12], involving genomic, transcriptomic, proteomic, and metabolomic researches [14]. As an inclusive discipline, systems biology attempts to organize, evaluate and interpret data sources *in silico* [15], gradually decoding detailed information of microorganisms. Moreover this field of knowledge facilitates the metabolic engineering process, deciphering mechanisms for the development of strains with improved phenotypes in lines derived from mutations or adaptive laboratory evolution [1]. In metabolic engineering, organisms are designed with improved capacities concerning the productivity of desired compounds [16]. The most effective

and used tool for the study of the metabolism is the *in silico* reconstruction of GSMMs, being nowadays an essential tool for this propose [7, 17].

2.3 Genome-scale metabolic models

Since Gregor Mendel discovered the laws of genetics, in which information is transmitted between generations and determines the form and function of an organism, the genotype-phenotype relationship has been essential in life sciences research [18]. Genomics studies have increased significantly with the emergence of new next-generation sequencing techniques [3]. A sequenced and annotated genome of a microorganism has many potential applications such as the reconstruction of GSMMs [4] which uses also biochemical information [4, 5], being one of the most interesting systems biology's tools to study the link between metabolism and phenotype [3, 19]. Therefore, GSMM can be simulated in order to achieve a desired goal, that may have applications in several areas such as biotechnological, medicine [20, 21], pharmaceutical, chemical, and environmental industries, in which interest has been increasing [21].

GSMMs are comprehensive metabolic models based on the well-known stoichiometry of biochemical reactions, which includes all known chemical reactions and their corresponding associated genes of a cell, organism, or microbial community [4, 22]. These models offer an ideal view of the whole system making them a perfect tool to predict phenotypes under different environmental and genetic conditions [4, 23, 24]. These constraint-based stoichiometric models can also incorporate different data types such as transcriptomic, proteomic and metabolomic information [4, 15]. Gene-Protein-Reaction (GPR) rules describe the relationship between genes that encode enzymes catalysing reactions [22]. These GPR associations are very important in GSMMs since they allow the phenotypic prediction consequences of a genetic modifications [22] such as finding potential drug targets candidates and assigning functions to unknown genes [4].

The firsts GSMMs published were used to understand the characteristics of microbial pathogens at a genome-scale, which was followed by the use of GSMMs to develop metabolic engineering strategies of microbial hosts for improved production of various bioproducts (for example, products with biotechnological interest, such as L-valine production by *Escherichia coli* applicable in cosmetic, pharmaceutical and animal feed industries) [21, 25]. However, the potential of GSMMs is still being unveiled and have been also used in other aspects, such as to contextualize the high-throughput data, to understand the complex biological phenomena, to orientate the hypothesis-driven discovery, to comprehend the microbial community relationships, and to discover network properties [24].

The first GSMM to be published was the GSMM of *Haemophilus influenza* in 1999 [4, 26]. Since then, great progress has been made in the reconstruction of GSMMs, and so efficiency and quality have been significantly improved [21, 24, 27, 28]. The number of GSMMs is greatly increasing, but more important than the quantity is the quality of these models. To

ensure the quality of a metabolic model, validation must be made by evaluating its ability to correctly predict the physiological characteristics of the given organism, comparing the predicted physiological characteristics of the metabolic model with experimental data [21, 24]. Through this comparison, the accuracy and prediction capabilities of the metabolic model can be improved. The whole process of reconstructing a GSMM is laborious and time consuming, and has been extensively described, stimulating the reconstruction of these models [7, 23, 24, 29]. The GSMMs reconstruction is still less expensive and time-consuming than wet lab experiments, where these models are already a common assistance [5]. Due to the increased interest in GSMMs, there is a community effort to reduce the time of the reconstruction process, by automating some steps and at the same time improve their quality [7, 23, 24, 29].

Therefore, the reconstruction of GSMMs continues a challenge for organisms that have a lack of biological data and information reported in the literature, as well as with poorly studied and uncharacterized metabolism [3], since this reconstruction process is only possible due to the availability of annotated and sequenced genomes [15]. Hence, the poor availability of literature and biological data remains an obstacle to the reconstruction of this type of models. Well-studied organisms with accessible biological and experimental data, such as *Escherichia coli* [30–33], already have several models available, therefore the reconstruction process is much easier [3]. Algorithms have been developed to examine metabolic models in various ways, for example by calculating the redistribution of metabolic flux in response to genetic or environmental disturbances [21]. Thus, to predict flux distributions in metabolic networks for product formation or optimal growth, for example, Flux Balance Analysis (FBA) can be performed depending on nutritional conditions or as response to enzyme/gene knockouts [20].

GSMMs have been successfully applied to the study of single microbes, however, in recent years has been developed extensive tools that analysis models from single strains to complex microbial communities [34].

2.4 Reconstruction of Genome-scale metabolic models

The reconstruction of GSMM has been studied and described over the years by several authors [5], representing two reconstruction approaches: the top-down and the bottom-up. The latter is the most common reconstruction approach [35]. The top-down approach consists in generating a universal model which is then manually curated. This model is ready to simulate, and it includes a universal biomass equation, transport reactions and do not contain blocked or unbalanced reactions. Then it is used a *carving* process where the universal model is used as a template to generate GSMMs of specific organism. This process preserves all the manual curation and the major structural properties of the original model, identifying the reactions and metabolites predicted to be present in the given organism through homology tools and orthology predictions. The top-down approach is a good option for organisms that do not grow in well-defined media or their defined media is not known, because it is able to deduce their

uptake and secretion capabilities from genetic information [35]. CarveMe is a tool example which uses the top-down approach [35, 36].

The bottom-up approach is the traditional method where given a genome of a specific organism and its functional annotation, based on homology tools and orthology predictions, a draft model can be automatically reconstructed. This draft model is then extensively manually curated. Bottom-up approach is time-consuming, including repetitive tasks. For every new reconstruction model, the reconstruction steps must be performed [35]. The most comprehensive bottom-up study was published by Thiele and B. Ø. Palsson in 2010 and describes a protocol to reconstruct a high-quality genome-scale metabolic model with about one hundred steps which are grouped in five mains stages: Genome annotation, Manual reconstruction refinement, Conversion to mathematical model and Network evaluation, followed by a fifth stage for prospective use. This reconstruction is an iterative process which comprises several steps, as represented in Figure 1 (adapted from I. Rocha et al. 2008) which can be summarised as follows: first, the genome annotation is performed using data from different data sources. Second, the metabolic network is refined (also using different data sources), by performing manual curation, adding the biomass production equation, filling the gaps and adding other constraints, such as reactions reversibility. In the third step, the reconstructed GSMM is converted into a stoichiometric model (also named mathematical model). Finally, to validate the model, the predictions are compared to experimental data. If the simulations do not agree with the experimental data, a debugging process must be performed, repeating the metabolic network refinement and conversion stage [4, 5, 7]. The final model should be exported in a standard format [5], accepted by the most tools, for example, the Systems Biology Markup Language (SBML) [37], to use the GSMM prospectively. Metabolic Models Reconstruction Using Genome-Scale Information (*merlin*) is a tool which uses bottom-up approaches for GSMM reconstruction.

2.4.1 Genome annotation

In the first stage, a draft reconstruction is generated based on the organism's genome annotation and in databases with biochemical information [7] such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [38] and BRaunschweig ENzyme DAtabase (BRENDA) [39]. Thus, the first step is the genome annotation whereby is collected genomic information from different databases[7], such as National Center for Biotechnology Information Entrez Gene (NCBI Entrez gene) [40] and Genomes OnLine Database (GOLD) [41]. E.C. numbers [42], T.C. numbers [43] and if available, associated genes and gene product names information may also be collected [7]. Overall, metabolic functions encoded by the genome are collected using tools which perform Basic local alignment search tool (BLAST) [44], such as Pathway tool [45, 46], and metabolic SearcH And Reconstruction Kit (metaSHARK) [47]. Nevertheless, a manual curation is indispensable because genome annotation may include false information due to missing, wrong, or incomplete annotations [7].The draft of the reconstruction, as well

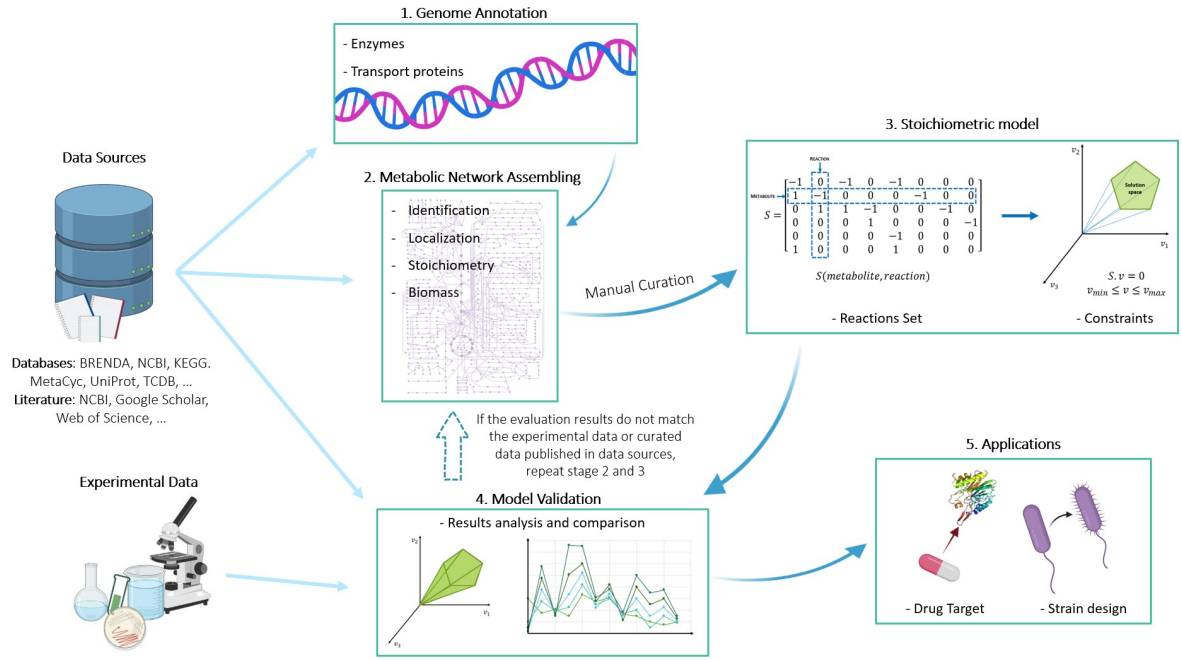


Figure 1: GSMM reconstruction iterative process.

Adapted from I. Rocha et al. 2008. Created with BioRender.com

as the cured reconstruction, depends essentially on a reliable annotation of the genome [7]. The genome annotation information was performed while the genome was sequenced, which may not be updated, therefore it is important to download the most recent version available [7]. Data related to gene expression regulation or genes associated to signal transduction are not regarded for these metabolic models [7]. Typically, a re-annotation is necessary because the collected information does not always comply with the GSMM requirements. Usually, genes identified as metabolic have missing E.C. number [5, 7]. Thus, in the second step, named candidate metabolic function, the metabolic genes can be automatically retrieved from the genome annotation, using for example the gene ontology [7]. The identified genetic products catalyse metabolic reactions that can be linked to the draft reconstruction through E.C. numbers and biochemical reaction databases[7] such as BRENDA [39] and KEGG [38]. A list of candidates is obtained which may not be complete or comprehensive, as many false positives may be present, such as proteins involved in Deoxyribonucleic acid (DNA) methylation or ribosomal Ribonucleic acid (rRNA) modification that have associated E.C. numbers but are not normally considered in metabolic models. Therefore, manual curation will be necessary[7]. There are several databases and online resources with available data which can be very helpful in manual curation, as is described in Table 1 (Adapted from I. Rocha et al. 2008 and Thiele and B. Ø. Palsson 2010).

Table 1: Online data sources used in GSMMs reconstruction*1

Database	Description	Reference
GOLD	GOLD is an open online resource, constantly updated, which contains genome and metagenome projects and their associated metadata	[41]
NCBI	NCBI is an online repository comprising several databases. Biomedical, biochemical and genomic information can be retrieved from NCBI website.	[40]
NCBI Entrez gene	NCBI Entrez gene is a database from NCBI for gene-specific information	[48]
KEGG	KEGG is an integrated database of genes and genomes. It has information about systems, genes, health and chemical compounds	[38]
BRENDA	BRENDA is the main database with enzyme and enzyme-ligand information. This database has manually annotated information and use E.C. classification system.	[39]
TCDB	TCDB is a database for transport protein research with freely accessible information about structure, function, mechanism, evolution and disease/medicine about all organism types transporters. In this database each transporter protein is identified with a T.C. number.	[49]
MetaCyc	MetaCyc is a database of metabolic pathways and enzymes, containing about 2749 pathways. Thus, it is the largest cured metabolic pathways database.	[50]
BiGG Models	BiGG Models is a repository of high-quality GSMMs. This database includes more then one hundred of GSMMs, which are identified by a BiGG IDs.	[51]

Continued on next page

Table 1 – continued from previous page

Database	Description	Reference
BioCyc	BioCyc is a microbial genome and metabolic pathway collection, with information inferred by software tools, imported from outher databases and curated from literature. The web portal provides tools for visualization and analysis of the metabolic pathways	[52]
ExPASy [*2]	ExPASy was one of the first life science web servers and now it is one of the main bioinfroamtics resources fro proteomics information. This database has evolved, comprising a wide range of resources in many different fields beyond proteomics, such as genomics, phylogeny/evolution, systems biology, population genetics and transcriptomics.	[53]
UniProt [*3]	UniProt is a database of sequences and annotations of millions proteins. UniProt Knowledgebase (UniProtKB) is a component of UniProt database, which comprise other two databases: UniProtKB/TrEMBL (Translated EMBL Nucleotide Sequence Data Library) (TrEMBL) and UniProtKB/Swiss-Prot (Swiss-Prot). TrEMBL has entries automatically annotated, while the Swiss-Prot entries are curated by a biocuration team.	[54]

[*1] (Adapted from I. Rocha et al. 2008 and Thiele and B. Ø. Palsson 2010)
[*2]- Expert Protein Analysis System (ExPASy)
[*3]- Universal Protein Resource (UniProt)

2.4.2 Manual reconstruction refinement

Manual curation defines the second stage of the GSMM reconstruction whereby the entire draft will be re-evaluated and refined. This process is important since not all annotations have a high confidence value (low e-value) and most biochemical databases do not have information specific to certain organisms. Adding organism-unspecific reactions, the predictive behaviour of the model can be compromised. Moreover, at this stage, information about

biomass composition, maintenance parameters, and growth conditions is collected in order to provide background for simulations in the coming stages [4, 7].

Genes, Proteins and Reactions

A very important step is the GPR association. Information about GPR association usually is provided from genome annotation through the gene function. Although verification and refinement is necessary in order to determine if the functional protein is a heteromeric enzyme complex, if the enzyme or the enzyme complex catalyzes more than one reaction and if more than one protein have the same function. Databases information, such as KEGG data, and literature search are used to collect this type of data. Afterward, spontaneous reactions, which have at least one metabolite connecting them to the rest of the draft reconstruction, must be included, in order to avoid dead-end metabolites (metabolites that are only reactants or products) [4, 7].

Spontaneous Reactions

In this step, the spontaneous reactions, reactions which favours the products formation under certain conditions, and nonenzymatic reactions, that occur spontaneously or catalysed by small molecules, are included in the network metabolic draft. Information about these reactions can be found in the literature and in databases such as KEGG [4, 7, 38]. Some tools such as *merlin* [5, 55] and MicrobesFlux [56] can automatically add those to the draft model[4, 7].

Demand and sink reactions

Demand reactions provide compound accumulation which is not permitted in steady-state models because of mass-balancing requirement, where for each metabolite, the influx sum is equal to the sum of efflux. At this point, these reactions should be added for compounds known to be produced by the target organism, for example, certain cofactors. By adding these reactions for a certain metabolite, the blocked reactions, in which no flux passes, can be converted into active reactions, in which flux passes. In general, most GSMMs drafts have few demand reactions and can be temporarily added in the validation stage during the evaluation process to test or verify certain metabolic functions. In the end, they will be removed. Sink reactions are similar to demand reactions and reversible, supplying the network with metabolites. These reactions are important for compounds that are not metabolized by metabolic cellular processes, but their production is necessary. The addition of many sink reactions can allow the model to grow without any need from medium resources and therefore these reactions should be added with care. They are also used in the debugging process in the validation stage, helping to identify the origin of a problem. When a gap is identified, it can be functionally replaced by a sink reaction [7].

Stoichiometry and Reaction directionality

The formula of each metabolite is fixed, as metabolites are generally listed with their uncharged formula, but in the medium and cells, most are protonated or deprotonated. Thus, the stoichiometry of the draft reconstruction reactions must be determined. Should be considered that unbalanced reactions can synthesize protons or energy (Adenosine triphosphate (ATP)). The reaction reversibility must be corrected using literature data or databases such as MetaCyc [50]. If no information is available, the reaction should be considered reversible [4, 7].

Localization

In this step, information about gene and reaction localization may be obtained [4, 7]. The compartments can be predicted using algorithms such as PSORT for prokaryotes [57], WoLF PSORT [58]for eukaryotic and LocTree [59] for both prokaryotic and eukaryotic, giving information about the protein cellular localization based on nucleotide or amino acid sequences. In prokaryotes, the compartments are limited to cytoplasmic and periplasmic space. In eukaryotes, proteins can be located in various compartments, including the mitochondrion, endoplasmic reticulum, or Golgi apparatus. For each compartment, the metabolite should be replicated, and either name and identifier should impress each localization. The intracellular transport reaction has to be included in the model in order to establish the connection between compartments [4, 7].

Biomass equation

Thereafter, the biomass composition is determined. Biomass reaction considers all known biomass constituents such as DNA, Ribonucleic Acid (RNA), proteins, lipids, carbohydrates and cofactors, and their fractional contributions to the overall cellular biomass. Biomass composition of an organism should be determined experimentally with cells growing in log phase, to obtain a more detailed composition. This reaction is very important for *in silico* simulations, where maximizing growth is the objective, because if a biomass precursor is not taking into account, synthesis reactions may not be needed for growth, affecting the essentiality reactions and their associated genes. The unit of all biomass precursor fractions is mmol per gram of dry weight (mmol/gDW), so the biomass reaction unit is per hour (h^{-1}) . Thus, the sum of the molar fraction of each precursor corresponds to the required to produce 1 g of dry cell weight. The biomass reaction can be represented by equation 1 [4, 7]:

$$\sum_{k=1}^{p} C_k.X_k \longrightarrow Biomass \tag{1}$$

Where,

- C_k represents the coefficient of the metabolite X_k.

However, in the absence of experimental data, the information obtained through genome annotation and from phylogenetically related organisms has to be taken into account. In turn,

GSMMs of phylogenetically close organisms can be used as a template to define biomass composition [3].

Energy requirements

The growth associated ATP maintenance reaction considers the energy in ATP form necessary for cell replication, embracing macromolecular synthesis and it must be added to the draft reconstruction. This energy requirement reaction should be determined experimentally, but if there is no experimental data available, it can be estimated by calculating the energy required for each macromolecular synthesis. The non-growth associated ATP maintenance reaction should be also added and represents cellular non-growth associated ATP requirements, by ATP hydrolysis. When no experimental data for energy requirements is available, the reaction values can be estimated from related organisms data or by fitting the model results to growth rate data [4, 7].

Growth requirements

The information about growth medium requirements should be collected. These requirements include metabolites present and if there are any auxotrophies, base medium (e.g. water, protons, ions) and rich medium compositions. This information is very relevant and crucial for the simulation and evaluation of the model. Uptake or secretion rates should be collected too, if available. The identification of growth requirements can be obtained by doing experimental work or by literature research [4, 7].

Manual Curation and Metabolic Network Assembly

In the reconstruction assembly, the reactions are evaluated in their metabolic context, identifying missing gene annotations in order to simplify the gap analysis and to debug the validation stage. Manual curation will also identify and obtain additional information about certain reactions which can be added to the model, or simply noted so that they can be easily recovered if necessary. Then metabolic function is verified using information about metabolic genes retrieved from genome annotation. This analysis must be supported by experimental data or literature, due to possible errors or incomplete genome annotation. Thus, the information about organisms phylogenetically close is used, when no organism-specific information can be found in the literature, receiving a lower confidence score to be easily identified in the case of possible problems during future simulations. Certain reactions that contain generic terms such as protein or electron acceptor must be excluded, because they are not specific enough. In this stage, it is important to identify the substrate and cofactor specificity, using organism-unspecific databases [4, 7], such as KEGG [38] and BRENDA [39] or organism-specific databases, such as EcoCyc [60], which list information about the enzymes and can have information about kinetic parameters (BRENDA [39]), and substrate and cofactor utilization (BioCyc [52]). Often, this curation is very time-consuming, and requires an intensive literature search [4, 7].

Before the next stage, the metabolites should be associated with at least one of the following identifiers: Chemical Entities of Biological Interest (ChEBI) [61], KEGG [38], Biochemical, Genetic and Genomic (BiGG) [51], ModelSEED [62, 63], or PubChem [40], in order to be recognized by software tools and even by other scientists [4, 7].

2.4.3 Converting the metabolic network to a stoichiometric model

In the third stage, commonly named conversion from reconstruction to a mathematical model, the GSMM reconstruction is converted into a mathematical format and certain specific constraints are defined. The conversion process is defined in three main steps. Firstly, a mathematical representation of the model is developed, more properly of the metabolic network reactions list, through the construction of a stoichiometric matrix, called S matrix, in which the columns represent the reactions while the lines represent the metabolites. A metabolic network of M metabolites and N reactions can be represented as in equation 2 [4, 7]:

$$\sum_{j=1}^{N} S_{ij}.v_j = 0, \quad i = 1, ...M \tag{2}$$

Where,

- v_j represents the rate of a reaction j;
- S_{ij} represents the stoichiometric coefficient of a metabolite i in a reaction j;
- i represents a metabolite concentration;
- j represents a reaction.

The values of the matrix represent the coefficients, where the reagents of each reaction are defined with a negative coefficient, while the products have a positive value. Metabolites that participate in, at least, one reaction have at least one non-zero entry in the S matrix. As mentioned before, a state of mass conservation is assumed, i.e. a steady-state [4, 7], which can be described as in equation 3:

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

Where,

- S corresponds to the stoichiometric matrix $S(MxN)$;
- v corresponds to a vector of reaction fluxes.

The limits of the system are defined, adding an exchange reaction to all metabolites that can be secreted or consumed by the target cell. These exchange reactions can be used in future simulations to define environmental conditions such as carbon sources. Finally, restrictions must be added to the mathematical model, such as thermodynamics or reaction directionality and enzymatic capacity. Then, as the GSMM reconstruction must be in a computer readable

format, these constraints are represented as inequalities by imposing a flux range for each reaction. Lower and upper bounds are defined to represent the model constraints in the inequalities [4, 7], as shown in equation 4:

$$\alpha_j \leq v_j \leq \beta_j, \quad j = 1, ...N \tag{4}$$

Where,

- v_j represents the flux vector;
- α_j represents the lower bound;
- β_j the upper bound.

To represent a reversible reaction the lower bound of the inequality can be -10000, representing minus infinity, and the upper bound can be defined as 10000, corresponding to plus infinity. To represent an irreversible reaction, one of the bounds is defined as zero. If the maximum and/or minimum of flux values are known, these should be set in the constraints as upper and/or lower bounds, respectively [4, 7].

Thus, a condition-specific model is obtained. Thereby, the set of solutions obtained will decrease and will be more confined, more stable and more viable [4, 7]. For the simulation's performances, the mathematical representation of the model should be saved in a standard and computational-friendly format, for example, SBML [37]. Some computational tools specialized in model simulations [4, 7] such as OptFlux [64] or COnstraint-Based Reconstruction and Analysis (COBRA) toolbox [65] need to be used in order to validate the model.

2.4.4 Metabolic model validation

The model evaluation, which is also called debugging mode, represents the fourth stage of GSMM reconstruction. In this stage, the model is verified, evaluated and validated. The model obtained in the previous stage is tested, among others, for its ability to synthesize biomass precursors, which is essential for cell survival. This evaluation identifies missing metabolic functions, namely network gaps, which are added by repeating the stage of curation and conversion. Consequently, the reconstruction process is considered an iterative procedure. Therefore, it is important to know when to stop the iterative process and to know when a reconstruction has been completed. This decision is normally based on the objective of the reconstruction [4, 7].

Initially, several reactions and functions are expected to be missing from the model, leading to dead-end metabolite production. So, a first gap analysis is done and then through intensive literature research and maybe a genome re-annotation, candidate genes and reactions are identified in order to fill the gaps. There are at least two approaches for identifying gaps in the reconstruction. In connectivity-based approach, it is possible to count the non-zero inputs on each line of the S matrix and identify the metabolites, which are only produced or consumed. A

second approach is based on the functionality of the model, in which the ability of the models to transport flux through each network reaction is tested, thus identifying blocked reactions, which are associated with one or more dead-end metabolites. For gap filling, KEGG [38] maps, biochemical textbooks, or other available biochemical maps can be used as resources in order to identify the metabolic pathway of the dead-end metabolites. A gap should be filled to assure the model's functionality (for example, biomass precursor synthesis reaction), to allow the model to perform a function that the organism is able to do, or when there is experimental evidence that a reaction is present in the organism. When no information is available, a gap reaction should be filled if necessary to ensure that the model is functional, because if the reaction is falsely added, the predictive capacity of the model can be affected [4, 7].

At this point, the model must be tested for its ability to produce each biomass component, initially in a standard medium condition and then in other growth media, i.e. in all known growth conditions of the target organism. Under most of these conditions, the model should be capable to synthesize biomass, otherwise, missing functions can be identified. By-products secretion information is very helpful to evaluate the phenotype simulated by the model. If this information is available, it must be added to the model, in order to further refine it. To identify other gap fillings, blocked reactions need to be determined [4, 7], for example through FVA [66]. Depending on the function of the blocked reactions, they can be connected to the model network. In the next step, a simulation with a single gene deletion can be performed. The analysis between the *in silico* results and the experimental data can provide a model improvement and consequently an improvement in its predictive capacity. A matrix with false positive and false negative predictions is constructed. A possible explanation to false positives is the missing of regulatory rules, reactions falsely included and/or an incomplete biomass reaction. The false negatives can be explained by missing metabolic transport reactions and missing enzyme reactions. At this point, the model has been tested for its ability to, for example, produce biomass and to grow under certain conditions [4, 7]. So now, the model must be tested if it can reproduce the known inabilities of the organism. If any errors are detected, caused by falsely added reactions or missing information, they should be corrected doing literature research. Finally, the growth rate is quantitatively evaluated. A too slow growth indicates that at least one biomass precursor cannot be adequately synthesized. Whereas, a higher growth rate has many possible explanations and there is not a general rule for the corrections. Values of the energy necessary for growth wrongly estimated, missing or incorrect constraints, such as missing regulation or incorrect reaction directionality, and also falsely added reactions are among the reasons for a higher growth rate than expected [4, 7].

Validation and simulation methodology

After the construction of the mathematical model and the constraints applied, phenotype and behaviour can be predicted, for each environmental condition established [4, 7]. In order to predict exchange fluxes, Metabolic Flux Analysis (MFA) can be performed. MFA is an approach used to estimate some intracellular fluxes of a defined metabolic network when other

experimentally measured fluxes are given as input. It allows the user to perceive the metabolism balance, that is, the way that the organism can convert substrates into biomass and chemicals, and its ability to predict the model metabolic capability after genetic or environmental changes. MFA is useful for predicting possible metabolic limitations [4, 67].

A commonly used approach to evaluate the model predictive potential is FBA, which uses linear programming and in a steady-state system determines the optimal reactions flux distribution through the optimization of an objective function, satisfying the set of restrictions imposed. FBA calculates the metabolite flux through the metabolic network, thus being able to predict organism growth rates or production rates of a metabolite with biotechnological interest. This approach is a restriction-based method and therefore restrictions are first imposed on a biochemical system, defining a space of possible phenotypes or solutions, and then an objective function is determined in order to be optimized within that space to determine the most likely phenotypic state of the system. The objective function can be the maximization of biomass production/growth rate, the minimization of energy use or the maximization of the production of the final product of interest, among others. The set of restrictions can be divided into four categories: physical-chemical (such as mass conservation), topological (such as compartmentalization and spatial restrictions associated with metabolites or enzymes), environmental (such as media composition, pH, temperature), and thermodynamic (for example reversibility of the reaction) [4, 68–70]. However, for the same objective function, it is possible to obtain a set of several optimal solutions with different fluxes distributions. In these cases, other simulation methods such as parsimonious Flux Balance Analysis (pFBA) should be used. pFBA was developed as an extension of FBA which, using the set of optimal solutions calculated by the FBA, selects the solution that minimizes the sum of all the flows. In a biological sense will correspond to the solution that maximizes the objective function, usually the maximization of the biomass production, using the minimum of resources required [71, 72]. The linear programming problem of FBA can be represented as in equation 5:

$$\begin{aligned} &\text{Maximize} && Z \\ &\text{subject to} && S \cdot v = 0 \\ & && \alpha_j \leq \mathrm{v}_j \leq \beta_j, \quad j = 1, ..., N \end{aligned} \tag{5}$$

Where,

- Z is the linear objective function;
- v is the flux vector;
- S is the stoichiometric matrix;
- α_j and β_j is the lower and upper bound, respectively.

After the model has been evaluated and considered to be realistic, other simulation tests can be performed. FVA is a simulation method used to test the robustness of a metabolic model. This approach finds the minimum and maximum flux of the model reactions, calculating the

range of flux variability which reaches optimal as well as sub-optimal objective states. It is commonly used in the optimization of antibiotic production [7, 66]. To simulate the phenotype of mutants, two algorithms are frequently used: Minimization of metabolic adjustment (MOMA) and Regulatory on–off minimization (ROOM) [4, 68]. MOMA is a quadratic programming algorithm based on the same FBA stoichiometric constraints, but the assumption of optimality for gene deletions is not justifiable. It tends to minimize the distance between wild type and mutant flux distributions [4, 68, 73, 74]. In turn, ROOM is also quadratic programming algorithm and predicts the organism metabolic steady-state after gene knockout. ROOM differs from MOMA once it tends to minimize the number of significant changes in flux distribution from the wild-type flux distribution [4, 68, 74, 75].

2.4.5 Applications

GSMMs are very useful, having a wide variety of different applications, such as predicting phenotypic behaviour under various genetic and environmental conditions, analysing the robustness of a network [4], improving the production of chemicals and materials, predicting drug targets and enzyme functions, analyzing pan-reactome, modeling interaction among multiple cells or organisms and understanding human diseases [76]. In the final stage, as the desired *in silico* capacity and the necessary content has been reached, the GSMM reconstruction can be used in a prospective form, applying it to a scientific study, such as the case of the unveiling of the lambic beer flavour[4, 7].

In conclusion, the complete reconstruction of a GSMM is a long and laborious process. Therefore, the development of bioinformatics and software tools was necessary and can greatly accelerate the reconstruction process by automating some steps [5].

2.5 Genome-scale metabolic models reconstruction tools

Over the years, the interest in reconstructing high-quality GSMM has been increasing and therefore the number of new tools developed for GSMMs reconstruction has also increased [7, 36]. Several GSMMs reconstruction tools have been developed to accelerate the reconstruction process by automating several steps and tasks, such as draft network generation and formulation of biomass equation [5, 36]. There is no perfect tool, each one has its own disadvantages and advantages, and the strengths in some tools can be the weaknesses in others. Some features have the same great performance in the different tools, such as in the use of updated databases and resources and in the continuous software maintenance [36]. Although in other features, all tools have a low performance, for example in traceability or automatic refinement using experimental data. In Table 2 (adapted from Mendoza et al. 2019), seven different GSMMs reconstruction tools and their main features are listed. These tools were chosen because they cover most of the software platforms available for free [36].

In 2019, Mendoza et al. [36] evaluated these seven tools. For example, when the automatic reconstruction process was tested, ModelSEED [62, 63] and CarveMe [6] were considered as outstanding while *merlin* [77] as poor because users should infer more in metabolic network refinement in order to get a model ready to perform simulations [36]. But on the other hand, *merlin* [5, 77] was considered as outstanding concerning the workspace for manual curation and information provided to help users during this stage. In this feature, CarveMe [6] and ModelSEED [62, 63] were assessed as poor, as they do not provide further information or a workspace for manual refinement and curation. ModelSEED [62, 63] and CarveMe [6] are examples of tools that can use media composition to fill the network gaps [36]. Pathway Tools [45] and AUtomatic REconstruction of MEtabolic models (AuReMe)[78], besides media composition, can also utilize known metabolic compounds. However, the use of biolog phenotype arrays, knockout experiments and different types of omics data, such proteomic, to automatically curate the draft model is not yet performed by any of the tools listed in Table 2 [36].

AuReMe [78], MetaDraft [79] and Reconstruction, Analysis and Visualization of Metabolic Networks (RAVEN) [80] are capable to use and import metabolic networks, using level 3 SBML files [81], but AuReMe [78] generates networks with slightly differences when level 2 SBML files [82] is used [36]. In relation to outputs, MetaDraft [79], *merlin* [5, 77] and RAVEN [80] are able to export the networks in level 3 SBML files with flux balance constraints annotations [83], SBML groups [84] and Minimum information requested in the annotation of biochemical models (MIRIAM) [85] compliant controlled vocabularies annotations [36]. CarveMe is only able to reconstruct models for prokaryotics [6], while the other six tools can reconstruct models for prokaryotics and eukaryotic organisms [5, 36, 77].

AuReMe [78] has traceability as a priority, keeping a register of changes made in the draft network. Its workspace is a command line. This tool is a good option to reconstruct high-quality drafts models, where models for a phylogenetically close species are available because the time required to obtain a manually curated model of high-quality is reduced [36]. AuReMe [78] also has some disadvantages such as the interface which is not user-friendly and it does not provide good assistance for manual refinement [36]. In turn, CarveMe [6] which also uses a command-line to reconstruct GSMMs, is one of the faster tools, providing GSMMs ready to be validated and simulated. However, many reactions must be checked manually because the gap-filling procedure is automatic and it can include reactions that are not really present in the organism [36]. Based on Python, CarveMe [6] can also be performed to study microbial communities[6]. The workspace is not suitable for manual refinment making this reconstruction step difficult to handle [36]. *merlin* [5, 55, 77] is an user-friendly Java™tool that has useful features such as the automatic genome re-annotation through BLAST or HMMER and a platform provided for manual refinement, guiding the user through the entire reconstruction process. Reactions localization can be automatically predicted as well as GPR associations [5, 36, 77]. *merlin* [5, 55, 77] has a feature which creates a pseudo biomass reaction with estimated values for proteins, DNA, RNA and cofactors [77]. Another reconstruction tool is MetaDraft [79], a user-friendly tool, which is Python-based and was designed to reconstruct

GSMMs from previously curated models [36]. MetaDraft [79] is more robust for handling models and very suitable for non-specialists due to the graphical user interface but is not good for manual curation [36]. In turn, ModelSEED [62, 63] has a short computational time to get a draft GSMM and it is able for the reconstruction of plant and microbial community models [36]. ModelSEED [62, 63] can fill automatically the gaps, but the reactions must be checked. Therefore, the draft obtained is ready for be validated and simulated. This tool is able to identify GPR associations, to automatically predict reactions localization and generate a biomass reaction [5, 36]. Nevertheless, ModelSEED [62, 63] does not provide a platform with good assistance for manual curation [36]. Pathway Tools [45] is a software to reconstruct and curate GSMMs and has an outstanding user-friendly interface [36], which provides a suitable workspace for manual curation. One of the disadvantages of this tool is that the final reconstructed models do not have all the fields required [36]. Finally, RAVEN [80] is a GSMMs reconstruction and curation tool which requires the use of the commercial tool MatLab®[5, 36]. It allows several reconstructions to be generated with great parameter flexibility and it is appropriate for model reconstructions of less studied organisms because it is possible to integrate information from KEGG and MetaCyc databases. The available interface for manual refinement is not good, which is a disadvantage [36]. RAVEN [80] is also able to predict automatically reactions localization [5].

Therefore, *merlin* [5, 55, 77] can be considered one of the most complete reconstruction tools because provides a user-friendly interface, facilitating manual curation which is indispensable in any reconstruction process. Moreover, has access to updated databases and resources. The reconstructed model obtained has all the fields required and can be exported in standard formats [5, 36, 77]. Additionally, *merlin* [5, 55, 77] provides excellent features that will help in the reconstruction process such as the automatic annotation, GPR association identification, reaction localization prediction and estimation of a pseudo biomass equation [5, 77].

There are some tools for automatic gap filling, for example, FastGapFilling and GapFind/-GapFill [86, 87]. FastGapFilling uses only Linear Programming (LP) [86, 87] and is capable to integrate gap-filling and, flux and stoichiometric consistency [87]. This tool creates a linear problem with all reactions, where an objective function is defined based on their weighted fluxes, and the biomass reaction has a variable weight, not using an integer variable. While the algorithm increases the biomass flux, it minimizes the weighted minimum sum of fluxes from new reactions [86]. GapFind/GapFill find gaps, for example metabolites unable to be produced, and fills the network, for example, adding reactions to the model in order to produce the metabolites [86]. This tool uses Mixed-Integer Linear Programming (MILP), which computes a minimum set of reactions to add, making it computationally expensive when the candidate reactions set is large. This approach is also used to integrate transcriptomic data into GSMMs in order to simulate the model under a certain physiological state [21, 86]. For both tools, the reactions added to fill the gaps may be analysed and validated for each organism [86].

Table 2: GSMMs reconstruction tools and their main features *

Reconstruction tool	Mapping method	Reactions source	Associated databases	Version	Type of software	Reference
AuReMe	Pantograph (Inparanoid and OrthoMCL)	Template model(s)	BiGG-MetaCyc	1.2.4	Command line	[78]
CarveMe	Diamond, eggNOG-mapper [88] (should be run externally by the user)	Template model	BiGG	1.2.1	Command line	[6]
merlin	Mapping from annotation with BLAST [44] or HMMER [89]	Database	KEGG	3.8	Standalone interface	[55, 77]
MetaDraft	Autograph (Inparanoid)	Template model(s)	BiGG	0.9.2	Standalone interface	[79]
ModelSEED	Annotation ontology map from RAST [90] data	Template model	ModelSEED	2.2–2.4	Online service	[62, 63]
Pathway Tools	Pathologic	Database	MetaCyc	22	Standalone interface	[46]
RAVEN	Autograph-type method from BLASTP [91] and Bidirectional BLASTP	Database-Template model(s)	KEGG-MetaCyc	2.0.1	Command line	[80]

* (Adapted from Mendoza et al. 2019)

FastGapFilling is considered more efficient because has to solve only a few LP problems, whereas MILP-based approaches needs to solve several LP problems, requiring much more computation time [86].

Although there are already some tools to facilitate and improve model reconstruction and simulation, it is still a time consuming and laborious process. So, there are some features that should be improved such as, algorithms that include genes and reactions in the models, through the use of experimental data. The lack of experimental information is one of the biggest problems, specially in respect to biomass composition since the quality of the model is compromised and hence its ability to accurately predict phenotypes [36].

2.6 Simulation tools

After the reconstruction process, models must be validated and simulated. Tools with no simulation features need the use of additional software such as COBRA Toolbox [65], OptFlux [64], ReFramed [92] and MEWpy [93]. COBRA is a framework that allows an integrative analysis of systems biology data and a quantitative prediction of viable phenotypes. COBRA Toolbox [65] is a software of interoperable COBRA methods, with applications in biology, biomedicine and biotechnology, because it has functions that can implement COBRA protocols to any biochemical network [65]. However, COBRA Toolbox methods require MatLab®to be used. COBRA for Python (COBRApy) is a Python package that allows the user to access the methods of COBRA Toolbox [65], and therefore does not require MatLab®[94] in contrast with COBRA Toolbox versions [65]. COBRApy facilitates the representation of the biological processes of metabolism and gene expression[94]. Nevertheless, COBRA Toolbox [65] has several tools and methods with different objectives, namely, visualization and reconstruction of network models, integration of omics data, metabolic engineering, gap-filling and FBA [65]. OptFlux [64] is an open-source software and user friendly, being the first tool to identify Metabolic Engineering targets using Evolutionary Algorithms/Simulated Annealing Metaheuristics or the previous algorithm proposed by OptKnock [64, 95]. It also allows stoichiometric metabolic models to simulate wild type and mutants phenotypes of an organism, using different methods, such as FBA, MOMA or ROOM [64]. It performs MFA by calculating the admissible flux space given measured fluxes set, and also it can analyze the pathways by calculating Elementary Flux Modes [64, 96]. OptFlux has several methods for simplification of the model and other pre-processing operations in order to optimize algorithms [64]. OptFlux allows the analysis of the model structure through a visualization module, which is compatible with the Cell Designer layout information [64, 97]. ReFramed [92] implements several constraint-based simulation methods including FBA, pFBA, MOMA , and contains interfaces to other COBRA libraries, such as escher, COBRApy, and optlang. In addition is possible to construct GSMMs for microbial communities [92] from existing ones and simulate them using the methods implemented. In specific for microbial communities, SteadyCom can be used to [98]. MEWpy [93] is an integrated Metabolic Engineering Workbench for strain design optimization, developed in Python [93]. For simulations, it allows the simulation of different types of models (such as metabolic, kinetic, community and Gecko), considering different phenotypes [93]. In the optimization process, it uses Evolutionary Computation based on the optimization of strain design by doing a knock out or forging an over/under expression in reactions, genes or enzymes. ReFramed [92] and COBRApy are the simulation environments that are currently available in MEWpy [93].

2.7 Acid lambic Beer

Belgium is known for its traditional spontaneous mixed fermentation beers, such as lambic beers [10, 99]. Originally brewed in Belgium, lambic beer is one of the oldest types of beer still produced today [10, 100, 101]. Becoming increasingly popular around the world, this Belgian beer has a unique alcoholic flavour profile characterised by a refreshing acidity with fruity notes and little residual carbohydrates [102]. Prominently more acidic and weakly carbonated, the characteristic flavours and aromas of this beer are created using distinct microorganisms [10], such as *Enterobacteriaceae*, *Pediococcus* and *Brettanomyces* , unable to be obtained by pure culture fermentations [10, 100, 101]. As mentioned before, sour beers are gaining interest worldwide [10], especially in the United States of America [100], and thus, the American craft-brewing beers are quite similar to the method of lambic beer production [100].

Lambic beer is traditionally produced during the winter by inoculating the wort in an open coolship. The lambic beer fermentation is a very long process that can last several years [99, 100, 103] (up three years [102]). This beer is characterized by its unique flavour, rich and complex associated with the rich microbial flora involving bacteria and yeasts that thrives throughout the fermentation [10, 100, 103]. In the production of lambic beer, the necessary microorganisms are obtained through the air and the successful production depends on the low temperature which is important to prevent the growth of undesired microorganisms [10, 100]. The fermentation and maturation phases of this type of beers take place in horizontal oak or brown barrels, contributing to microbial inoculation [10, 100]. In lambic beer production, the extraction of wood compounds has no interest, unlike the production of port wine and whiskey, so the old barrels are reused, preserving the resident microbiota [10, 100]. *B. bruxellensis* yeast and *P. damnosus* Lactic acid bacterium, which are related to the beer maturation phase, have been isolated from lambic beer barrels [10, 100]. Since the casks are used constantly and they are only cleaned superficially with high pressure water between production batches, these microorganisms are probably originated from previous lambic beer production processes [10, 100]. *P. damnosus* has been identified as a key microorganism in the fermentation of lambic beer [100, 104–106] and *B. bruxellensis* has been identified by Kufferath and Van Laer in 1921 as the yeast that confers to lambic beer its characteristic flavour [100].

The beer is composed of many (about eight hundred) different volatile and semi- volatile compounds, but only a few are active flavourings (between ten and thirty) [108].

2.7.1 Acid lambic beer production

Lambic acid beer is produced through the spontaneous fermentation of the wort (water, barley malt, unmalted wheat, and aged dry hops [102]) in horizontal wooden casks [102], performed together with different species of yeasts, for example species of *Brettanomyces* and *Candida* genus, and by some bacterias, such as Acetic acid bacteria (AAB) and LAB [100, 109]. The lambic beer production starts with the wort preparation, which is a mixture of barley malt

and unmalted wheat, in a proportion of approximately 60-40, respectively [100, 109]. After the addition of yeast culture, spontaneous fermentation is initiated, but it is performed by microorganisms from the environment and/or from the previous beer. The cooked wort is cooled overnight in open recipients. In the next day, the wort that was inoculated during the night is transferred to wooden barrels, initializing the fermentation which can last up to two years, with temperatures between 0 and 25°C, because the room atmosphere is not regulated and therefore depend on the seasons. Lambic is the resulting product and a refermentation of this product gives rise to another type of beer called gueuze. Therefore, lambic is considered the basis for other types of beers [100, 109].

This production is usually performed in the coldest months to avoid contamination, starting the wort fermentation in September and finishing in April of the following year. The total duration of lambic beer fermentation can take up to three years [100, 109].

The production of lambic beer can be divided into four phases, each characterized by the action of specific microorganisms [100, 109], as observed in figures 2 and 3 (both adapted from Spitaels et al. 2014):

1. ***Enterobacterial* phase** - During the cooling of the wort and during the first month of fermentation, it predominates microorganisms such as *Enterobacteriaceae* (*Enterobacter*, *Escherichia*, *Klebsiella*, *Hafnia*, *Citrobacter* and non-fermental maltose yeasts (e.g. *Kloeckera apiculata*). LAB are minimally present. Thus, butanediol, dimethyl sulfide and formate production prevail, with lower production of ethanol, acetic acid and lactic acid [10, 100, 109];

2. ***Saccharomyces* phase** - After glucose exhaustion, the non-fermental maltose yeasts disappear, the pH decreases below 4.4 and the ethanol concentration increases above 2%, completely inhibiting the activity of *Enterobacteria*. This stage is the main one of alcoholic fermentation, dominated by *Saccharomyces* (*Saccharomyces cerevisiae*, *Saccharomyces pastorianus*, *Saccharomyces bayanus*) which has an important role in the total attenuation of the wort. Fermentation is not as intense as in a normal beer, which may be due to the fact that *Enterobacteria* have consumed various amino acids and peptides at the beginning of wort's fermentation. In this phase prevails the production of ethanol, higher alcohols and esters [10, 100, 109];

3. **Acidification phase** - After about four months of fermentation, the LAB, especially *P. damnosus*, and the AAB such as *Acetobacter* and *Gluconobacter*, appear in large quantities in the wort, leading to the production of lactic acid and acetic acid. The number of LAB reaches its peak during the transition from spring to summer, corresponding to the time between the sixth and the eighth month of fermentation, because these microorganisms need higher temperatures for their growth. *Brettanomyces* yeasts, in particular *B. bruxellensis*, replaces *Saccharomyces* yeasts, approximately in the eighth month. *B. bruxellensis* is responsible for the further fermentation of the wort, occurring in lambic fermentation for another eight months. The acidification phase is characterized by

the strong production of acid compounds such as lactate, ethyl lactate and also, but to a lesser extent, the production of acetate, ethyl acetate, diacetyl and isobutyrate [10, 100, 109];

4. **The maturation phase** - Approximately after ten months of fermentation, the number of LAB decreases gradually, as does the activity of Brettanomyces, which decreases later. In this phase occurs the fermentation of superior dextrins, the reduction of diacetyl and dimethylsulfide and the wort is gradually attenuated. The beer maturation takes about one year and therefore the number of LAB increases again with the increase of temperature, coinciding with the hottest summer months, corresponding to the second year of fermentation [10, 100, 109].

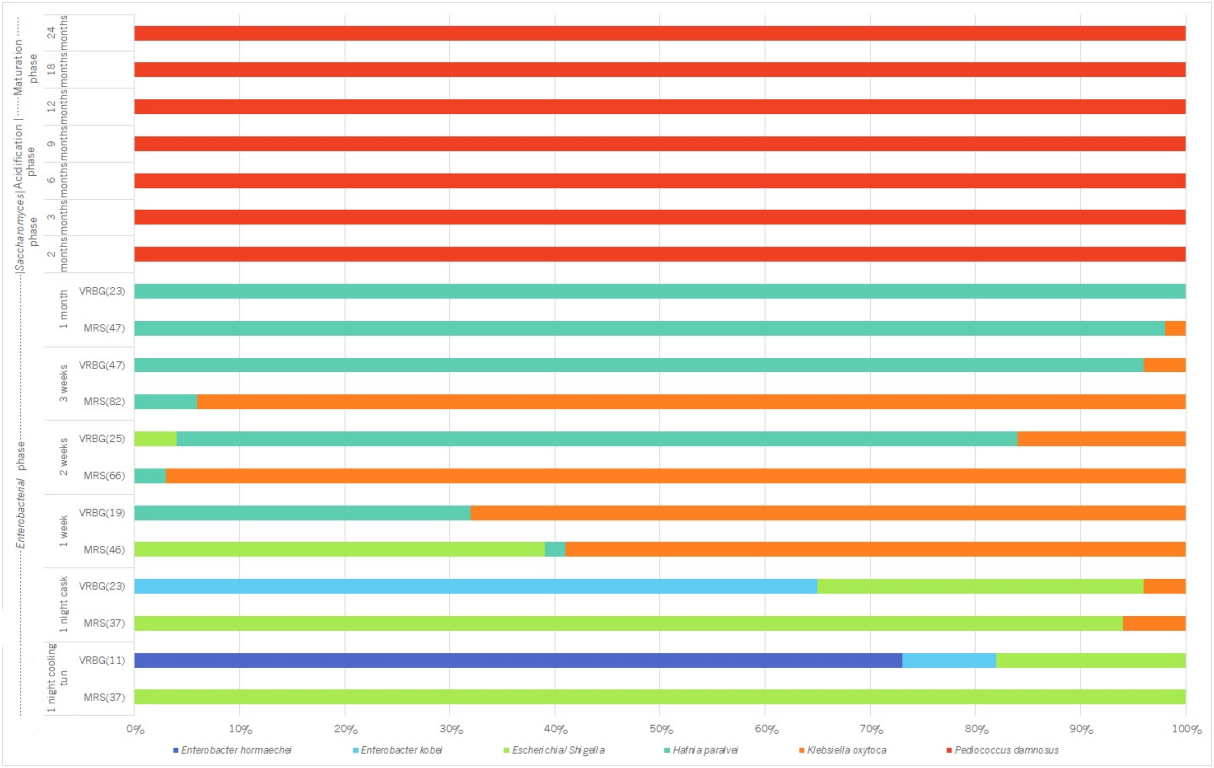


Figure 2: Bacteria present in lambic beer isolated from two different agar culture media.

- Bacteria were isolated in two different agar media namely Man-Rogosa-Sharpe (MRS) and Violet red bile glucose (VRBG). The isolates' number is given between brackets.
Adapted from Spitaels et al. 2014 [100].

The lambic beer fermentation process has a worldwide great interest, and therefore the study of the microbial community present in its production is important. Therefore, in this work we will study the LAB *P. damnosus* and *B. bruxellensis* yeast, which play an important role in the final stages of fermentation, being isolated in the acidification and maturation phase [10, 100, 109].

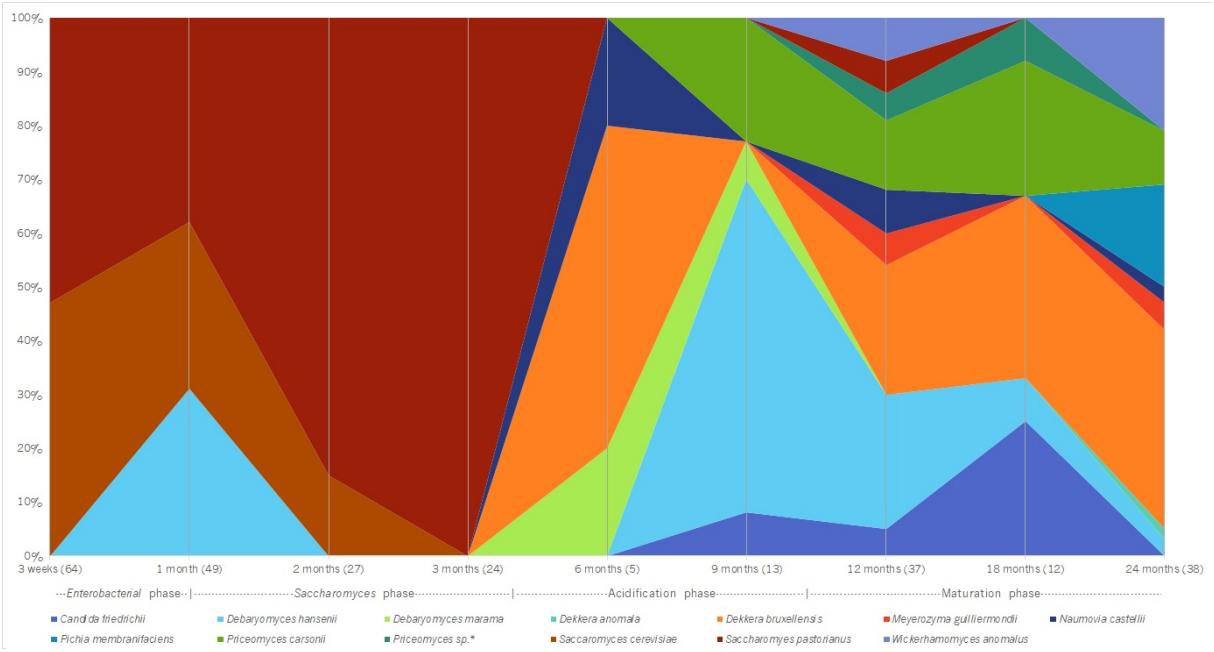


Figure 3: Yeasts present in lambic beer isolated from two different agar culture media.

- Yeats were isolated in two different agar media namely DYPAI (2% glucose, 0.5% yeast extract, 1% peptone and 1.5% agar; weight/volume) and UBAGI which is a general yeast agar isolation medium for beer environment. The isolates' number is given between brackets.
Adapted from Spitaels et al. 2014 [100].

2.8 Lactic acid bacteria

LAB are a group of Gram-positive bacteria, which metabolize sugars, converting them mostly into lactic acid [110]. Classified in the Firmicutes phylum, class *Bacilli* and order *Lactobacillales*, the LAB belong to several genera, including *Pediococcus*, *Lactobacillus*, *Lactococcus*, *Streptococcus*, *Oenococcus* and *Enterococcus* [110, 111], as shown in Table 3. LAB are defined by a ubiquitous and heterogenous family of microorganisms, characterized by similar metabolic capabilities [112]. Non-spore forming bacteria and microaerophilic, LAB can be cocci or rods, having a high tolerance for low pH. Being considered as catalase-negative and fastidious organisms, these bacteria reside in several different habitats, including human cavities, such as gastrointestinal or oral cavity, as well as plants, meat, vegetables and processed dairy products [111, 112]. With a low DNA content of G+C [111, 112], they are strictly fermentative and facultatively anaerobic [10]. LAB are characterized by the production of growth inhibition substances such as diacyls, hydrogen peroxide and bacteriocins, which have a preservative and antimicrobial action, preventing the proliferation of bacteria and pathogens responsible for the spoilage of food, also acting in the control of human pathogens [111]. Nevertheless, more information is needed on the LAB antimicrobial potential against human pathogens, especially in the current era where antibiotic resistance is increasing. Therefore, through the new technologies, it is important to characterize and identify new LAB strains that have more powerful antimicrobial activities [111].

LAB belong to the most important microorganisms used in food fermentations, being extensively used in food industry as promoters in the fermentation of, for example, fruit, drinks, meat, vegetables and milk, playing a key role in food preservation, as well as in the development of the texture and taste of food, improving them [110, 111]. Since they are present in the gastrointestinal tract, some species of this bacteria group are used as probiotics, because they are of human origin and confer beneficial effects on human health [112].

To obtain energy, LAB ferment carbohydrates and as the final electron acceptor, it utilizes endogenous carbon sources instead of oxygen [111]. They are protected against oxygen by-products, for example, hydrogen peroxide by peroxidases, considered thereby as aerotolerants [111]. LAB has an optimal growth at pH between 5.5 and 5.8 [111], with a temperature range from 30ºC, corresponding to mesophilic LAB species, to a maximum of 45ºC, being considered as thermophilic LAB species [10]. The nutritional requirements of these bacteria are complex for macromolecules, nucleotides and amino acids as well as vitamins and minerals [111]. Based on the fermentation end-products, these microorganisms can be classified into homofermentatives and heterofermentatives [111]. When the main product produced by LAB through fermentation of sugars is lactate, the microorganism is homofermentative, whereas heterofermentatives LAB, besides lactate, produce carbon dioxide and acetic acid or alcohol [111].

Several LAB genomes have been sequenced and are available to public users, leading to advances in industry, based on genome analysis [113]. Thus, the knowledge on LAB has greatly increased, contributing to the industrial progress of the LAB applications [113]. As mention in the section 2.3, GSMM has become a necessary tool to study and design new improved strains in order to achieve desired goals [113]. Nevertheless, few GSMMs for LAB strains have been reconstructed, being the first GSMM for *Lactococcus lactis* (*Lc. lactis*) *ssp. Lactis* IL1403, composed by 621 reactions and 509 metabolites [113, 114]. Thereafter, metabolic models of other LAB were also reconstructed such as *Lactobacillus plantarum* (*Lb. plantarum*), *Streptococcus thermophilus* and *Bifidobacterium prausnitzii* [113].

LAB are present in the acid lambic beers fermentation, characterising the acidification phase of the traditional production process, between two and ten months of fermentation and the maturation process [10]. They are responsible for the acidic tart flavour of the beer due to their lactic acid production, making them indispensable because acidity is one of the main characteristics of lambic beer [10]. *Lactobacillus brevis* and *P. damnosus* are microorganisms belonging to the LAB that are often found in beers fermentation, especially the second one which plays a very important role in lambic beer fermentation [10]. Other spontaneously fermented acidic beers, such as American coolship ales (American beer which use the same production process of the Belgian lambic beer) and red-brown acidic ales, also require LAB in their microbiota core, participating in their fermentation and consequent production [10].

Table 3: Number of classified species of the LAB genus based on Taxonomy NCBI Database

Genus	Number of classified species	Reference
Lactobacillus	224	[115]
Lactococcus	16	[116]
Leuconostoc	16	[117]
Pediococcus	11	[118]
Streptococcus	128	[119]
Aerococcus	8	[120]
Alloiococcus	1	[121]
Carnobacterium	12	[122]
Dolosigranulum	1	[123]
Enterococcus	63	[124]
Oenococcus	4	[125]
Tetragenococcus	6	[126]
Vagococcus	13	[127]
Weissella	25	[128]

2.8.1 *Pediococcus damnosus*

P. damnosus is a Gram-positive bacterium belonging to LAB group, which was previously identified as *Pediococcus cerevisiae* (*P. cerevisiae*) [10, 129]. The genus *Pediococcus* belong to the *Lactobacillacea* family [129] and it is composed of eleven species [130] among *Pediococcus*

acidilactici, *Pediococcus pentosaceus*, *P. damnosus* and *Pediococcus parvulus* [129]. The strains which were formerly known as *P. cerevisiae*, are currently distributed by three different species: *P. damnosus*, *Pediococcus acidilactici* and *Pediococcus pentosaceus* [129]. Vandamme et al. in 2014 verified the clustering of the genus *Pediococcus* among *Lactobacillus*, through analysis of the 16S rRNA [131]. Thus, in the same year, Franz et al. confirmed through a phylogenetic relationship analysis based on 16S rRNA, that *Pediococcus* is phylogentically closer to *Lactobacillus* genus, being the *Lb. plantarum* one of the *Lactobacillus* species most similar to *P. damnosus* [132]. Also in 2014, Wassenaar and Lukjancenko established a phylogenetic tree of species from *Lactobacillus* genus and other LABs. Interpreting this tree and using *Lb. plantarum* as a reference, *Lc. lactis* and *Streptococcus thermophilus* are other species from different genera that are phylogenetically close to *P. damnosus* [133].

Over the years, the *P. damnosus* genome has been sequenced. There are nine published genomes from different strains: TMW 2.1532 [134], TMW 2.1535 [134], TMW 2.125 [134], TMW 2.1639 [134], TMW 2.1643 [134], VTT E-123212 [135], VTT E-123216 [135], P58 [135], and LMG28219 [136]. Overall, the genome size ranges from 2.07 Mega base pairs (Mbp) to 2.27 Mbp and the Guanine-cytosine (GC) percentage is between 38.1 and 38.9 [134–136].

P. damnosus is a facultative anaerobic and an immotile bacterium which does not sporulate under any provided culture conditions [129]. Is considered as a homofermentative LAB specie, since only lactate is produced by the sugar metabolism [10, 129]. Due to the lack of a respiratory chain, *P. damnosus* has to reduce the pyruvate for NAD^+ recuperation by the action of a lactate dehydrogenase enzyme in the Embden-Meyerhof-Parnas pathway [10, 129, 134]. Besides producing the D- and L-lactate configuration, *P. damnosus* also produces acetoin or a diacetyl which are responsible for the sarcina odour in spoiled beers [132, 137, 138]. The optimal growth temperature of this specie is 22°C and it has an optimum growth pH of 5.5 [129, 139]. *P. damnosus* cannot grow in an environment with a NaCl concentration of 4%, a temperature above 35°C, or at a pH of 8.5 [129, 139]. Ribose, arabinose, xylose, rhamnose, starch and lactose are not metabolizing by these bacteria, mostly fermenting glucose, fructose, cellobiose, mannose and galactose and only some strains can hydrolyze maltose, trehalose and sucrose [129, 132, 137, 138]. This LAB species is capable to split sucrose into glucose and fructose due to the sucrose-6-phosphate hydrolase gene and also it could convert L-malic acid in L-lactate by the malolactic enzyme [140]. *P. damnosus* is chemo-organotrophic and has a complex growth factor as well as amino acid requirements [139], from the twenty amino acids, *P. damnosus* is auxotrophic to all amino acids except for asparagine and glutamine [141, 142]. Needs nicotinic acid, pantothenic acid, biotin, riboflavin and pyridoxine for growth, not being able to use ammonium as a source of nitrogen [139, 141, 142].

P. damnosus is found in brewery environments, occurring in wine, beer and cider [10, 129]. Therefore, it has a great research interest and many studies have been reported that *P. damnosus* produces EPSs and bacteriocins, specifically pediocins [129]. EPSs contributes to a unique physicochemical property applied in food industry, being responsible for the food viscosity and contributing to many biological activities, such as anti-oxidation, anti-bacterial, cholesterol-lowering, immunoregulatory function, anti-tumor, anti-coagulant and even antiviral activities

[143]. Pediocins are bacteriocins produced by species of the genus *Pediococcus*, such as *P. damnosus* [129]. They have a bactericidal effect to sensitive Gram-positive bacteria. Pediocins have an important application in food industry too, due to their preservative property [129]. Implementation of pediocins as preserving food offer better solutions because it is a natural compound which protects food against undesired pathogens, and is well accepted by the consumers [129]. Besides its preservative feature, pediocins have an antimicrobial property which can be used as complements of antibiotics, facing the bacterial resistance [129], as studies have shown for nisin [144–146].

Despite the lack of information about the cell wall composition of *P. damnosus*, there are several studies on LAB. Therefore, based on the cell structure of gram-positive bacteria and LAB studies, it is possible to suppose that *P. damnosus* has a cell wall composed by three major components: Teichoic acids (TAs), Peptidoglycan and polysaccharides [110, 147]. An adapted scheme from Delcour et al. of *P. damnosus* cell wall is illustated in figure 4.

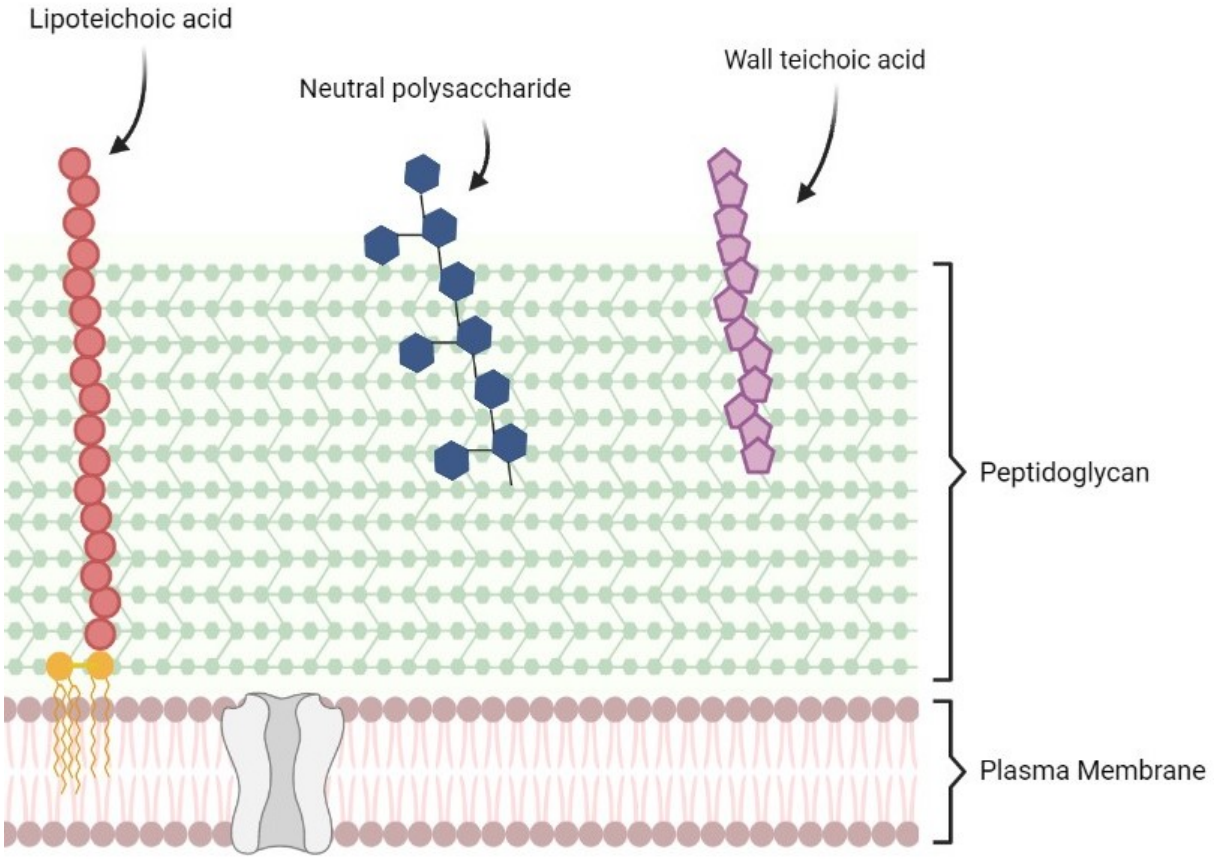
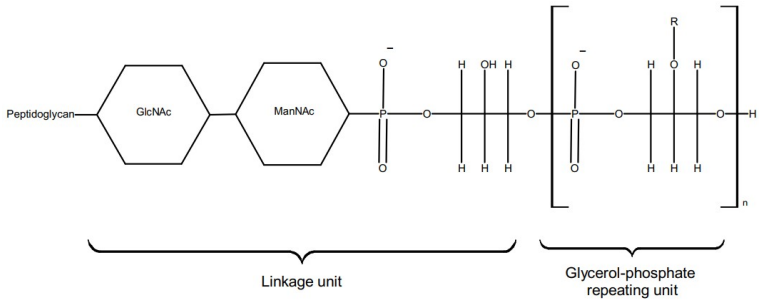


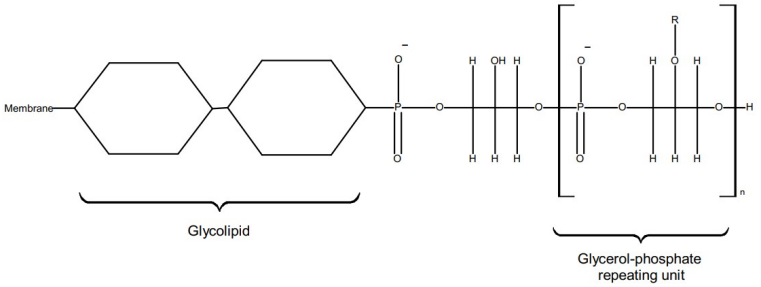
Figure 4: Cell wall structure of *Pediococcus damnosus*.

The plasma membrane has proteins integrated and it is covered by a PG multilayer. There are also present WTA, LTA and neutral polysaccharides.
Adapted from Delcour et al. 1999 [147]. Created with BioRender.com.

TA are presented in the cell wall of several Gram-positive bacteria [110]. In 2005, Fujii et al. has reported the presence of two types of TA in *P. damnosus*: WTAs and LTAs . Nevertheless, no studies about the composition of these cell wall compounds have been published. The structure and abundance of WTA and LTA differ between species and strains, and also depend on the medium in which the bacteria grow [147]. Usually, the structure of a WTA are a poly-glycerophosphate or a poly-ribitolphosphate chain, covalently linked to PG via a linkage unit of a disaccharide and a glycerol-phosphate unit [110]. Based on a study of *P. cerevisiae* WTAs [149], *P. damnosus* have WTA composed by poly-glycerophosphate chain instead of poly-ribitolphosphate chain [110, 149]. In turn, LTA are composed by a poly-glycerophosphate attached to a glycolipid anchor, typically a diglucosyldiacylglycerol [110]. The poly-glycerophosphate repeating unit has D-Alanine or monosaccharides substitutions in both WTA and LTA, being glucose the major saccharide found in *P. cerevisiae* WTAs. [110, 149]. Figure 5 shows a schematic representation of both WTA (a) and LTA (b) structures. These TA are very important to cell wall functionality playing different and important roles. Due to their anionic character and also their distribution on cell wall, the enzymes function properly, being responsible for generating a pH gradient along the cell wall. In addition, TA are responsible for maintaining cell morphology and controlling autolysins. The recognition of bacteriophages and the interaction with the host immune system as well as their colonization is part of the role played by TA. The D-Alanine substitutions have an important impact on cell wall functionality because they modify the TAs charge [110, 147].



(a) WTA



(b) LTA

Figure 5: Both WTA and LTA structures represented schematically.

Adapted from Chapot-Chartier and Kulakauskas 2014 [110].

PG, also called murein [147], is a polymer composed by a glycan chain of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) linked alternately via β -1,4 linkages. A pentapeptide made of L- and D-aminoacids is attached on the lactyl group of MurNAc [110, 147]. In LAB, this peptide is composed by L-Ala-γ-D-Glu-X-D-Ala, where the X represent a di-amino acid which differs from species to species [110]. In *P. damnosus*, the di-amino acid is L-Lysine and D-Aspartate, and this latter is used in the interpeptide cross-bridge [110, 132, 137, 138], as represented in Figure 6.

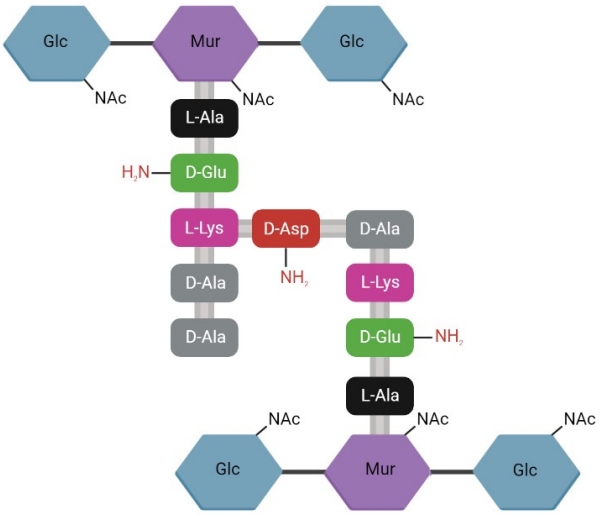


Figure 6: PG scheme of *P. damnosus*.

L-Ala = L-Alanine; D-Glu = D-Glutamate; L-Lys = L-Lysine; D-Ala = D-Alanine; D-Asp = D-Aspartate
Adapted from Chapot-Chartier and Kulakauskas 2014 [110]. Created with BioRender.com.

As mentioned before, *P. damnosus* produces EPSs which has important industrial applications [129, 143]. This cell wall compound is composed by sugar (or derivatives) monosaccharide residues and it can be associated to cell surface or secreted into the extracellular environment [110, 143, 150]. EPSs are classified into homopolysaccharides, if the polymer is formed of identical monosaccharides; and heteropolysaccharides, if two or more types of monosaccharides are present [143, 150]. The EPSs of *P. damnosus* are homopolysaccharides composed by a trisaccharide repeating unit of D-glucose, forming a β-glucan structure [151–153], as shown in Figure 7:

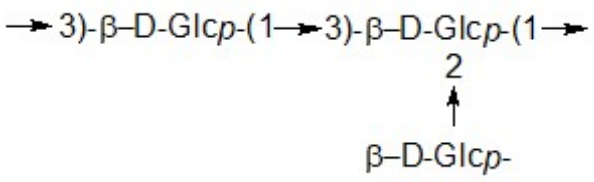


Figure 7: EPS structure of *P. damnosus*.

Glc = Glucose
Adapted from Dueñas-Chasco et al. 1997 [151].

2.9 Brettanomyces bruxellensis

The fourth phase of the lambic beer production process, after eight months of fermentation, is characterized by the presence of *Brettanomyces* yeasts [10]. *B. bruxellensis* is facultatively anaerobic [10], ethanol and sulfur dioxide tolerant [154, 155], resistant to cycloheximide [10] and produces ethanol by alcoholic fermentation [156]. *Dekkera* is the teleomorphic form (sexual) of the anamorphic (asexual) genus *Brettanomyces*. Presently, this genus incorporates six species, namely the anamorphs *Brettanomyces anomalus*, *B. bruxellensis*, *Brettanomyces custersianus*, *Brettanomyces naardenensis*, *Brettanomyces nanus*, and *Brettanomyces acidodurans* [10, 157]. The first four species also have teleomorphs form, named as *Dekkera anomala* [10, 158], *Dekkera bruxellensis* [10, 159], *Dekkera custersiana* [160] and *Dekkera naardenensis* [161]. In 1996, Cai et al. constructed a phylogenetic tree based on 18S RNA, using the neighbor-joining method, where it can be verified that *Candida albicans* (*C. albicans*), *Candida tropicalis* (*C. tropicalis*), *Kluyveromyces lactis* (*K. lactis*), *Candida glabrata* (*C. glabrata*) and *Saccharomyces cerevisiae* (*S. cerevisiae*) are phylogenetically close species to *B. bruxellensis*.

Currently, there are five *B. bruxellensis* genomes published from different strains: UCD 2041 [163], AWRI1613 [164], NRRL Y-12961 [165], AWRI1499 [166] and CBS 2499 [167]. Nevertheless, there are twelve genomes assembly and annotation reports published in NCBI [168]. The genomes size ranges from 11.77 Mbp to 26.99 Mbp and the GC percentage is between 39.80 and 41.60 [163–167].

B. bruxellensis was first taxonomically named and described in 1921 by Kufferath and Van Laer, in a study about lambic beer. This yeast plays an important role during the final maturation stage of lambic beer process production, being responsible for its typical aroma and, attenuation and over-attenuation [10]. It is possible to obtain a better over-attenuation if LAB is also present which indicates a synergistic effect [10]. *B. bruxellensis* produces several volatile phenolic compounds which is undesired in wine, but crucial in lambic beer [10, 169, 170]. Subsequently, this yeast has been isolated from other alcoholic fermentations such as wine and tequila, from soft drinks and even dairy products [10]. *B. bruxellensis* metabolism is nutritionally efficient, making it capable of surviving to the severe conditions during lambic beer production process [10]. In the maturation phase, a biofilm is formed, preventing oxygen influx and thereby inhibiting beer oxidation, which is produced by *B. bruxellensis* and other yeast such as *Candida* and *Pichia* [10]. The secondary fermentation of lambic beer, which starts in the maturation phase, is facilitated by *B. bruxellensis*, contributing to the flavour profile [10]. *P. damnosus* and *B. bruxellensis* metabolise the dextrins that were not consumed by *Saccharomyces* yeasts, taking part in the development of the acidity and flavour of the lambic beer [10].

The optimal growth temperature of *B. bruxellensis* ranges from 25 to 28ºC [10], and it is resistant at low pH values such as between 1.5 and 2 [156]. Is resistant to large variations of pH and temperature, hence, under oxygen limited conditions, it may also have an energy efficient metabolism [154]. *B. bruxellensis* express a negative Pasteur effect, also named Custer effect, inhibiting the alcoholic fermentation under anaerobic conditions and stimulating it in presence of oxygen [10]. Thereby, when a *B. bruxellensis* yeast is in an aerobic environment

and is introduced in an anaerobic environment, glycolysis is temporarily stopped, introducing a transitional lag phase before the slow restart of glucose fermentation [10]. The Custer effect mechanism is not yet unveiled, but there is a consensus hypothesis [10]. Acetaldehyde is continually converted, in an irreversible way, into acetic acid, using a NAD^+-dependent aldehyde dehydrogenase, and subsequently continuous NAD^+ drainage occurs [10]. Due to the lack of NAD^+, glycolysis is forced to stop [10]. Anaerobically, *B. bruxellensis* cannot restore the redox imbalance by re-oxidation of NADH to NAD^+ via the glycerol pathway, since the glycerol 3-phosphate phosphatase activity is limited or none, resulting in lag phase [10]. The yeast circumvented this phase by the recuperation of NAD^+ via the production of other reduced metabolites such as ammonium and ethyl derivates (e.g. 4-ethylphenol) [10, 169, 170]. Furthermore, *B. bruxellensis* is a Crabtree-positive yeast, producing ethanol through carbohydrates degradation, in the presence of high carbohydrate conditions and under aerobic conditions [10]. Acetic acid is another product produce by *B. bruxellensis* in an aerobic environment [10]. Nevertheless, under oxygen limitation *B. bruxellensis* favours alcoholic fermentation, while in aerobic conditions, oxidative metabolism seems to be stimulated and consequently the acetic acid production levels increase closer to those of ethanol [156]. *B. bruxellensis* utilize several carbon sources [171] which include glucose, fructose, maltose and mannose [154]. As nitrogen sources, this *Brettanomyces* species uptake ammonium, proline, arginine, other amino acids, nitrate and nitrite [154]. *B. bruxellensis* is considered a yeast with low nutritional requirements as it can appear when nutrients are scarce [154]. Unlike other common yeasts, *B. bruxellensis* is capable to survive and be well adapted to post fermentation conditions [154].

The yeast cell wall is an essential structure for the maintenance of cell integrity and viability, protecting the cell against osmotic and mechanical stress, and controlling its permeability [172]. There is not much information available about the *B. bruxellensis* cell wall. Therefore, it was presumed based on studies about the cell wall of other yeasts such as *S. cerevisiae* [173, 174] and *C. albicans* [172]. The yeast cell wall structure is composed by different layers where the most conserved structure is the inner layer. Major components of these cell walls are chitin, β-1,3-glucan, β-1,6-glucan and mannan [172, 173]. Chitin represents 1% to 3% of wall mass and it is synthesized from GlcNAc with a polymerization degree of 120 [172, 173, 175]. Located in the inner layer of the cell wall, is important to the cell stability [176] and it can be found free, linked to β-1,3-glucan mainly present in the neck between mother and daughter cell and, in lesser amounts, it occurs in lateral walls attached to β-1,6-glucan, which are bounded to mannan and β-1,3-glucan[172, 173].

Representing 50% to 60% of cell wall dry weight, glucans are the most important polysaccharide of this yeast structure. β-1,3-glucan is the main glucan polymer accounting for about 80% of all glucans whereas β-1,6-glucan only accounts for about 20% [172, 173]. The latter plays an important role as a linker of the different cell wall components, contributing to its stability [175, 177], and the β-1,3-glucan although its biological function is still unknown, it confers some elasticity to the cell wall structure [174, 177]. They have different degrees of polymerization where β-1,3-glucan has 1500 and β-1,6-glucan has 150 [174, 175].

Finally, mannan represents 40% of the dry weight wall mass and has different important biological functions such as signal transmission, cellular protection, maintenance of the cell shape, reorganization of cell wall components and is also responsible for the adhesion process [172, 175]. Figure 8 shows a scheme of the cell wall structure of the yeast *B. bruxellensis*.

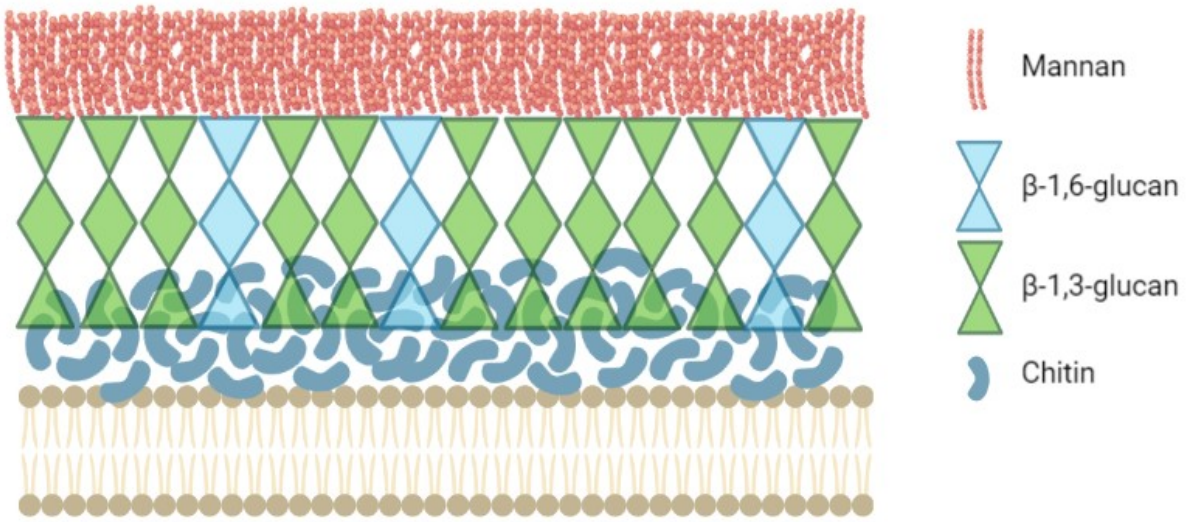


Figure 8: Scheme of the cell wall structure of *Brettanomyces bruxellensis*.

Adapted from Garcia-Rubio et al. 2020 [172]. Created with BioRender.com.

2.10 Microbial communities

In nature, microorganisms usually appear in communities [20]. Understanding how microbes interact and how this interaction influences the community dynamics and robustness, as well as the individual genotype affects ecosystems properties, is a huge challenge [20]. With the advance of meta-omics methods, it becomes easier to characterize, in a detailed way, microbial communities [20]. The use of GSMMs to study microbial communities enhances the development of new applications [34], such as probiotic design for restoring a diseased intestinal microbiota [178]. Moreover, it enables a better understanding of microbial communities, creating new insights about their development, interaction and evolution [34]. The application of GSMMs to communities is more complex because it is considered not only all metabolic reactions of each microorganism, but also the exchange of metabolites between species and/or biomass abundance of each microorganism [20]. Additionally, growth nutrient requirements are specific for each species, and in a community can occur metabolic cross-feeding, nutrient competition, or nutrients uptake from the environment [20]. Selective pressures in single organisms can influence the metabolic interactions between species which can force the shape of microbial communities' structure [20].

2.10.1 Microbial community models

Microbial communities are very important for many ecosystems, for health and for several different industries. Therefore, is necessary to reconstruct predictive models of microbial communities in order to understand its structure and function [34]. To study microbial communities, methods including evolutionary models, thermodynamically based models, Lotka-Volterra models, non-linear regression models and stoichiometric modelling (such as GSMM) have been used [179]. As mentioned above, GSMMs have been successfully implemented to study of single organisms and consequently, methods and tools have been developed over the years to be applied in complex microbial communities. In 2007, Stolyar et al. published the first community model of the mutualistic relationship between *Desulfovibrio vulgaris* and *Methanococcus maripaludis* [34]. Identically to single-organism models, community modelling is based on top-down or bottom-up approaches, which are defined as Population-level models (PLMs) and Individual-based models (IBMs) respectively [179]. PLMs instead of directly describing the microorganisms in a community, describe the changes at the population level and therefore they are classified as top-down. IBMs are considered bottom-up because they describe the individual behavior of each microorganism present in order to predict the behavior of the community [181]. In general, PLMs should be used when homogeneous environmental conditions are considered and IBMs should be applied to heterogeneous environments [179].

Population-level models

PLMs are traditionally used in microbial systems modelling. These models use a mass action approach generating simple models of interaction between different species, being typically applied to reveal general explanations about the communities [181]. Thus, PLMs can be used as a first step to study in detail a complex system or to study, for example, the tendency to create oscillatory population dynamics in predator-prey systems. Conversely, is not advisable for ecosystem management because it does not adequately predict specific populations in specific ecosystems [181]. PLMs are based on ordinary differential equations, when applied to spatially homogeneous environments, or partial differential equations for spatially structured environments. This type of modelling considers the dynamics of resources or assumes a density that depends on the growth rate, representing that at a higher population density the resources deplete faster. PLM has provided important knowledge on microbial ecology, for example in the study of biofilms, where it is possible to simulate density distributions over time [181].

PLMs are able to simulate the system dynamics, such as changes in the microbial distribution of a certain population. Although to verify such changes at the individual level of the microorganisms present in the community is not possible [181].

Individual-based models

Unlike PLMs, IBMs simulate the behavior of microorganisms on an individual level, as well as the changes they cause in the environment and how they respond to it. In this way, this

approach can overcome the limitations of PLMs. Individual-based modeling assumes that the action of each microorganism determines the properties of the population or community, i.e., any change at the population level arises from all interactions between individuals and the environment. IBMs describe the interactions, activities and properties of each microorganism and may therefore include:

- As interactions: competition, synergy or parasite among community microorganisms;
- As activities: substrate absorption rate, secretion rate or synthesis of new biomass;
- And as properties: biomass, size or physiological state of the microbe.

These characteristics should be classified as continuous and based on equations, such as growth kinetic models, or discrete and based on rules, such as describing cell division. IBMs have advantages over PLMs when the objective is to make individual simulations, to simulate specific interactions or adaptive behavior, which are important features in microbial ecology. Although to evaluate behavior at the population level, PLMs are more advisable even because they are typically simpler models.

2.10.2 Microbial community models construction

Over the years, methods to reconstruct GSMMs of microbial communities have been developed. These methods can be divided into constraint-based methods and topological methods, embracing several different approaches. The main approaches will be described below. First, the approaches belonging to the constraint-based methods will be discussed, starting with compartmentalization. Compartmentalization was the first approach developed for modelling microbial communities and is still the most used. This framework is a logical extension of the compartmentalization method used for eukaryotic GSMM. It combines multiple species-level GSMM generating a large stoichiometric matrix with a defined compartment for each species, transport reactions for the metabolic fluxes between species and normally it is inserted in a major compartment as a representation of the microbial community environment [34, 179, 182]. FBA is used to estimate growth rate and metabolite fluxes, by defining a linear objective function based on the weighted combination of the biomass equation for each species, which relies on experimental data. This framework allows the study of growth limits as well as of metabolite fluxes according to the metabolic network structure [34]. For commensalism or mutualism relationships between species it allows the computational design of environment and media conditions [34, 179, 180, 183] and it can also be used to understand the impact of a microbiotic on host metabolism [34, 179, 180, 184, 185]. As mentioned above, the first community GSMM used this approach and it depicted a mutalistic interaction between *Desulfovibrio vulgaris* and *Methanococcus maripaludis* [34, 180]. It was also implemented in others studies such as to understand the host-pathogen relationship between *Mycobacterium tuberculosis* and an alveolar macrophage [34, 179, 184] and to predict the competitive and cooperative potential among

several pairs of species [8, 34]. The compartmentalization framework may limit several analyses that could be performed. Due to steady-state constraints in FBA, it is forced an assumption of balanced growth, therefore metabolic accumulation in a certain environment is not take into account. Another limitation is to assume that each species in the community grows optimally, because the objective function defined in the FBA problem is the combination of the objective functions of each species. Finally, the quantity of species in the community is assumed to be fixed, not representing changes in the microorganism abundance in response to their interactions [34].

Another approach is based on community objectives, which is an approach implemented in the OptCom tool and it is considered as an extension of the compartmentalization strategy, but it uses an objective function defined at the community level [34, 179, 186]. Throught a nested, bi-level optimization, it is able to simulate several metabolic interactions such as mutualism, synergism, commensalism, parasitism and competition [34, 179]. A main advantage of this approach is the ability to verify and interpret trade-offs between the microorganisms and community objectives. Overall, OptCom allows the exploration of several types of communities and their interactions, respecting the trade-offs between objectives [34, 186]. This tool was used in different studies, including the simulation of two gut bacteria interactions - one an acetate producer and other an acetate consumer and butyrate producer [187]. FVA was used in order to obtain a range of flux values for shared metabolites (acetate, in this study case) and subsequently to explore these possibilities [34, 187]. OptCom is expensive at computational-level and due to the nonlinear constraints, it is not appropriate to certain optimizations solvers. This approach is sensitive to the optimization functions and flux constraints that are defined by the user. Therefore it is not advisable the use of this tool for poorly defined communities with metabolic interactions not well known [34, 186].

Dynamic analysis is also an approach which belongs to the constraint-based methods. This dynamic approach relies on dFBA where the fluxes of metabolite consumption/production rates are integrated over time, allowing the accumulation or depletion of compounds, in contrast to FBA which assumes a steady-state condition. dFBA allows to simulate changes in metabolite consumption and production over time, as well as shifts in biomass and in metabolism as response to environmental changes. To implement dFBA is necessary kinetic parameters, specially related to uptake rates of certain metabolites, such as oxygen or glycose, which are considered limiting metabolites. Thus, this method provides an entire time course whereby the adaptation to modifications at conditions level and nutrients availability is described, instead of a static snapshot of metabolites states as it is obtained by FBA [34, 179, 188, 189]. To perform a dynamic analysis, an extended version of OptCom has been developed [186]. Thereby a community objective function must be defined and the fluxes bounds of interspecies interactions must also be set as well as the proper known kinetic parameters [34, 179]. dFBA is often used to optimize community to produce a specific compound [189, 190] or to understand and study the behaviour of a biofilm [191]. It can also be used to estimate biomass production or shifts in nutrient concentrations due to, for example, metabolites diffusion between compartments [192], or to study the emergent biosynthetic capacity of a community [193]. This approach was also

implemented to explore the interactions between *Escherichia coli* strains [194] and to simulate the response of a two bacteria community to nutrient modulation [195]. dFBA should be used to model small communities with well-defined characteristics. Being an underexplored approach, it allows to obtain information about many communities and can take into account their spatial dynamics. One of the advantages is that it does not require as many initial assumptions as FBA although is more computational demanding due to increases in computations over time. Moreover, is based on kinetic parameters, information that is more difficult to obtain [34, 179].

Lastly, there is a constraint-based method approach named enzyme soup which is a completely different approach from the others presented. As the name suggests, the microbial community is treated as an enzyme soup [179], in which reactions are not identified to the different microorganisms. Therefore, no boundary concept between species is assumed. In an enzyme soup community GSMM, the enzymes are annotated according to a meta-omic dataset and the associated reactions are clustered, forming a single set [34, 179, 196]. Thus, these reactions are not divided by species. The biomass equation has a general formulation that represents the biomass of the whole community, since the various components are shared by the community [179]. In the first studies developed with this method, the constraints associated with reactions stoichiometry were not considered and therefore the differences between networks reconstructed from different metagenomic data were examined. Recently, methods of analysis based on constraints, such as FBA, have been used to predict biomass production and substrate consumption [34, 197]. Due to its assumptions, this method focuses on exploiting the metabolic potential of the microbial community and not on interactions between species [34, 179, 197]. Thus, with the support of experimental meta-omic data, the application of this approach to large and complex communities is quite easy [34, 197]. In 2015, Tobalina et al. [198] applied this approach to the study of a naphthalene-degrading community [34, 179]. In order to obtain more detailed and specific solutions within defined boundaries, other approaches, such as compartmentalization, can be used. In 2009, Taffs et al. [199] used different modelling approaches, including compartmentalization and enzyme soup, to study the interaction between three microbial guilds. The enzymatic soup approach is advised when *a priori* knowledge of a community is limited [34, 199]. It must be considered that using this method, the precision decreases due to the absence of boundaries between microorganisms. Therefore, the main advantage is the lack knowledge about the community needed, being well applicable to communities that are poorly studied or poorly understood [34].

The reconstruction of community-level GSMMs using the topological method uses approaches such as graph-based, network expansion and comparative. In order to identify patterns of competition and cooperation between species, graph-based methods have been developed [34, 179, 200, 201]. A graph of all substrate-product pairs is generated from the stoichiometric matrix of the GSMM. Thus, seed sets of each GSMM are identified at species-level and species-specific limits are maintained. The graph shows connections between metabolites with direction from substrates to products but does not contain information about stoichiometry. Metabolites which are consumed but not produced are called seed sets and they are represented by nodes with an in-degree/out-degree rate equal to zero and therefore, must be

supplied to the metabolic network [34, 179, 200]. By evaluating this set of seeds for several species it is possible to unveil the metabolic basis and then estimate the competition or co-operation potential between species [201]. Although it shows that species tend to coexist in nature with mutual competitors, the graph-based method is not useful for predicting fluxes because it ignores the stoichiometry of reactions. Taking this into account, it is advised for more generalised analyses of network similarity and for draft models with low accuracy for FBA problems or similar analyses [34, 179].

On the other hand, network expansion is an agglomerative algorithm that exploits the metabolic potential of any set of reactions [34, 179, 202]. Initially the algorithm receives as input a set of metabolites designated as the environment and then, the reactions that can use these metabolites as substrates are added to the metabolic network. The network is interactively expanded: as the products of the reactions added in the previous step become available, new reactions that have part of the metabolites' set as substrate present in the network are added [34, 203]. In order to extend this approach to a microbial community-level analysis, the reaction sets of the two microorganisms are considered and it is assumed that any intermediate metabolites can be shared. Thus, the network expansion is done by conjugating the reactions set of each microorganism. This approach is used to identify emergent properties such as biosynthetic capabilities, of many microorganisms' pairs. For example, this approach is a good choice to study microbes that together produce certain metabolites that are not synthesised by any of the parent species. This approach presents a great advantage over the graph-based method since it preserves the reactions stoichiometric information. Moreover, it is advisable to identify the compounds produced from a given set of substrates and consequently using a specific set of reactions. Furthermore, it is also advised to identify the essential reactions of a species or community in a given environment and thus understand which reactions are redundant in the metabolic network [34, 179, 203, 204].

The comparative approach focuses on identifying functional differences between GSMMs [34]. This approach has already been applied in some studies, as in the comparative analysis between two species of the genus *Burkholderia* [205] and between two strains of the species *Lactobacillus casei* [206]. These comparative analyses facilitate the identification of functional properties of species present in large communities and thus understand redundancy as well as the differences between the metabolic capacity of the microorganisms present in the community [34]. For example, in the first study mentioned, exclusive reactions associated with each species of the genus *Brukholderia* were revealed, as well as metabolic differences associated with the capacity to produce the virulence factor [205]. In the second study, the result obtained was of great industrial importance and it showed functional differences between strains of *Lactobacillus casei* [206]. The comparative approach ignores the interactions between microorganisms in a community and therefore the only objective is to find the functional differences between different GSMMs. Therefore, this approach is not advisable for studies based on the interaction between species [34].

In addition, to build metabolic community models, RedCom approach allows to build a reduced model of a balanced microbial community [207]. Based on the compartmentalization

model approach [34, 179, 207], has shown a higher predictive power compared to the full community models [207]. As mentioned in section 2.6, ReFramed also allows the community GSMM assembly and simulation. SteadyCom can be used in ReFramed tool [92, 98].

Simulation methods specific for microbial communities have also been developed [20]. Community Flux Balance Analysis (cFBA) is a simulation method that was translated of FBA for single species to microbial communities and requires a few constraints which are specific to the community. In steady-state and with an exponential growth rate that needs to be fixed, cFBA predicts the values of biomass, the distribution of the intra- and extracellular fluxes and metabolic exchanges of all the microorganisms present in the community. Thereby, this method predicts, under specific environmental conditions, the optimal flux distribution, growth rate as well as the abundance of all organisms involved, the exchange fluxes between organisms and the environment where is the microbial community [20]. Another simulation method is dynamic Flux Balance Analysis (dFBA) which is also an extension of FBA. It allows the simulation, analysis and optimization of microorganism models in several environmental contexts. This method has been extended to the study of microbial communities where each microbe has the GSMM available. The GSMM are incorporated in a dynamic model, providing predictions of dynamic changes in metabolism, interaction between species and metabolites concentrations [188, 208, 209].

Dynamic Multispecies Metabolic Modeling (DyMMM) is a tool that is performing in MatLab®and it is based on dFBA [195, 210]. Dynamic Flux Balance Analysis laboratory (DFBAlab) is a simulation tool similar to DyMMM, also based in dFBA. In order to obtain unique exchange fluxes, it uses a feasibility problem of linear program to prevent infeasible linear programs while the simulation is running [210]. Other tool based on dFBA is μbialSim that allows the simulation of microbiomes under batch and chemostat conditions. Due to the numerical integration scheme which does not require additional objectives or compounds' pre-allocation, it provides simulations with high numerical accuracy [211]. OptCom simulation tool allows the simulation of microbial communities with any number of microorganisms. Based on FBA, it relies on multiple objective structures in order to be able to understand the individual and community interactions and behavior [186]. In turn, SteadyCom is an optimization framework to microbial communities that predicts metabolic flux distributions in a steady state. Is compatible with FVA and can predict the microbial community composition in a certain environment [98]. MICOM is a Python package which use strategies based on COBRApy Python package. This tool integrates GSMMs with information about import flux bounds and abundance values estimated from genetic information. Assuming that growth rates and relative abundances in the microbial community are at steady state, MICOM predicts *in silico* exchanges between microorganisms and their environment, as well as ecological interactions within communities. Other features are the prediction and formulation of assumptions about the interference of environmental conditions on metabolic interactions. MICOM mathematical formulation is similar to OptCom and SteadyCom frameworks' formulation [212]. An example of an optimization approaches is OptDeg which requires two objective functions [213]. Finally, Species Metabolic Interaction Analysis (SMETANA) is a MILP method which allows the prediction of community

resource competition and metabolic cross-feeding by assembling a microbial community model from the single-species models [214].

Finally, all modelling approaches have their advantages and disadvantages, consequently the best option is to use a combination of various modelling and experimental approaches [34, 179].

To unveil the flavour profile of lambic beer it is necessary to study the microorganisms present in the lambic beer production process, which requires a mixed yeast-bacteria microbial community [10], and it is also important to study the community interactions, in order to understand the metabolites which confer flavour and how the interaction between species influence the taste of the beer.

3

METHODOLOGY

The Methodology used in the reconstruction of GSMMs is described in the current chapter.

3.1 Tools

The GSMMs were reconstructed using *merlin*, which is an open-source Java™computational tool. Besides having an user-friendly interface, *merlin* contains several tools that automate some reconstruction steps, thus allowing a semi-automated reconstruction [5, 55].

To test the synthesis of biomass precursors a *merlin*'s feature named Biological networks constraint-based *In Silico* Optimization (BioISO) [215] was used. The tool used to validate both GSMMs was MEWpy [93] using ReFramed [92] as simulation environments, namely to assemble the community model, and COBRApy [94] for pFBA simulations.

SMETANA [214] was used to predict community interactions and SteadyCom to predict the abundance of which organism under certain environmental conditions [98].

3.2 Genome files

The RefSeq assembly genome files of *Pediococcus damnosus* LMG 28219, with NCBI assembly accession number ASM96287v1, were retrieved automatically using a *merlin* framework that needs only the organism NCBI taxonomy ID (1448143).

For the yeast *Brettanomyces bruxellensis* with NCBI taxonomy ID of 5007, the assembly genome files from GenBank where used, which has the NCBI assembly accession number of dekkera_v2.

3.3 Genome annotation

The first stage of the GSMMs reconstruction process is genome annotation. Providing the NCBI Taxonomy Identifier, *merlin* can import genome data files from the NCBI database. High quality models require high quality homology results, and *merlin* stores and provides a friendly interface facilitating annotation review. Therefore, a homology search was performed for each coding sequence encountered in the genome using BLAST [44] with the default parameters [5, 55].

merlin displays an enzymes board with data correspondent to organisms' genome functional annotation. The score calculated by *merlin*, which considers both hits frequency and source organism taxonomy, was automatically associated to each gene, and also a status, a name, a product and an E.C. number, if available. Another *merlin*'s feature named automatic workflow was used to obtain more reliable genome functional annotation. This tool annotates and confers a confidence label to each gene with at least one homologous gene associated with an E.C. number based on hits revision status and source organisms. The user selects an ordered list of organisms which usually are phylogenetically close organisms. This feature starts by searching for the first organism chosen in the BLAST results, creating an interactive process: if a match is found, the candidate gene is annotated according to hit's information; otherwise the next organism of the list is considered. Since BLAST results are organized according to databases and reviewed genes were found to be more reliable than unreviewed genes, the automatic workflow was performed for Swiss-Prot results and then for TrEMBL [5, 55].

Tables 4 and 5 represent the data used in the automatic workflow tool performed against Swiss-Prot and TrEMBL, for each organism. The organisms' list was created based on phylogenetic proximity and the availability of information on the organism. The first entry is the species used in this work, and the second is its genus. In sections 2.8.1 and 2.9, the phylogenetically closest species and genera were discussed.

Table 4: Automatic workflow data for *P. damnosus* GSMM.

Organism	Confidence label	e-Value
Pediococcus damnosus	A	1.00E-20
genus *Pediococcus*	B	1.00E-20
Lactobacillus palntarum	C	1.00E-20
genus *Lactobacillus*	D	1.00E-30
Lactococcus lactis	E	1.00E-30
Lactococcus lactis subsp. Cremosis	F	1.00E-30
genus *Lactococcus*	G	1.00E-30
Streptococccus thermophilus	H	1.00E-40
genus *Streptococcus*	I	1.00E-40

Therefore, after integrating the enzymes functional annotation, a draft GSMM has been generated.

3.4 Metabolic network reconstruction

The second stage of the GSMMs reconstruction process is the assembling the metabolic network. The methods used are described in the following sections.

Table 5: Automatic workflow data for *B. bruxellensis* GSMM.

Organism	Confidence label	e-Value
Brettanomyces bruxellensis	A	1.00E-20
genus *Brettanomyces*	B	1.00E-20
genus *Debaryomyces*	C	1.00E-20
genus *Candida*	D	1.00E-30
genus *Kluyveromyces*	E	1.00E-30
genus *Saccharomyces*	F	1.00E-30
genus *Yarrowia*	G	1.00E-30
genus *Schizosaccharomyces*	H	1.00E-40
genus *Cryptococcus*	I	1.00E-40

3.4.1 Metabolic data

First, it is necessary to collect metabolic data including metabolites, enzymes, reactions and pathways information. *merlin* retrieved this data from KEGG. In addition, spontaneous reactions were retrieved and automatically integrated into the model, which was later complemented by integrating enzyme and transporter annotations.

3.4.2 Transport reactions and exchanges reactions

The Transport Systems Tracker (TranSyT) is a tool implemented in *merlin* that generates and integrates transport reactions, using TCDB as the primary information source. This tool also retrieves data from MetaCyc and KEGG to complete information of the entries found in TCDB [216]. TranSyT identifies genes that encode transport systems as well the corresponding transported compounds. Therefore, transport reactions are automatically created with their Gene-Protein-Reaction (GPR) associations [55, 216]. Besides these reactions, other transport reactions were added to the model, with the following reasoning:

- If a metabolite belongs to the growth medium, but there is no uptake reaction associated to the metabolite in the model;
- If there is experimental or literature data that validates the production of a specific metabolite, but there are no excretion reactions associated;
- If a transport system is described in the literature, for example, transport reactions between mitochondria compartment and cytoplasm compartment, but TranSyT did not predict and therefore was not included automatically;
- If a metabolite is transported by simple diffusion.

Databases such as TCDB and BiGG were used to retrieve data about mechanisms, substrates and genes associated with manually added transport reactions. Nevertheless, some transport reactions were added without a gene association to ensure model functionality due to lack of information.

Since there are metabolites in the outside compartment, exchange reactions were automatically generated. These exchange reactions are responsible for defining the culture medium and other environmental conditions by manipulating the reactions' bounds.

3.4.3 Gene-Protein-Reactions associations

GPR associations are mandatory to reconstruct a high-quality GSMM as these affect simulations and future applications, such as mutant phenotype predictions. As mentioned in section 2.4.2, these associations are defined according to literature and databases [4, 7]. Hence, *merlin* has an automatic tool that generates GPR rules using KEGG Orthology data. All reactions with enzymes are analysed against KEGG orthologues database. The associated genes were evaluated for each orthologue, and a homology search was performed against the whole genome [5, 55].

3.5 Biomass and Energy requirements

The biomass reaction in GSMMs represents the cell composition of an organism. A high-quality formulation is essential to avoid future problems in the model validation stage. As aforementioned in section 2.4.2, biomass composition should be defined by experimental data from literature. However, in the absence of information about biomass composition, these data can be obtained from the organism's genome (specifically, amino acids, deoxynucleotides and nucleotides) or adapted from studies of phylogenetically close organisms.

For both *P. damnosus* GSMM and *B. bruxellensis* GSMM, the biomass equations represent each macromolecule's relative abundance in a cell, where the reactants are complex macromolecules with a specific stoichiometric coefficient defined as grams of macromolecule per gram of biomass. For both organisms, no information was found in the literature about the composition of their biomass. Therefore, experimental data of closely related organisms was used as a template. For the LAB *P. damnosus*, experimental data from *Lb. plantarum* WCFS1 [217] model and *Lc. lactis* ssp. lactis IL1403 [114] model was used. In turn, for the yeast *B. bruxellensis* *C. tropicalis* [218] model, *C. glabrata* [219], *K. lactis* [220] model and *S. cerevisiae* S288C [221] experimental data was used.

merlin has a feature named *e-biomass equation*, which automatically creates a draft biomass reaction and four other reactions, each representing the synthesis of DNA, RNA, protein and cofactor in a pseudo-pathway labelled "Biomass pathway". Using the genome files, this tool can determine the contents of deoxynucleotide, nucleotide and amino acid for the synthesis of DNA, RNA and protein, respectively [5, 55, 222].

Finally, the precursors of the automatically generated cofactor reaction belong to a predefined set showed to be present in prokaryotic organisms [223], although these were changed according to literature and network metabolic capabilities.

The complex macromolecules included in the biomass reaction of *P. damnosus* are DNA, RNA, protein, lipids, Exopolysaccharide (EPS), Peptidoglycan (PG), Wall teichoic acid (WTA), Lipoteichoic acid (LTA) and cofactors [114, 217]. The precursors of EPS and PG were defined based on literature [110, 114, 132, 137, 138, 147, 217], as well as the lipid, WTA and LTA composition [110, 114, 147, 149, 217]. Lastly, the cofactors and vitamins were set from *merlin* and literature [224].

In *B. bruxellensis*, biomass reactions include DNA, RNA, protein, lipids, carbohydrates and cofactors [218–220]. The composition of lipids and carbohydrates was defined based on literature [172–176]. Cofactors and vitamins compositions were based on *merlin* cofactor template reaction, *C. glabrata* and *K. lactis* experimental data [219, 220].

Although fatty acids are not directly included in the biomass reaction, they are components present in the synthesis of lipids and TA (in the case of the *P. damnosus*). Therefore, in the biomass pathway, a reaction representing the composition of fatty acids was added in each model, where the product of the reaction is a generic fatty acid metabolite. In the bacterium model, the fatty acids were defined based on literature [114, 217, 225]. In the *B. bruxellensis* model, these components were also determined based on literature [219–221]. The molecular weight of the fatty acid metabolite corresponds to the average weight of all fatty acids present in each organism, allowing to determine the molecular weight of both lipids and TA (in the case of the *P. damnosus*).

Due to a lack of information on energy requirements for *P. damnosus* and *B. bruxellensis*, the growth-associated energy and the non-growth-associated energy requirements were recovered from experimental data from *Lb. plantarum* WCFS1 for *P. damnosus* and *S. cerevisiae* iMM904 [221] for *B. bruxellensis*. The growth-associated energy was added to the biomass reaction and the non-growth-associated energy was included in the ATP hydrolysis reaction, defining the lower and upper bounds according to the set value. These energy requirements values can be adjusted in the model validation stage according to experimental data on growth rate.

3.6 Manual Curation

During the manual curation stage, different strategies described in the following sections were employed.

3.6.1 Reactions directionality and balance

After the metabolic data has been loaded, the draft model shows the KEGG reactions with default reversibility. The reversibility was corrected usind *merlin*'s "Correct Reversibility" tool that automatically corrects the reactions' reversibility and directionality. For this work, the

"GramPositive" and "Fungi" options for *P. damnosus* and *B. bruxellensis*, were respectively selected [5, 55]. However, several reactions were manually curated using different databases such as MetaCyc, BRENDA, and BiGG to obtain a feasible and high-quality model.

3.6.2 Growth Medium

A literature search was performed to determine a Chemically Defined Medium (CDM) for both organisms. A minimal growth medium was defined for each organism containing all the necessary compounds. This medium includes carbon, nitrogen, sulfur and phosphorus sources, and all the auxotrophies of each species. The minimal CDM of *P. damnosus* and *B. bruxellensis* is available in support material (Table S3).

As mentioned before, in a metabolic model, the growth medium is defined by constraining the exchange reactions bounds:

- For metabolites present in the CDM, the lower bound was set to -10000, and the upper bound was maintained as 10000;
- For the growth-limiting source (e.g. carbon), the lower bound correspond to an uptake value described in the literature or obtained from experimental data;
- Finally, for anaerobiosis simulation, the lower bound of the oxygen exchange reaction is zero, whereas, in aerobiosis simulations, the lower bound can be defined by a specific uptake value or unconstraining oxygen uptake.

3.6.3 Model troubleshooting

Several gaps compromised the synthesis of multiple biomass precursors. Therefore, a tool implemented in *merlin* named BioISO was used to identify and solve these gaps. Using BioISO, if a biomass precursor is not being synthesized or is not available in the defined medium, a traceback was generated identifying gaps, such as errors in genome annotation, absence of specific enzymatic or transport reactions and incorrect reversibility or direction of reactions. Once the gap is identified, a search in databases (KEGG, BRENDA, MetaCyc and BiGG) and literature was performed to solve the problem. Thus, an iterative process is developed and repeated until all biomass precursors can be synthesized [5, 55, 215].

3.6.4 Gap-Filling

Although the draft GSMM can produce biomass, there are still many gaps and dead-end metabolites, which affects the model feasibility. For this gap-filling performance, a *merlin* tool named "Blocked reactions" was used. This tool evaluates the model connectivity and identifies reactions which consume or produce dead-end metabolites. *merlin*'s "Draw in Browser" feature

opens a selected KEGG pathway map on the web browser with all enzymes, reactions, and dead-end metabolites present in the model highlighted, making it easy to detect gaps [5, 55]. Figure 9 shows the "Pentose phosphate pathway" as example of output in which each E.C. number was highlighted with different colours. More information in O. Dias et al. and Capela et al. [5, 55].

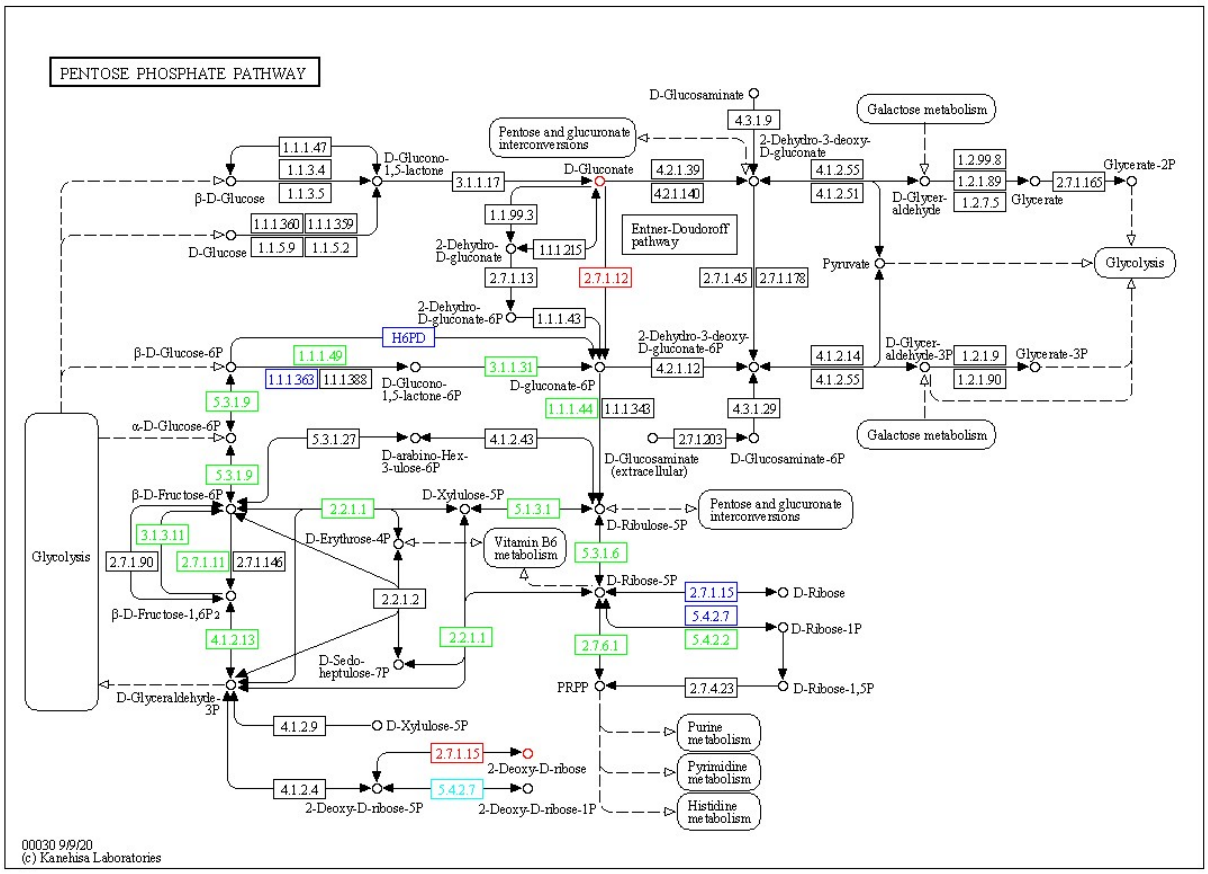


Figure 9: Example of a KEGG pathway map ("Pentose phosphate pathway") coloured by *merlin*'s "Draw in browser" tool.

Metabolic information was collected for each organism, especially for pathways containing blocked reactions and dead-end metabolites. This information was retrieved from the literature and different databases.

3.6.5 Reactions balance

merlin has a tool named "Unbalanced reactions" which allows their identification [5, 55]. Information on reaction stoichiometry metabolites formula was collected and the stoichiometry of unbalanced reactions was corrected, as shown below:

- Metabolites with variable formula: If metabolites are essential to keep model functionality, it was maintained. Example: "Fatty acid" and "Acyl carrier-protein";

- Metabolites without a formula or with a wrong one : In *merlin*'s "Metabolites" board, add or correct the formula based on information retrieved from literature or reviewed databases;
- Polymerization reactions, where $Polymer + Monomer \longrightarrow Polymer$: Remove the polymer from the reactants, $Monomer \longrightarrow Polymer$, and correct the metabolites formulas and reaction stoichiometry;
- Despolymerization reactions, where $Polymer \longrightarrow Polymer + Monomer$: Remove the polymer from the products, $Polymer \longrightarrow Monomer$, and correct the metabolites formulas and reaction stoichiometry;
- Generic reactions : Remove the reaction.

3.6.6 Compartmentalization

The compartments of an organism can be predicted using algorithms such as PSORTb3 [57], WolfPsort [58] or LocTree3 [59], according to the organism's characteristics. Therefore, for the bacterium *P. damnosus*, PSORTb3 [57] was used, while for the yeast *B. bruxellensis* the WoLFPSORT [58] was used. The prediction results were loaded into *merlin*, allowing the integration of the compartments' annotation. Thus, the compartments were assigned to the reactions in the model [5, 55].

3.7 Model validation

Once a feasible model is obtained, the model validation stage begins. Using pFBA, several simulations were performed for each model and the results were compared qualitatively and, if possible, quantitatively with the data described in the literature. The models were considered validated when the predicted values were in agreement with the literature values. The aspects evaluated in the model validation stage for both organisms were the following:

- No growth and no reactions flux when no medium is supplied;
- Growth under aerobic and anaerobic conditions since both organism are facultatively anaerobic;
- Growth in different carbon sources;
- Growth in different nitrogen sources for *B. bruxellensis* model;
- Amino acid auxothrophies.

The bounds of the exchange reactions were changed to simulate these different environmental conditions, and simulations were performed using the MEWpy framework [93].

3.7.1 No uptake

First, each model was not supplemented with any medium to test the model viability. The model should not produce any metabolites with no medium provided and no reactions should have flux. Hence, all exchange reactions were set to zero on the lower bound and 10000 for the upper bound.

3.7.2 Minimal Medium

Providing a minimal defined medium, each model was simulated under aerobic and anaerobic conditions. In all simulations, glucose was the carbon source and the limiting nutient. Under anaerobic conditions, the value of O_2 was restricting to zero, i.e. the lower bound was set to zero. Whereas under aerobic conditions, for *P. damnosus* model, the exchange reaction was left unbounded, but for *B. bruxellensis* model, an uptake value was defined.

P. damnosus

In 1954, Jensen and Seeley described a minimal defined medium for *P. damnosus*. In Table 6, the aerobic and anaerobic environmental conditions for a minimal medium are detailed [141]. Amino acids uptake values were retrieved from *Lc. lactis* ssp. lactis IL1403 [114] model simulations.

B. bruxellensis

In 2016, Von Cosmos and C. G. Edwards described a minimal defined medium for *B. bruxellensis*. In Table 7, the aerobic and anaerobic environmental conditions for a minimal medium are detailed [226]. It was necessary, for anaerobic conditions, to supply medium with some unsaturated fatty acids and sterols present in the biomass equation, based on the procedure described in the literature [220].

3.7.3 Different carbon sources

The models' predictive capabilities using different carbon sources were tested. Under the minimal medium changing only the carbon source (glucose), several simulations were performed and evaluated qualitatively for aerobic and anaerobic conditions.

P. damnosus

In 2002 and 2006, Carr et al. and Hammes and Hertel, respectively, studied the capabilities of *P. damnosus* to ferment different carbon sources [138, 227]. Since *P. damnosus* shows similar behaviour under aerobic and anaerobic conditions, it was assumed that under anaerobic

conditions, the ability to ferment and grow is the same as under aerobic conditions, as shown in Table 8.

B. bruxellensis

In 1998 and later in 2016, M. T. Smith and B. D. Smith and Divol, respectively, evaluated the ability of *B. bruxellensis* to grow on different carbon sources. Table 9 shows the information presented in these studies [154, 228]. According to literature, *B. bruxellensis* can grow on different carbon sources in aerobiosis and anaerobiosis [154, 228].

3.7.4 Different nitrogen sources

M. T. Smith and B. D. Smith and Divol also studied the growth profile of *B. bruxellensis* on different nitrogen sources under both aerobic and anaerobic conditions [154, 228]. The following Table 10 shows different nitrogen sources and the *B. bruxellensis*' ability to grow.

3.7.5 Amino acid auxothrophies

The bacterium *P. damnosus* is auxotrophic for almost all amino acids [141], while for *B. bruxellensis*, auxotrophy has only been reported for three amino acids [226]. Besides testing the auxotrophy of amino acids, the need for other compounds present in the minimal medium was also evaluated. Therefore, several simulations were performed, in which single deletions of the minimal medium components were performed.

After validation, the GSMMs are ready to be applied in scientific studies in context of several areas. In this case, from the two GSMMs, a community GSMM will be reconstructed to analyse the unique flavour of lambic beer further.

3.8 Community model

Firstly, the interactions between the single models of each organism were predicted using the SMETANA [214] tool and then compared with the simulation results. The community model was reconstructed using ReFramed [92], based on the compartmentalization method described in section 2.10.2. Based on the minimum media used for validating each model, the community simulation was performed using COBRApy [94]. Since the aim of the community model reconstruction is the study of lambic beer flavour, the environment in which the *P. damnosus* and *B. bruxellensis* are present was considered. As these microorganisms are mostly active during the acidification and maturation phase of the beer, which takes place in horizontal wooden barrels [102], the simulation was performed in aerobic conditions, with restricted oxygen value. Finally, the abundance of each microorganism in the community was predicted using SteadyCom [98].

Table 6: Environmental conditions used to simulate in a minimal medium for *P. damnosus*.

Compound	Formula	KEGG ID	Aerobic conditions		Anaerobic conditions	
			Lower bound	Upper bound	Lower bound	Upper bound
Adenine	$C_5H_5N_5$	C00147	-10000	10000	-10000	10000
alpha-D-Glucose	$C_6H_{12}O_6$	C00267	-1.9	10000	-1.9	10000
Biotin	$C_{10}H_{16}N_2O_3S$	C00120	-10000	10000	-10000	10000
Fe^{3+}	Fe	C14819	-10000	10000	-10000	10000
Folinic acid	$C_{20}H_{23}N_7O_7$	C03479	-10000	10000	-10000	10000
Glycine	$C_2H_5NO_2$	C00037	-0.2	10000	-0.2	10000
Guanine	$C_5H_5N_5O$	C00242	-10000	10000	-10000	10000
L-Alanine	$C_3H_7NO_2$	C00041	-0.2	10000	-0.2	10000
L-Arginine	$C_6H_{14}N_4O_2$	C00062	-0.2	10000	-0.2	10000
L-Aspartate	$C_4H_7NO_4$	C00049	-0.2	10000	-0.2	10000
L-Cysteine	$C_3H_7NO_2S$	C00097	-0.2	10000	-0.2	10000
L-Glutamate	$C_5H_9NO_4$	C00025	-0.2	10000	-0.2	10000
L-Histidine	$C_6H_9N_3O_2$	C00135	-0.2	10000	-0.2	10000
L-Isoleucine	$C_6H_{13}NO_2$	C00407	-0.2	10000	-0.2	10000
L-Leucine	$C_6H_{13}NO_2$	C00123	-0.2	10000	-0.2	10000
L-Lysine	$C_6H_3N_2O_2$	C00047	-0.2	10000	-0.2	10000
L-Methionine	$C_5H_{11}NO_2S$	C00073	-0.2	10000	-0.2	10000
L-Phenylalanine	$C_9H_{11}NO_2$	C00079	-0.2	10000	-0.2	10000
L-Proline	$C_5H_9NO_2$	C00148	-0.2	10000	-0.2	10000
L-Serine	$C_3H_7NO_3$	C00065	-0.2	10000	-0.2	10000
L-Threonine	$C_4H_9NO_3$	C00188	-0.2	10000	-0.2	10000
L-Tryptophan	$C_{11}H_{12}N_2O_2$	C00078	-0.2	10000	-0.2	10000
L-Tyrosine	$C_9H_{11}NO_3$	C00082	-0.2	10000	-0.2	10000
L-Valine	$C_5H_{11}NO_2$	C00183	-0.2	10000	-0.2	10000
Nicotinate	$C_6H_5NO_2$	C00253	-10000	10000	-10000	10000
Orthophosphate	H_3PO_4	C00009	-10000	10000	-10000	10000
Pantothenate	$C_9H_{17}NO_5$	C00864	-10000	10000	-10000	10000
Pyridoxine	$C_8H_{11}NO_3$	C00314	-10000	10000	-10000	10000
Riboflavin	$C_{17}H_{20}N_4O_6$	C00255	-10000	10000	-10000	10000
Thymine	$C_5H_6N_2O_2$	C00178	-10000	10000	-10000	10000
Uracil	$C_4H_4N_2O_2$	C00106	-10000	10000	-10000	10000
Xanthine	$C_5H_4N_4O_2$	C00385	-10000	10000	-10000	10000
Oxygen	O_2	C00007	-10000	10000	0	10000

Adapted from Jensen and Seeley 1954 [141] and Oliveira et al. 2005 [114].

Table 7: Environmental conditions used to simulate in a minimal medium for *B. bruxellensis*.

Compound	Formula	KEGG ID	Aerobic conditions		Anaerobic conditions	
			Lower bound	Upper bound	Lower bound	Upper bound
(9Z)-Hexadecenoic acid	$C_{16}H_{30}O_2$	C08362	0	10000	-10000	10000
(9Z)-Octadecenoic acid	$C_{18}H_{34}O_2$	C00712	0	10000	-10000	10000
(9Z)-Tetradecenoic acid	$C_{14}H_{26}O_2$	C08322	0	10000	-10000	10000
4-Aminobenzoate	$C_7H_7NO_2$	C00568	-10000	10000	-10000	10000
alpha-D-Glucose	$C_6HC_{12}O_6$	C00267	-2.6	10000	-1.8	10000
Ammonia	NH_3	C00014	-10000	10000	-10000	10000
Biotin	$C_{10}H_{16}N_2O_3S$	C00120	-10000	10000	-10000	10000
Ergosterol	$C_{28}H_{44}O$	C01694	0	10000	-10000	10000
Fe^{2+}	Fe	C14818	-10000	10000	-10000	10000
Folate	$C_{19}H_{19}N_7O_6$	C00504	-10000	10000	-10000	10000
Lanosterol	$C_{30}H_{50}O$	C01724	0	10000	-10000	10000
L-Histidine	$C_6H_9N_3O_2$	C00135	-0.2	10000	-0.2	10000
Linoleate	$C_{18}H_{32}O_2$	C01595	0	10000	-10000	10000
L-Methionine	$C_5H_{11}NO_2S$	C00073	-0.2	10000	-0.2	10000
L-Tryptophan	$C_{11}H_{12}N_2O_2$	C00078	-0.2	10000	-0.2	10000
Molybdate	H_2MoO_4	C06232	-10000	10000	-10000	10000
myo-Inositol	$C_6H_{12}O_6$	C00137	-10000	10000	-10000	10000
Nicotinate	$C_6H_5NO_2$	C00253	-10000	10000	-10000	10000
Orthophosphate	H_3PO_4	C00009	-10000	10000	-10000	10000
Pantothenate	$C_9H_{17}NO_5$	C00864	-10000	10000	-10000	10000
Pyridoxal	$C_8H_9NO_3$	C00250	-10000	10000	-10000	10000
Riboflavin	$C_{17}H_{20}N_4O_6$	C00255	-10000	10000	-10000	10000
Sulfate	H_2SO_4	C00059	-10000	10000	-10000	10000
Thiamine	$C_{12}H_{17}N_4OS$	C00378	-10000	10000	-10000	10000
Zymosterol	$C_{27}H_{44}O$	C05437	0	10000	-10000	10000
Oxygen	O_2	C00007	-0.1	10000	0	10000

Adapted from Von Cosmos and C. G. Edwards 2016 [226] and O. Dias et al. 2014 [220].

Table 8: List of the different carbon sources used to simulate under aerobic and anaerobic conditions, for *P. damnosus*.

Compound	Formula	KEGG ID	Experimental data
alpha,alpha-Trehalose	$C_{12}H_{22}O_{11}$	C01083	+/-
alpha-D-Glucose	$C_6H_{12}O_6$	C00267	+
Cellobiose	$C_{12}H_{22}O_{11}$	C00185	+
D-Fructose	$C_6H_{12}O_6$	C00095	+
D-Galactose	$C_6H_{12}O_6$	C00124	+/-
D-Mannose	$C_6H_{12}O_6$	C00159	+
Sucrose	$C_{12}H_{22}O_{11}$	C00089	+/-
Dextrin	$(C_{12}H_{20}O_{10})_n$	C00721	-
D-Ribose	$C_5H_{10}O_5$	C00121	-
D-Xylose	$C_5H_{10}O_5$	C00181	-
Glycerol	$C_3H_8O_3$	C00116	-
Lactose	$C_{12}H_{22}O_{11}$	C00243	-
L-Rhamnose	$C_6H_{12}O_5$	C00507	-
Starch	$(C_{12}H_{20}O_{10})_n$	C00369	-

Note: (+), the organism grow in the medium; (-), the organism do not grow in the medium; (+/-), some strains grow in the medium, but others do not.
Adapted from Carr et al. 2002 [138] and Hammes and Hertel 2006 [227].

Table 9: List of the different carbon sources used to simulate under aerobic and anaerobic conditions, for *B. bruxellensis*.

Compound	Formula	KEGG ID	Experimental data
alpha,alpha-Trehalose	$C_{12}H_{22}O_{11}$	C01083	+/-
alpha-D-Glucose	$C_6H_{12}O_6$	C00267	+
D-Fructose	$C_6H_{12}O_6$	C00095	+
D-Galactose	$C_6H_{12}O_6$	C00124	+/-
D-Mannose	$C_6H_{12}O_6$	C00159	+
D-Ribose	$C_5H_{10}O_5$	C00121	-
Glycerol	$C_3H_8O_3$	C00116	+/-
Lactose	$C_{12}H_{22}O_{11}$	C00243	-
L-Rhamnose	$C_6H_{12}O_5$	C00507	-
Maltose	$C_{12}H_{22}O_{11}$	C00208	+/-
Raffinose	$C_{18}H_{32}O_{16}$	C00492	+/-
Starch	$(C_{12}H_{20}O_{10})_n$	C00369	-

Note: (+), the organism grow in the medium; (-), the organism do not grow in the medium; (+/-), some strains grow in the medium, but others do not.
Adapted from B. D. Smith and Divol 2016 [154] and M. T. Smith 1998 [228].

Table 10: List of the different nitrogen sources used to simulate under aerobic conditions, for *B. bruxellensis*.

Compound	Formula	KEGG ID	Experimental data	
			Aerobic conditions	Anaerobic conditions
Ammonia	NH_3	C00014	+	+
L-Arginine	$C_6H_{14}N4O_2$	C00062	+	+
L-Cystein	$C_3H_7NO_2S$	C00097	+	-
L-Glutamate	$C_5H_9NO_4$	C00025	+	+
L-Methionine	$C_5H_{11}NO_2S$	C00073	+	+/-
L-Proline	$C_5H_9NO_2$	C00148	+	+/-[*1]
L-Tryptophan	$C_{11}H_{12}N2O_2$	C00078	+	-
Nitrate	HNO_3	C00244	+[*2]	+[*2]
Nitrite	HNO_2	C00088	+	+

Note: (+), the organism grow in the medium; (-), the organism do not grow in the medium; (+/-), some strains grow in the medium, but others do not.
Adapted from B. D. Smith and Divol 2016 [154] and M. T. Smith 1998 [228].
[*1]- Crauwels et al. 2015 [229],Conterno et al. 2006 [230];
[*2]-A. R. Borneman et al. 2014 [231].

4

RESULTS AND DISCUSSION

The results of GSMMs reconstruction are presented and discussed in the current chapter.

4.1 Genome annotation

P. damnosus

The genome annotation identified 840 candidate genes out of 2023 genes encoded in *P. damnosus* LMG 28219 genome, in which 461 genes were annotated with reviewed information and the remaining 379 annotated based on unreviewed annotations. The automatic workflow results are described in Figure 10, reporting gene count for each label split by both Swiss-Prot and TrEMBL databases.

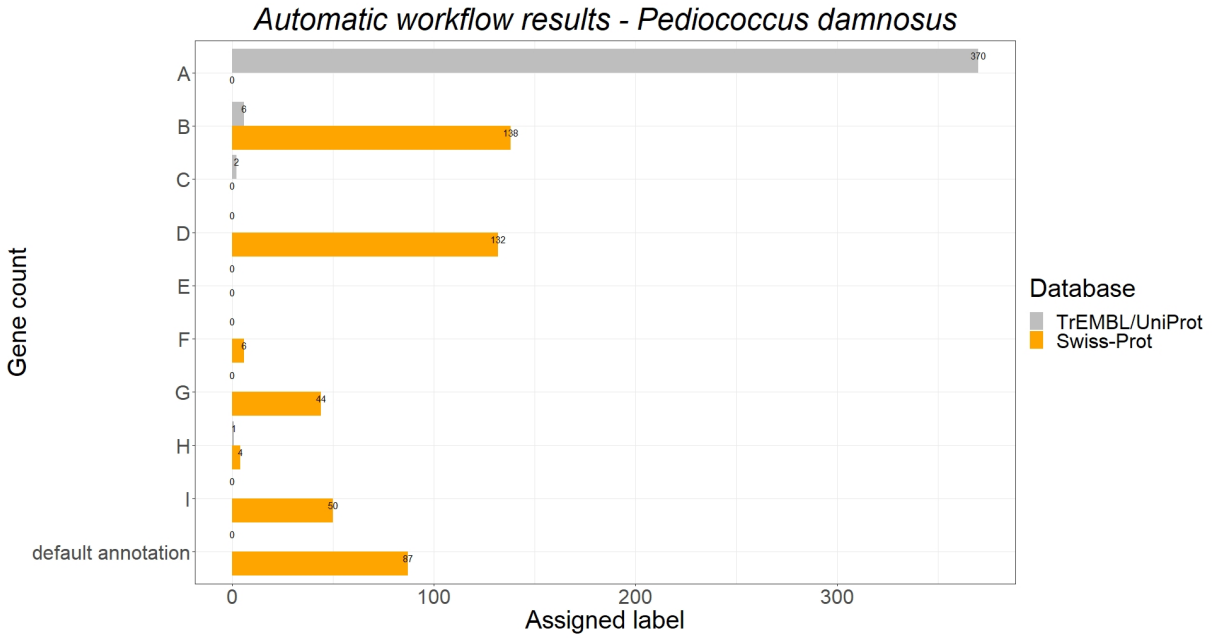


Figure 10: Automatic workflow results for *P. damnosus* GSMM.

B. bruxellensis

The genome annotation performed identifies 1200 candidate genes out of 4304 genes encoded in *B. bruxellensis* genome, in which 1182 genes were annotated with reviewed information, whereas 18 were annotated based on unreviewed annotations. Figure 11 exhibits the automatic workflow results where the gene counts are present for each label and database.

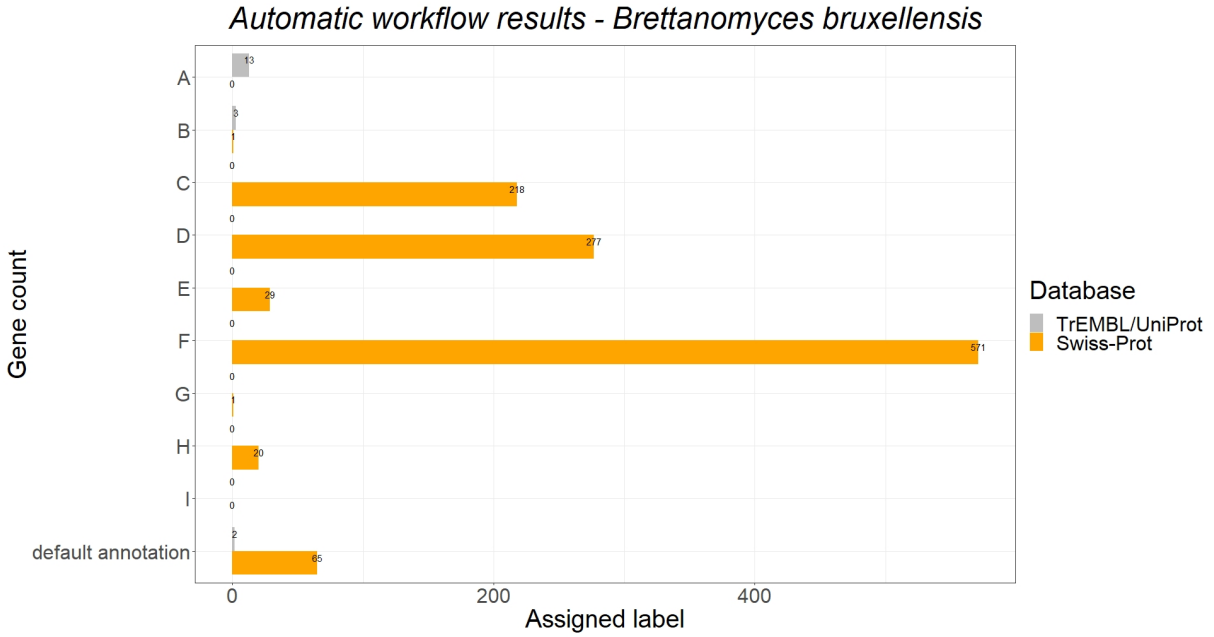


Figure 11: Automatic workflow results for *B. bruxellensis* GSMM.

The *P. damnosus* model contains 514 different E.C. numbers, whereas 748 E.C. numbers are annotated for *B. bruxellensis*. In 2015, Snauwaert et al. analysed the draft genome sequence of *P. damnosus* and identified four Coding sequence (CDS) (AH70_07835, AH70_01405, AH70_01410, and AH70_01415) involved in EPS biosynthesis. However, these CDS are not present in the genome annotation due to assembly file updates [136].

Table 11 shows the enzymes distribution among EC classes in *P. damnosus* and *B. bruxellensis*. In both models, transferases and hydrolases are the most common classes and, therefore, the most represented. With values between 2% and 3%, translocases is the least represented class.

Representing 14 % for *P. damnosus* model and 15% *B. bruxellensis* model, oxidoreductases catalyse the reduction and oxidation of many substrates [232, 233]. Lactate dehydrogenase, alcohol dehydrogenase and acetaldehyde dehydrogenase are enzymes that belong to this class. Genes encoding L-lactate and D-lactate dehydrogenase, essential enzymes for *P. damnosus* metabolism, were identified in the genome since they are involved in producing both lactate isomers, as reported in the literature [132, 137, 138]. Alcohol dehydrogenase and acetaldehyde dehydrogenase catalyse the reactions of ethanol and acetate production, respectively, from acetaldehyde. Therefore, these enzymes are essential for the *B. bruxellensis* metabolism. Genes encoding both enzymes were identified in agreement with the literature [154].

Table 11: Percentage of enzymes present in each model according to the E.C. number classification system.

E.C. number class	Percentage of E.C. numbers (%)	
	P. damnosus	B. bruxellensis
E.C. 1: Oxidoreductases	11.4%	15.8%
E.C. 2: Transferases	33.1%	37.1%
E.C. 3: Hydrolases	34.2%	32.1%
E.C. 4: Lyases	5.3%	4.4%
E.C. 5: Isomerases	5.7%	3.2%
E.C. 6: Ligases	7.6%	5.1%
E.C. 7: Translocases	2.7%	2.3%

Between 33% and 38%, transferases is one of the most representative classes in both models. Phosphate transferring enzymes (E.C. 2.7) represents around 50% of this class in both models. This subclass of enzymes is associated with nucleotide metabolism and carbohydrate phosphorylation [232, 233].

The other class which is also one of the most representative is hydrolases. These enzymes are important for the catabolism of molecules such as amino acids, carbohydrates and lipids. This E.C. class catalyses mostly hydrolysis reactions, acting on different bounds such as ester bounds (E.C. 3.1), sugar bounds (E.C. 3.2) and peptide bounds (E.C. 3.4) [232, 233].

Lyases catalyze the breaking of chemical bounds between a carbon atom and another atom, such as oxygen, without resorting to hydrolysis or oxidation mechanisms. Enzymes from this class play an essential role in glycolysis (e.g. E.C. 4.2.1.11 - Phosphopyruvate hydratase) and, acetic acid and ethanol fermentation (e.g. E.C. 4.1.1.1 -pyruvate decarboxylase) [232, 233].

Isomerases represents 5.7% and 3.2% in *P. damnosus* and *B. bruxellensis* respectively. This class includes epimerases, mutases and racemases [232, 233]. The latter is responsible for the conversion of L- amino acids to their respective D- isomer. To the *P. damnosus* model, D-alanine and D-aspartate are necessary for the biosynthesis of PG, a biomass component.

With a representation of 7.6% for *P. damnosus* and 5.1% for *B. bruxellensis*, ligases have the role of attaching two molecules or part of them, being involved in condensation reactions. In both models, the predominant subclasses are E.C. 6.1 and E.C. 6.3. The E.C. 6.1 groups the enzymes responsible for tRNA (transfer Ribonucleic acid) acylation, the amino acid-tRNA ligases present in the "Aminoacyl-tRNA biosynthesis" pathway. In turn, the ligases that catalyze the formation of carbon-nitrogen bounds are grouped in the subclass E.C. 6.3, playing an essential role in amino acids biosynthesis [232, 233].

Finally, translocases represent values below three for both models. This recent E.C. class is involved in the translocation of ions and molecules across membranes. It is often associated with ATP hydrolysis. The ATPases belong to the translocases class, being only present in the *B.bruxellensis*, as there is no scientific evidence of the presence of ATPase in *P. damnosus*. The

low value of the specific growth rate for this bacterium, mentioned in section 4.5.2, could also be justified by the absence of the ATPase mechanism [234].

4.2 Metabolic network reconstruction

The assembly of the metabolic network starts with collecting metabolite data such as metabolites, enzymes, reactions, and pathways information based on genome annotation. This process was automated by *merlin* and the results are shown in Table 12.

Table 12: Draft GSMMs details.

Draft GSMMs	Reactions	Metabolites	Pathways
P. damnosus	985	1066	116
B. bruxellensis	2230	1698	126

4.2.1 Compartmentalization

After performing PSORTb 3.0 tool for the *P. damnosus* model, four compartments were identified: cytoplasm (928 proteins), cytoplasmic membrane (598 proteins), cell wall (26 proteins) and extracellular space (35 proteins). Although the localization of 436 protein was not possible to predict, so it was assumed as cytoplasmatic. However, the reactions and metabolites are only located in three of these compartments. Enzymatic and spontaneous reactions and the metabolites involved are located in the cytoplasm, transport reactions in the cytoplasmatic membrane and exchange reactions and their metabolites in extracellular space. In the *B. bruxellensis* model, WOLFPSORT predicts nine compartments: cytoplasm (717 protein), cytoskeleton (38 protein), nuclear (2087 protein), endoplasmic reticulum (26 protein), Golgi apparatus (6 protein), peroxisome (4 protein), cytoplasmic membrane (519 protein), and extracellular space (152 protein). It was impossible to identify the localization of 58 proteins in the *B. bruxellensis* model, these proteins were assumed as cytoplasmic. Furthermore, all proteins located in the cytoskeleton were also considered cytoplasmic.

4.2.2 Transport Reactions

In the *P. damnosus* GSMM, TranSyT identified 129 transport reactions, whereas in the *B. bruxellensis* GSMM was identified 495 transport reactions. However, 23 transport reactions was added to the *P. damnosus* GSMM, based on literature information, including four simple diffusion reactions, six uniporter reactions involved in the minimal medium uptake, six reactions that allow the uptake or secretion of sugars and two reactions which secrete compounds of interest. In the B. bruxellensis, 240 reactions were added according to experimental data and information found in the literature. As TranSyT only predicted transport reactions between

extracellular space and cytoplasm, the added reactions to *B. bruxellensis* GSMM include 99 reactions involved in transport to the mitochondrion, 19 involved in the transport to nucleus, 45 responsible for transport to the peroxisome, three with the transport to the Golgi apparatus, 24 with the transport to the endoplasmic reticulum, and 16 responsible for the transport to the vacuole (compartment added, as there is scientific evidence that *B. bruxellensis* stores amino acids in the vacuole [235, 236]). In both models, most of the transport reactions were added without associated genes, as these reactions were essential to the model. All changes were based on literature and databases:

- Reactions for the uptake of compounds from the minimal medium [141, 226];
- Reactions for the compounds secretion, for which there is scientific evidence of their production (e.g. Acetoin in *P. damnosus* [132, 137, 138] and 4-ethylphenol in *B. bruxellensis* [10]);
- Reactions between compartments cytoplasmic and extracellular space in *B. bruxellenis* GSMM (e.g. Transport of compounds participating in the Krebs cycle which takes place in the mitochondrial matrix [237]);
- Reactions involved in the uptake of metabolites used as carbon or nitrogen source (e.g. Trehalose transport reaction in *P. damnosus* [138, 227] and glycerol in *B. bruxellensis* [154, 228]).

Table 13 describes the number of transport reactions obtained with TranSyT and the added reactions for each GSMM grouped according to the T.C. number classification system.

Table 13: Distribution of transport mechanism in each model according to the T.C. number classification system.

T.C. class	TranSyT results		Transport reactions added	
	P. damnosus	B. bruxellensis	P. damnosus	B. bruxellensis
T.C. 1: Channels	7	16	4	41
T.C. 2: Secondary transporters	71	363	13	183
T.C. 3: ABC-binding cassette	30	116	2	16
T.C. 4: Group translocators (PTS)	21	0	3	0
T.C. 5: Electron carriers	0	0	0	0

The transport of several simple compounds, such as oxygen, water, orthophosphate and ammonia, is performed through channels. The metabolites cross the membrane by passive diffusion or facilitated diffusion, involving no energy consumption [238]. The transport mechanism classified as T.C. 2 uses active transport and can be divided into uniporter, symporter and even antiporter. The uniporter mechanism transports a single type of compound across the membrane through of a protein integrated with the membrane, and it can also operate

through facilitated diffusion. On the other hand, the symporter mechanism uses an integrated membrane protein to transport two metabolites across the membrane in the same direction, unlike the antiporter mechanism, which transports two metabolites in opposite directions.

Table 14 shows the number of reactions obtained by TranSyT and the number of reactions added, for each organism, grouped according to their subclass [239].

Table 14: Distribution of transport reactions according to T.C. 2 subclass.

T.C. class	TranSyT results		Transport reactions added	
	P. damnosus	B. bruxellensis	P. damnosus	B. bruxellensis
Uniporter	9	115	8	124
Symporter	56	199	5	36
Antiporter	8	49	0	22

The third T.C. class uses the ABC-binding cassette mechanism, transporting amino acids, sugars, lipids and other metabolites through a system that consumes ATP [240]. As shown in Table 13, T.C. 4 is very common in bacteria and is responsible for the transport of free sugars. The PTS mechanism uses phosphoenolpyruvate as a phosphate donor, converting the sugar into a sugar-phosphate, which is then converted into glucose 6-phosphate, following the Embden-Meyerhof-Parnas pathway [241].

4.3 Biomass and Energy requirements

4.3.1 *P. damnosus*

As mentioned before, *P. damnosus* has a reaction of the biomass with nine macromolecules entities. Due to the lack of information, the macromolecules and their respective fraction were defined based on the existing *Lb. plantarum* WCFS1 [217] model, *Lc. lactis* ssp. lactis IL1403 [114] model, and literature [110, 147]. The protein, DNA, RNA, lipid, PG and EPS contents were obtained from the average of the *Lb. plantarum* WCFS1 [217] and *Lc. lactis* ssp. lactis IL1403 [114] models. The value of the WTA fraction was taken from *Lb. plantarum* WCFS1 [217] model as this was experimentally calculated. Likewise, the LTA and cofactor values were obtained experimentally in the reconstruction of the *Lc. lactis* ssp. lactis IL1403 [114] model and therefore, the stoichiometry of these entities were deduced to *P. damnosus* model. Table 15 presents the biomass composition of *Lb. plantarum* WCFS1 [217], *Lc. lactis* ssp. lactis IL1403 [114] models, and the inferred macromolecules and their respective calculated fraction of *P. damnosus* model.

As there is no experimental information about the content of each macromolecule, it was assumed that *P. damnosus* has a protein content of 34% of the biomass, EPS represents 10% of the biomass, PG content is about 12%, lipids represent a value of 0.05 %, WTA and LTA account for a value around 22% and cofactors represents around 6% of the biomass.

Table 15: Biomass composition of *Lb. plantarum* WCFS1 [217], *Lc. lactis* ssp. lactis IL1403 [114] and *P. damnosus* models.

Macromolecule	Lb. plantarum	L. lactis	Source	P. damnosus
DNA	0.019	0.023	Average	0.020
RNA	0.090	0.107	Average	0.093
Protein	0.261	0.460	Average	0.339
Lipids	0.063	0.034	Average	0.046
EPS	0.099	0.120	Average	0.103
Peptidoglycan	0.145	0.118	Average	0.123
WTA	0.138		*Lb. plantarum* WCFS1	0.138
LTA	0.041	0.080	*Lc. lactis* ssp. lactis IL1403	0.080
Cofactors/vitamins/rest	0.144	0.058		0.058

Adapted from Oliveira et al. 2005 [114] and Teusink et al. 2006 [217].

Tables 16 and 17 summarise the DNA and RNA synthesis reactions obtained using the *e-biomass equation merlin*'s tool [222]. The deoxyribonucleotide and ribonucleotide compositions are represented in the triphosphate form. The coefficients of these reactions are in millimoles of molecule necessary to produce one mole of macromolecule. These values were determined using the *P. damnosus* genome.

Table 16: DNA synthesis reaction for *P. damnosus* model created with *merlin*.

Reactants		
Precursor	**KEGG ID**	**Coefficient (mmol/mol$_{DNA}$)**
dCTP	C00458	0.164
dATP	C00131	0.335
dTTP	C00459	0.305
dGTP	C00286	0.196
Products		
Diphosphate	C00013	1.000
DNA		1.000

The protein synthesis reaction was also generated through the *e-biomass equation merlin*'s tool [222] that uses the *P. damnosus* proteomics data. Table 18 shows the amino acid content obtained.

Table 17: RNA synthesis reaction for *P. damnosus* model created with *merlin*.

Reactants		
Precursor	**KEGG ID**	**Coefficient (mmol/mol$_{RNA}$)**
ATP	C00002	0.270
GTP	C00044	0.292
UTP	C00075	0.211
CTP	C00063	0.227
Products		
Diphosphate	C00013	1.000
RNA		1.000

Table 18: Protein composition for *P. damnosus* model provided by *merlin*.

Precursor	**KEGG ID**	**Coefficient (mmol/g$_{Protein}$)**
Glutaminyl-tRNA	C02282	0.403
Glycyl-tRNA(Gly)	C02412	0.580
L-Alanyl-tRNA(Ala)	C00886	0.659
L-Arginyl-tRNA(Arg)	C02163	0.346
L-Asparaginyl-tRNA(Asn)	C03402	0.459
L-Aspartyl-tRNA(Asp)	C02984	0.516
L-Cysteinyl-tRNA(Cys)	C03125	0.042
L-Glutamyl-tRNA(Glu)	C02987	0.473
L-Histidyl-tRNA(His)	C02988	0.197
L-Isoleucyl-tRNA(Ile)	C03127	0.642
L-Leucyl-tRNA(Leu)	C02047	0.852
L-Lysyl-tRNA(Lys)	C01931	0.644
L-Methionyl-tRNA(Met)	C02430	0.234
L-Phenylalanyl-tRNA(Phe)	C03511	0.3929
L-Prolyl-tRNA(Pro)	C02702	0.314
L-Seryl-tRNA(Ser)	C02553	0.573
L-Threonyl-tRNA(Thr)	C02992	0.569
L-Tryptophanyl-tRNA(Trp)	C03512	0.092
L-Tyrosyl-tRNA(Tyr)	C02839	0.322
L-Valyl-tRNA(Val)	C02554	0.629

Fatty acids are not directly represented in the biomass equation but are essential for the composition of lipids and LTA. The fatty acids composition was obtained using experimental data [225], *Lb. plantarum* WCFS1 [217] model, and *Lc. lactis* ssp. lactis IL1403 [114] model. Thus, a compound representing the average fatty acid composition was created and included in the model, named "e-Fatty_acid". In the following Table 19, the fatty acid composition of *P. damnosus* is described. The coefficient was calculated based on the *Lb. plantarum* WCFS1 [217] and *Lc. lactis* ssp. lactis IL1403 [114] model values average.

Table 19: Fatty acid composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG iD	$g_{Precursor}/g_{Fatty\ acid}$	Reference
Tetradecanoic acid (14:0)	$C_{14}H_{28}O_2$	C06424	0.061	[114, 217]
Hexadecanoic acid (16:0)	$C_{16H32}O_2$	C00249	0.278	[114, 217, 225]
Hexadecenoic acid (16:1)	$C_{16H30}O_2$	C08362	0.074	[114, 217, 225]
Octadecanoic acid (18:0)	$C_{18}H_{36}O_2$	C01530	0.016	[114, 217]
Octadecenoic acid (18:1)	$C_{18}H_{34}O_2$	C00712	0.379	[114, 217, 225]
Methylene octadecanoic acid (cyc 19:0)	$C_{19}H_{36}O_2$		0.192	[114, 217, 225]

Methylene octadecanoic acid (cyc 19:0) is a cyclopropane fatty acid commonly found in LAB, associated with cell membrane stability and acid shock resistance [242–245]. Due to the phylogenetic proximity to lactobacilli species and the results obtained by Beverly et al. in 1997, this cyclopropane fatty acid may be lactobacillic acid (11, 12 methylene octadecanoic acid) or dihydrosterculic acid (9, 10-methylene octadecanoic acid) [225, 242–244]. As there is no specific information about this cyclopropane for *P. damnosus*, a generic compound (Methylene octadecanoic acid (cyc 19:0)) was assumed to be produced from octadecenoic acid, and S-adenosylmethionine [243, 246] as the synthesis mechanism is similar.

The e-Lipid metabolite incorporates metabolites necessary for the cell wall composition. Specifically the phospholipids molecules contain several fatty acid chains and therefore depend on their synthesis. Thus, an average reaction of fatty acids synthesis was created and the coefficients of each precursor were determined based on *Lb. plantarum* WCFS1 [217], and *Lc. lactis* ssp. lactis IL1403 [114] models.

The lipid composition of *P. damnosus*, described in Table 20 is specifically phospholipidic. The components and their coefficients were determined based on the *Lb. plantarum* WCFS1 [217], and *Lc. lactis* ssp. lactis IL1403 [114] models. Phosphatidylglycerol is the most abundant precursor and it participates in the biosynthesis of the other two phospholipids. The 3-O-L-Lysyl-1-O-phosphatidylglycerol is obtained from phosphatidylglycerol through a lysyltransferase (E.C. 2.3.2.3), and cardiolipin is also produced from phosphatidylglycerol through the action of cardiolipin synthase (E.C. 2.7.8.-).

A synthesis reaction of PG, an important cell wall component, was manually created based on literature [110, 132, 137, 138], as described in section 2.8.1. PG consists of a glycan chain of GlcNAc and MurNAc linked to a pentapeptide of L-alanine, D- glutamate, L-lysine, D-aspartate

Table 20: Lipid composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG ID	$g_{Precursor}/g_{Lipid}$
Cardiolipin	$C_{13}H_{18}O_{17}P_2R_4$	C05980	0.036
3-O-L-Lysyl-1-O-phosphatidylglycerol	$C_{14}H_25N_{2O11}PR_2$	C04482	0.252
Phosphatidylglycerol	$C_8H_{13}O_{10}PR_2$	C00344	0.712

and D-alanine. Table 21 shows the final metabolite of the "Peptidoglycan biosynthesis" pathway, composed of all the mentioned precursors .

Table 21: Peptidoglycan composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG ID	$mol_{Precursor}/PG$	$g_{Precursor}/g_{PG}$
Undecaprenyl-diphospho-N-acetylmuramoyl--(N-acetylglucosamine)- L-alanyl-D-glutamyl -meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	$C_{95}H_{156}N_8O_{28}P_2$	C05898	1	1

As mentioned in section 2.8.1, *P. damnosus* produces EPSs, a molecule that has important industrial applications [129, 143]. EPSs are homopolysaccharides composed of three molecules of D-glucose, forming a β-glucan structure [151–153].The EPS composition is described in Table 22.

Table 22: Exopolysaccharide composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG ID	$mol_{Precursor}/EPS$	$g_{Precursor}/g_{EPS}$
UDP-glucose	$C_{15}H_{24}N_2O_{17}P_2$	C00029	3	1

The genes involved in EPS biosynthesis were not found in the genome annotation. However, in 2015, Snauwaert et al. [136] identified four CDS (AH70 07835, AH70 01405, AH70 01410, and AH70 01415) responsible after analysing the draft genome for the EPS production. As mentioned in section 4.1, the genes are not present in the model because the assembly files have been updated in the meantime.

Based on the literature [110, 147–149], *Lb. plantarum* WCFS1 [217] model and the *Lc. lactis* ssp. lactis IL1403 [114] model, the WTA and LTA composition and their coefficients were determined, as shown in Tables 23 and 24, respectively. In section 2.8.1, the precursors of WTA and LTA were discussed. Therefore, WTA is composed of a poly-glycerophosphate covalently attached to PG through a disaccharide linkage and a glycerol-phosphate. In contrast, LTA is composed of a poly-glycerophosphate linked to a diglucosyldiacylglycerol. The mole fraction of each precursor of WTA entity was determined based on the *B. subtillis* 168 [247] model. Table 24 shows the final LTA metabolite of the "Glycerolipid metabolism" pathway composed of all the precursors mentioned.

Table 23: Wall teichoic acids composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG ID	$mol_{Percursor}/WTA$	$g_{Precursor}/g_{WTA}$
CDP-glycerol	$C_{12}H_{21}N_{3}O_{13}P_2$	C00513	45	0.640
UDP-glucose	$C_{15}H_{24}N_2O_{17}P_2$	C00029	15	0.224
UDP-N-acetyl-alpha-D-glucosamine	$C_{17}H_{27}N_3O_{17}P_2$	C00043	1	0.019
D-Alanine	$C_3H_7NO_2$	C00133	15	0.098
UDP-N-acetyl-D-mannosamine	$C_{17}H_{27}N_3O_{17}P_2$	C01170	1	0.019

Table 24: Lipoteichoic acid composition of *P. damnosus* GSMM.

Precursor	Formula	KEGG ID	$mol_{Precursor}/LTA$	$g_{Precursor}/g_{LTA}$
Glycerophosphoglycoglycerolipid	$C_{20}H_{33}O_{20}P(C_{18}H_{34}O_2)_2$	C20897	1	1

e-biomass equation merlin's tool [222] also generates a Cofactor synthesis reaction that included universal organic compounds essential for prokaryotic organism. Therefore, based on *Lb. plantarum* WCFS1 [217] model, *Lc. lactis* ssp. lactis IL1403 [114] model and in the literature [224], the precursors were filtered according to available information and described in Table 25.

Table 25: Cofactors composition of *P. damnosus* GSMM obtained from *merlin*, Teusink et al. [217] and Xavier et al. [224]

Precursor	KEGG ID	$mmol/g_{Cofactor}$
Biotin	C00120	0.315
CoA	C00010	0.100
di-trans,poly-cis-Undecaprenyl diphosphate	C04574	0.083
FAD	C00016	0.098
FMN	C00061	0.169
Glutathione	C00051	0.250
NAD+	C00003	0.116
NADPH	C00005	0.103
Pyridoxal phosphate	C00018	0.311
Riboflavin	C00255	0.204
S-Adenosyl-L-methionine	C00019	0.193
Tetrahydrofolate	C00101	0.173
Thiamin monophosphate	C01081	0.223

Due to lack of information about the energy requirements of *P.damnosus*, the *Lb. plantarum* WCFS1 [217] values were used as reference. The value for growth-associated energy was 56 mmol h^{-1} $g_{DW}{}^{-1}$ (based on total ATP production on the basis of lactate and acetate data) and the non-growth-associated energy was 0.6 mmol $h^{-1}g_{DW}{}^{-1}$.

4.3.2 *B. bruxellensis*

B. bruxellensis has a reaction of the biomass with six macromolecules entities. Both macromolecules and their respective fraction were defined based on the existing *C. tropicalis* [218], *C. glabrata* [219], *K. lactis* [220] and *S. cerevisiae* S288C [221] models. Table 26 presents the inferred macromolecules and their respective calculated fraction. The fractions of each macromolecule were obtained by averaging the stoichiometry of the precursors in the other models.

Table 26: Biomass composition of *C. tropicalis* [218], *C. glabrata* [219], *K. lactis* [220], *S. cerevisiae* S288C [221] and *B. bruxellensis* models.

Macromolecule	S. cerevisiae	C. tropicalis	C. glabrata	K. lactis	Source	B. bruxellensis
RNA	0.063	0.063	0.059	0.04959	Average	0.058
DNA	0.004	0.004	0.0037	0.00328	Average	0.004
Cofactor			0.0087	0.01338	Average	0.011
Protein	0.461	0.53	0.5046	0.26765	Average	0.438
Lipids	0.029	0.0395	0.0502	0.00064	Average	0.030
Carbohydrates	0.407	0.4	0.3738	0.14504	Average	0.459

Due to the lack of experimental information about the specific biomass composition, literature data was considered and it was assumed that protein represents around 44% of the biomass, carbohydrates account for a value of 46%, RNA and DNA content is approximately 0.06%, lipids represent a value of 0.03% and cofactors are responsible for 0.01% of the biomass, as shown in Table 26.

In Tables 27 and 28, the DNA and RNA synthesis reactions are summarised. The DNA synthesis reaction was obtained using the *e-biomass equation merlin*'s tool [222]. As these reactions are generated using the *B. bruxellensis* genome, and no genomic information on RNA is available in the assembly files, the RNA synthesis reaction was defined based on *C. glabrata* [219]. The coefficients of these reactions are in millimoles of molecules necessary to produce one mole of macromolecule.

The *e-biomass equation merlin*'s tool [222] also generated the protein synthesis reaction, using the *B. bruxellensis* proteomics data. Table 29 shows the protein composition, namely amino acid content obtained.

Although not represented directly in the biomass reaction, fatty acids are essential for lipid and phospholipid biosynthesis. In 1992, Razes et al. [248] analysed and quantified the fatty acids from two strains of *B. bruxellensis*: *D. bruxellensis* intermedia and *D. bruxellensis* lambicus.

Table 27: DNA synthesis reaction for *B. bruxellensis* model created with *merlin*.

Reactants		
Metabolite	**KEGG ID**	**Coefficient (mmol/mol$_{DNA}$)**
dCTP	C00458	0.365
dATP	C00131	0.125
dTTP	C00459	0.161
dGTP	C00286	0.349
Products		
Diphosphate	C00013	1.000
DNA		1.00

Table 28: RNA synthesis reaction for *B. bruxellensis* model obtained from N. Xu et al. [219].

Reactants		
Metabolite	**KEGG ID**	**Coefficient (mmol/mol$_{RNA}$)**
ATP	C00002	0.351
GTP	C00044	0.113
UTP	C00075	0.399
CTP	C00063	0.137
Products		
Diphosphate	C00013	1.000
RNA		1.000

Years later, in 2015, Galafassi et al. [249] quantified some of these fatty acids from another strain of *B. bruxellensis*. Thus, the composition of the fatty acids from *B. bruxellensis* was determined, as well as their mole fraction by averaging the experimental values reported in the mentioned studies.

The "e-Lipid" metabolite incorporates phospholipids and sterols, which are essential for cell viability. Although representing a small part of the lipids, sterols are involved in plasma membrane fluidity [251], having phospholipids as the most representative species. Phospholipids contain several fatty acid chains and are thus dependent on their synthesis. The precursors and their coefficients values were determined based on *C. glabrata* [219] and *S. cerevisiae* S288C [221] models. Table 31 summarises the lipid composition of *B. bruxellensis* GSMM.

The cell wall components were added to "e-Carbohydrates" reaction. As discussed in section 2.9, the cell wall is composed mainly of chitin, β-1,3-glucan, β-1,6-glucan and mannan [172, 173]. Whereas β-1,3-glucan has 1500 degrees of polymerization, β-1,6-glucan has 150 and chitin has 120, based on *S. cerevisiae* cell wall components [175].

Table 29: Protein composition reaction for *B. bruxellensis* model provided by *merlin*.

Precursor	KEGG ID	Coefficient (mmol/$g_{Protein}$)
Glutaminyl-tRNA	C02282	0.330
Glycyl-tRNA(Gly)	C02412	0.505
L-Alanyl-tRNA(Ala)	C00886	0.543
L-Arginyl-tRNA(Arg)	C02163	0.453
L-Asparaginyl-tRNA(Asn)	C03402	0.453
L-Aspartyl-tRNA(Asp)	C02984	0.529
L-Cysteinyl-tRNA(Cys)	C03125	0.124
L-Glutamyl-tRNA(Glu)	C02987	0.572
L-Histidyl-tRNA(His)	C02988	0.185
L-Isoleucyl-tRNA(Ile)	C03127	0.555
L-Leucyl-tRNA(Leu)	C02047	0.827
L-Lysyl-tRNA(Lys)	C01931	0.656
L-Methionyl-tRNA(Met)	C02430	0.203
L-Phenylalanyl-tRNA(Phe)	C03511	0.387
L-Prolyl-tRNA(Pro)	C02702	0.371
L-Seryl-tRNA(Ser)	C02553	0.779
L-Threonyl-tRNA(Thr)	C02992	0.477
L-Tryptophanyl-tRNA(Trp)	C03512	0.090
L-Tyrosyl-tRNA(Tyr)	C02839	0.310
L-Valyl-tRNA(Val)	C02554	0.523

B. bruxellensis seems to have only glucose and mannose in its biomass [252]. Therefore, based on literature and in *C. glabrata* [219], and *S. cerevisiae* S288C [221] models, the "e-Carbohydrate" reaction was formulated, as shown in Table 32.

Cofactor reaction synthesis was generated by *e-biomass equation merlin*'s tool [222]. Nevertheless manual curation was performed according to the cofactors reported in the literature [253, 254], *C. tropicalis* [218], *C. glabrata* [219], *K. lactis* [220] and *S. cerevisiae* S288C [221] models. Although molybdenium is not present in *S. cerevisiae*, *B. bruxellensis* requires it to reduce nitrate [253]. Table 33 shows the precursors of cofactors reaction and the coefficient of each precursor. As heme is not synthesised in the absence of oxygen, two cofactor reactions had to be created. The ubiquinone present in *B. bruxellensis* is 95% ubiquinone-9 [254]; thus, it was assumed as the only metabolic species in the model .

Due to a lack of information on the energy requirements of *B. bruxellensis*, the *K. lactis* [220] model was used for reference. In the model validation stage, these parameters were adjusted.

Table 30: Fatty acid composition of *B. bruxellensis* GSMM.

Precursor		Formula	KEGG ID	Molar fraction	Reference
Octanoic acid	C8:0	$C_8H_{16}O_2$	C06423	0.0008	[248]
Decanoic acid	C10:0	$C_{10}H_{20}O_2$	C01571	0.0112	[248]
Dodecanoic acid	C12:0	$C_{12}H_{24}O_2$	C02679	0.0345	[248, 250]
Tetradecanoic acid	C14:0	$C_{14}H_{28}O_2$	C06424	0.0158	[248–250]
(9E)-Tetradecenoic acid	C14:1	$C_{14}H_{26}O_2$		0.0063	[248, 250]
Hexadecanoic acid (Palmitic acid)	C16:0	$C_{16}H_{32}O_2$	C00249	0.2007	[248–250]
(9Z)-Hexadecenoic acid (palmitoleic acid)	C16:1	$C_{16}H_{30}O_2$	C08362	0.3695	[248–250]
Octadecanoic acid (Stearic acid)	C18:0	$C_{18}H_{36}O_2$	C01530	0.0826	[248–250]
(9Z)-Octadecenoic acid (oleic acid)	C18:1	$C_{18}H_{34}O_2$	C00712	0.1774	[248–250]
(9Z,12Z)-Octadecadienoic acid (linoleic acid)	C18:2	$C_{18}H_{32}O_2$	C01595	0.0957	[248–250]
Long-chain fatty acid (Icosanoic acid)	C20:0	$C_{20}H_{40}O_2$	C06425	0.0055	[248, 249]

Table 31: Lipid composition of *B. bruxellensis* GSMM.

Precursor	Formula	KEGG ID	$g_{Precursor}/g_{Lipid}$
Cardiolipin	$C_{13}H_{18}O_{17}P_2R_4$	C05980	0.051
Phosphatidate	$C5H_7O_8PR_2$	C00416	0.021
Phosphatidylglycerol	$C_8H_{13}O_{10}PR_2$	C00344	0.003
Phosphatidylethanolamine	$C_7H_{12}NO_8PR_2$	C00350	0.095
Phosphatidylserine	$C_8H_{12}NO_{10}PR_2$	C02737	0.038
Phosphatidylcholine	$C_{10}H_{18}NO_8PR_2$	C00157	0.227
1-Phosphatidyl-D-myo-inositol	$C_{11}H_{17}O_{13}PR_2$	C01194	0.177
Triacylglycerol	$C_6H_5O_6R$	C00422	0.366
Zymosterol	$C_{27}H_{44}O$	C05437	0.009
Ergosterol	$C_{28}H_{44}O$	C01694	0.010
Lanosterol	$C_{30}H_{50}O$	C01724	0.002
Squalene	$C_{30}H_{50}$	C00751	0.001

The value for growth-associated energy was 59 mmol h^{-1} $g_{DW}{}^{-1}$ and the non-growth-associated energy was 2 mmol $h^{-1}g_{DW}{}^{-1}$.

4.4 Manual Curation

The current section presents all the changes made to both models during the model curation.

Table 32: Carbohydrates composition of *B. bruxellensis* GSMM.

Precursor	Formula	KEGG ID	$g_{Precursor}/g_{Carbohydrates}$
UDP-glucose	$C_{15}H_{24}N_2O_{17}P_2$	C00029	0.000244
GDP-mannose	$C_{16}H_{25}N_5O_{16}P_2$	C00096	0.00013
Mannan	$C_6H_{10}O_5$	C00464	0.000517
1,3-beta-D-Glucan	$C_{9000}H_{15000}O_{7500}$	C00965	0.975215
1,6-beta-D-Glucan	$C_{900}H_{1500}O_{750}$	C02493	0.019902
Chitin	$C_{960}H_{1560}N_{120}O_{600}$	C00461	0.003991

Table 33: Cofactors composition of *B. bruxellensis* GSMM.

Precursor	Formula	KEGG ID	$mmol_{Precursor}/g_{Cofactor}$		Reference
			Aerobiosis	Anaerobiosis	
Biotin	$C_{10}H_{16}N_2O_3S$	C00120	0.273	0.292	*merlin*/[219]
CoA	$C_{21}H_{36}N_7O_{16}P_3S$	C00010	0.087	0.093	*merlin*/[220]
FAD	$C_{27}H_{33}N_9O_{15}P_2$	C00016	0.085	0.091	*merlin*/[219]
FMN	$C_{17}H_{21}N_4O_9P$	C00061	0.146	0.157	*merlin*/[219]
Glutathione	$C_{10}H_{17}N_3O_6S$	C00051	0.217	0.232	*merlin*/[220]
Heme	$C_{34}H_{32}FeN_4O_4$	C00032	0.108		*merlin*/[220]
Molybdenum	$C_{10}H_{12}MoN_5O_8PS_2$	C18237	0.128	0.137	[253]
NAD+	$C_{21}H_{28}N_7O_{14}P_2$	C00003	0.100	0.108	*merlin*/[219]
NADPH	$C_{21}H_{30}N_7O_{17}P_3$	C00005	0.089	0.096	*merlin*/[219]
Pyridoxal Phosphate	$C_8H_{10}NO_6P$	C00018	0.270	0.289	*merlin*/[219]
Riboflavin	$C_{17}H_{20}N_4O_6$	C00255	0.177	0.190	*merlin*/[221]
S-adenosyl-L-methionine	$C_{15}H_{22}N_6O_5S$	C00019	0.167	0.179	*merlin*
Tetrahydrofolate	$C_{19}H_{23}N_7O_6$	C00101	0.150	0.160	*merlin*/[220]
Thiamin Diphosphate	$C_{12}H_{19}N_4O_7P_2S$	C00068	0.157	0.168	*merlin*/[219]
Ubiquinone-9	$C_{54}H_{82}O_4$	C01967	0.084	0.090	*merlin*/[254]

4.4.1 Reversibility and directionality

For both draft models, *merlin*'s automatic correction of reactions reversibility tool was used. Then, manual curation was performed because the production of several biomass precursors was compromised. Thus, in *P. damnosus*' GSMM a total of 155 reactions were corrected: 148 to irreversible and seven to reversible. A total of 266 reactions was corrected in the *B. bruxellensis* GSMM, where 215 reactions to irreversible, and 51 reactions to reversible.

4.4.2 Gap-Filling

Both drafts contained gaps specifically in the cofactors, lipids and amino acids biosynthesis pathways. After the compartmentalization process, in both models several reactions were duplicated and present in several compartments. Overall, 40 reactions were added to the *P. damnosus* model, and 337 reactions were removed. Whereas for the *B. bruxellensis* model 48 reactions were added and 669 removed. In *P. damnosus*' genome annotation, 29 E.C. numbers were modified, whereas for *B. bruxellensis* 37 E.C. numbers were changed (support material Table S1 and Table S2). Therefore, reactions added without an associated gene were always supported by scientific evidence in the literature.

As mentioned in section 2.8.1, *P. damnosus* can produce acetoin and diacetyl. Thus, the ADU72_1545 gene was reanotated to E.C. 2.2.1.6 and R02946, which is catalysed by the enzyme E.C. 1.1.1.4, was added without an associated gene [132, 137, 138]. In *P. damnosus*' model, an alternative pathway to "One carbon pool by folate" was added since the metabolite Tertahydropteroyltri-L-glutamate was being consumed but not produced. Hence, according to Shane in 1989 [255], seven reactions were added to obtain a model with a connected network.

For *P. damnosus*' model to produce the fatty acids, reactions catalysed by the enzymes E.C. 4.2.1.59, E.C. 1.3.1.9, and E.C. 1.14.19.2 were added, without associated genes, as these enzymes are not annotated in the genome. Beverly et al. in 1997 [225] reported that a cyclopropane fatty acid, methylene octadecanoic acid (cyc 19:0), is present in *P. damnosus*. Therefore, based on phylogenetics proximity, this cyclopropane fatty acid may be lactobacillic acid (11, 12 methylene octadecanoic acid) or dihydrosterculic acid (9, 10-methylene octadecanoic acid) identified in many *Lactobacilli* and associated with cell membrane stability and acid shock resistance [242–245, 256]. This two methylene octadecanoic acids have similar production mechanisms: S-adenosylmethionine is required as methyl donor, generating also S-adenosylhomocysteine. The Autoinducer-2, which can be found in several gram-positive and gram-negative bacteria [246], was produced to convert S-adenosylhomocysteine to homocysteine, maintaining the model in a pseudo-steady state. This Autoinducer-2 is a quorum-sensing signaling molecule and quorum-sensing signaling has been identified in pediocins [257]. A transport reaction for this compound was added and associated to T.C. number 2.A.86.

In *B. bruxellensis*' model, there are five unsaturated fatty acids in the fatty acid synthesis reaction. Thus, due to the lack of information in KEGG, 38 reactions based on the *S. cerevisiae* S288C [221] model were added, in order to synthesise these fatty acids. The E.C. numbers involved in the fatty acid biosynthesis are present in the model, thus all the added reactions have associated genes. Althought, the E.C. 6.4.1.2, the E.C. 3.1.2.21 and E.C. 1.14.19.2 were required in the "Fatty acid biosynthesis" pathway. Since they are not annotated in genome annotation, the reactions catalysed by them were added without associated genes.

As already mentioned, *B. bruxellensis* needs a molybdenum cofactor to reduce nitrate [253]. R12621 and a sink reaction for metabolite 5-Deoxy-D-ribose (C22288) were added to obtain a model with a connected network, according to *Escherichia coli* str. K-12 substr. MG1655 model [33], since 5'-Deoxyadenosine (C05198) was a dead-end.

B. bruxellensis does not appear to have any 4-hydroxybenzoate synthesis mechanisms, nor is it one of the essential metabolites in the minimal medium. Thus, it was assumed that the biosynthesis process of ubiquinone-9 occurs from the 4-aminobenzoate. This mechanism was considered similar to the ubiquinone-6 biosynthesis mechanism from para-aminobenzoate reported in *S. cerevisiae* and described in the literature [258].

There is scientific evidence that *B. bruxellensis* produces volatile phenols namely 4-ethylphenol, 4-ethylguaiacol, 4-vinylphenol (or 4-hydroxystyrene) and 4-vinylguaiacol . The p-coumaric acid, ferulic acid and caffeic acid are converted into hydroxystyrenes (4-vinylguaicol, 4-vinylphenol) wich are then reduced into ethyl derivatives(4-ethylphenol, 4-ethylguaiacol) [169, 170]. Although enzymes responsible for biosynthesis of these compounds are not available in the genome annotation, six reactions without associated genes were added: RXN25879 and RXN25875 (ModelSEED ID), R11071, R02952, R03366 and R07826. Only these last three reactions are associated with complete E.C. numbers.

4.4.3 GPR

merlin's tool established 110 and 191 GPR rules for *P. damnosus* and *B. bruxellensis* GSMMs, respectively.

Therefore, the GSMMs are cured and ready for the validation process. Table 34 shows GSMMs details.

Table 34: GSMMs details.

GSMMs	Reactions	Metabolites	Pathways
P. damnosus	809	589	67
B. bruxellensis	2095	1249	79

Tables S4 and S5 (support material) show the pathways present in each model and the respective number of reactions.

4.5 Model validation

In this section, the results obtained during the *P. damnosus* and *B. bruxellensis* model validation are shown.

4.5.1 No uptake

The first simulation for both models verifies if the model can produce biomass even if the medium provides no metabolites. Therefore, the lower bounds of all the exchange reactions were set to zero, and the biomass growth rate obtained for both models was zero, not having any reaction with flux. Consequently, the models proceed to next validation stage.

4.5.2 Minimal Medium

P. damnosus

At a minimal medium, the growth rate of *P. damnosus* was calculated based on experimental data obtained from the literature [142]. This growth rate was obtained under anaerobic conditions. Still, since *P. damnosus* does not appear to have oxidative phosphorylation (no information about oxidative phosphorylation was found, and the genome annotation did not contain any information about enzymes involved in this pathway), the same specific growth rate for anaerobiosis and aerobiosis was considered . The specific growth rate experimentally determined was 0.04h^{-1} [142]. As mentioned in 2.8.1, *P. damnosus* can produce L- and D-lactate, diacetyl (such as butanedione), and acetoin [132, 137, 138]. During model validation, the non-growth-associated energy was set to 0.60 mmol $h^{-1}g_{DW}^{-1}$ and the growth-associated energy requirements to 56 mmol g_{DW}^{-1}. The uptake value for glucose was set to -1.9 mmol h^{-1} g_{DW}^{-1}, whereas the amino acids uptake value was set to -0.2 mmol h^{-1} g_{DW}^{-1} according to *Lb. plantarum* WCFS1 [217]. The exchange reactions bounds of the remaining metabolites were defined as unconstrained, i.e. with values of -10000 mmol $h^{-1}g_{DW}^{-1}$ for the lower bound and 10000 mmol $h^{-1}g_{DW}^{-1}$ for the upper bound. The simulations were performed using mainly pFBA. Yet, whenever a specific compound was not produced, contradicting the experimental data, FVA was performed (Table 37). If the maximum value obtained in the FVA is different from zero, it means that the model can produce the metabolite under these conditions and thus validate the model. Tables 35 and 36 show the consumption and production values, and the compounds produced under aerobic and anaerobic conditions.

During *in silico* simulations, the two lactate configurations can not be secreted simultaneously. However, the model is capable of producing both. Acetoin is completely converted to butanedione; hence not secreted in pFBA simulations. The degradation of L-cysteine by cysteine-S-conjugate beta-lyase originates hydrogen sulfide and ammonia, which are secreted. Thereby, the *P. damnosus* model can produce the expectable metabolites in a minimal medium, being apt for the simulation using different carbon sources.

B. bruxellensis

In 2011, Rozpedowska et al. [11] calculated the specific growth rate of *B. bruxellensis* under aerobic and anaerobic conditions in a minimal defined medium including the secreted

Table 35: *In silico* consumption rates of metabolites present in the minimal medium by *P. damnosus* model.

Metabolite	KEGG ID	Consumption rate (mmol h^{-1} g_{DW}^{-1})		Metabolite	KEGG ID	Consumption rate (mmol h^{-1} g_{DW}^{-1})	
		Aerobiosis	Anaerobiosis			Aerobiosis	Anaerobiosis
Adenine	C00147	0.0054	0.0054	L-Phenylalanine	C00079	0.0054	0.0053
Biotin	C00120	0.0007	0.0007	L-Proline	C00148	0.0043	0.0043
D-Glucose	C00031	1.9000	1.9000	L-Serine	C00065	0.0148	0.0147
Fe_3^+	C14819	0.0000	0.0010	L-Threonine	C00188	0.0078	0.078
Folinic acid	C03479	0.0004	0.004	L-Tryptophan	C00078	0.0013	0.0013
Glycine	C00037	0.0000	0.0000	L-Tyrosine	C00082	0.0044	0.0044
Guanine	C00242	0.0039	0.0039	L-Valine	C00183	0.0086	0.0086
L-Alanine	C00041	0.0245	0.0244	Nicotinate	C00253	0.0005	0.0005
L-Arginine	C00062	0.0047	0.0047	Orthophosphate	C00009	0.0000	0.0691
L-Aspartate	C00049	0.0185	0.0184	Oxygen	C00007	0.0140	0.0000
L-Cysteine	C00097	0.2000	0.2000	Pantothenate	C00864	0.0002	0.0002
L-Glutamate	C00025	0.0151	0.0151	Pyridoxine	C00314	0.0012	0.0012
L-Histidine	C00135	0.0032	0.0032	Riboflavin	C00255	0.0011	0.0011
L-Isoleucine	C00407	0.0088	0.0088	Thymine	C00178	0.0000	0.0000
L-Leucine	C00123	0.0116	0.0116	Uracil	C00106	0.0148	0.0148
L-Lysine	C00047	0.0088	0.0087	Xanthine	C00385	0.0000	0.0000
L-Methionine	C00073	0.0036	0.0036				

compounds. In aerobiosis, *B. bruxellensis* has a specific growth rate of 0.12 h^{-1}, and besides ethanol, acetate is also secreted. Under anaerobic conditions, the specific growth rate is 0.07 h^{-1}, and ethanol and a small amount of glycerol are secreted [11]. As discussed in section 2.9, *B. bruxellensis* produces volatile phenols such as 4-ethylphenol [10, 169, 170]. In aerobiosis, the lower bound of the oxygen exchange reaction was constrained to -0.1 mmol $h^{-1}g_{DW}^{-1}$ as the experimental data for specific growth rate was obtained in an environment with a controlled dissolved oxygen concentration [11]. The exchange reactions of the three amino acids present in the minimal medium were restricted to -0.2 mmol $h^{-1}g_{DW}^{-1}$. The glucose uptake was defined as -2.6 mmol $h^{-1}g_{DW}^{-1}$. In anaerobiosis, the exchange reactions bounds of the sterols and unsaturated fatty acids included in *B. bruxellensis* biomass were defined as unconstrained, with the lower bound value of -10000 mmol $h^{-1}g_{DW}^{-1}$ and the upper bound value of 10000 mmol $h^{-1}g_{DW}^{-1}$, as reported in the literature [220]. The exchange reactions bounds of the three amino acids were restricted as in aerobiosis, and the glucose uptake was set to -1.8 mmol $h^{-1}g_{DW}^{-1}$. The exchange reactions bounds of the remaining metabolites present in the minimal medium were not restricted, having values of -10000 mmol $h^{-1}g_{DW}^{-1}$ for lower bound and 10000 mmol

Table 36: *In silico* secreted compounds by *P. damnosus* model, in a minimal medium.

Metabolite	Formula	KEGG ID	Production rate (mmol h^{-1} g_{DW}^{-1})	
			Aerobiosis	Anaerobiosis
D/L-Lactate	$C_3H_6O_3$	C00256/C00186	3.437	3.446
(R,R)-Butane-2,3-diol	$C_4H_{10}O_2$	C03044	0.106	0.104
Ammonia	NH_3	C00014	0.174	0.175
Autoinducer2			0.002	0.002
Biomass			0.040	0.040
Fe_2^+	Fe	C14818	0.000	0.001
H^+	H	C00080	0.400	0.385
HCO_3^-	HCO_3	C00288	0.416	0.399
Hydrogensulfide	H_2S	C00283	0.198	0.198

Table 37: FVA results - *P. damnosus*

Metabolite	Formula	KEGG ID	Aerobiosis		Anaerobiosis	
			Minimum	Maximum	Minimum	Maximum
D-Lactate	$C_3H_6O_3$	C00256	0	3.880	0	3.884
L-Lactate	$C_3H_6O_3$	C00186	0	3.880	0	3.884
(R)-Acetoin	$C_4H_8O_2$	C00810	0	0.380	0	0.240

$h^{-1}g_{DW}^{-1}$ for upper bound, for both aerobic and anaerobic conditions. During model validation, the non-growth-associated energy requirements were redefined to 0.45 mmol $h^{-1}g_{DW}^{-1}$ and the growth-associated energy to 10 mmol g_{DW}^{-1} by fitting the model results to experimental data on growth rate. Therefore, after setting all constraints, the model was simulated using pFBA. FVA was used to check if the model is capable of producing a particular metabolite that should have been excreted in the simulation with pFBA (Table 40), for example glycerol in anaerobiosis and acetate in aerobiosis. Tables 38 and 39 show the uptake and export fluxes and the compounds produced, under aerobic an anaerobic conditions.

The exchange reactions to L-homoserine and phosphatidylcholine were constrained to zero, i.e. the exchange reactions lower bound was set to zero; otherwise, these metabolites would be secreted, leading to a not expected flux diversion. In aerobiosis, acetate is not secreted by pFBA simulation because acetaldehyde is converted to ethanol. In anaerobiosis, glycerol was not secreted in pFBA simulation, but the FVA analysis noted that it could be produced and excreted.

Table 38: *In silico* consumption rates of metabolites present in the minimal medium by *B. bruxellensis*

Metabolite	KEGG ID	Consumption rate (mmol h^{-1} g_{DW}^{-1})		Metabolite	KEGG ID	Consumption rate (mmol h^{-1} g_{DW}^{-1})	
		Aerobiosis	Anaerobiosis			Aerobiosis	Anaerobiosis
(9Z)-Hexadecenoic acid	C08362	—	0.00262	L-Methionine	C00073	0.01822	0.01113
(9Z)-Octadecenoic acid	C00712	—	0.00126	L-Tryptophan	C00078	0.04563	0.00286
(9Z)-Tetradecenoic acid	C08322	—	0.00005	Molybdate	C06232	0.00018	0.00012
4-Aminobenzoate	C00568	0.00000	0.00000	myo-Inositol	C00137	0.00070	0.00043
Ammonia	C00014	0.59806	0.41319	Nicotinate	C00253	0.00027	0.00018
Biotin	C00120	0.00039	0.00025	Orthophosphate	C00009	0.03046	0.01863
D-Glucose	C00031	2.60000	1.80000	Oxygen	C00007	0.10000	0.00000
Ergosterol	C01694	—	0.00005	Pantothenate	C00864	0.00012	0.00008
Fe_2^+	C14818	0.00015	0.00000	Pyridoxal	C00250	0.00038	0.00025
Folate	C00504	0.00021	0.00014	Riboflavin	C00255	0.00058	0.00038
Lanosterol	C01724	—	0.00001	Sulfate	C00059	0.00000	0.00000
L-Histidine	C00135	0.00971	0.00590	Thiamine	C00378	0.00022	0.00015
Linoleate	C01595	—	0.00068	Zymosterol	C05437	—	0.00005

Table 39: *In silico* secreted compounds by *B. bruxellensis* model in a minimal medium.

Metabolite	Formula	KEGG ID	Production rate (mmol h^{-1} g_{DW}^{-1})	
			Aerobiosis	Anaerobiosis
4-Ethylguaiacol	$C_9H_{12}O_2$	62465 (PubChem CID)	0.00430	0.00270
5-Deoxy-D-ribose:	$C_5H_{10}O_4$	C22288	0.00018	0.00012
Biomass			0.12001	—
Biomass Anaerobic			—	0.07295
CO_2	CO_2	C00011	2.69726	1.81735
Ethanol	C_2H_6O	C00469	3.22081	2.24316
H_2O	H_2O	C00001	2.07351	1.43500
HCO_3^-	HCO_3	C00288	0.47414	0.26955
Succinate	$C_4H_6O_4$	C00042	0.44337	0.36847

Table 40: FVA results- *B. bruxellensis*

Metabolite	Formula	KEGG ID	Aerobiosis		Anaerobiosis	
			Minimum	Maximum	Minimum	Maximum
Glycerol	$C_3H_8O_3$	C00116			0.000	0.360
Acetate	$C_2H_4O_2$	C00033	0.000	0.539		

4.5.3 Different carbon sources

To evaluate the growth of both models in different carbon sources, simulations using minimal medium were preformed in aerobiosis and anaerobiosis. During these simulations, the carbon source (initially glucose) were changed but the uptake rate was maintained. Due to the lack of information for both organisms, only a qualitative comparison was performed between the *in silico* simulation results and experimental data. Therefore, if the growth rate is greater than zero, the model can grow.

P. damnosus

The different carbon sources [138, 227] are listed in section 3.7.3. Table 41 shows the *in silico* simulation results.

Analyzing the results, the *P. damnosus* model grew on seven of the different carbon sources used, namely alpha,alpha-trehalose, cellobiose, D-fructose, D-galactose, D-glucose, D-mannose and sucrose. When alpha,alpha-trehalose, galactose and sucrose were used as the main carbon source, the experimental data showed that only a few strains could grow [138, 227]. Notwithstanding that fact, this model was able to grow in such a carbon source. With the other carbon sources used in the simulation, the model could not produce biomass. The simulation with glycerol as a carbon source was considered infeasible, and the model has no exchange reactions for the other carbon sources. In general, the *P. damnosus* GSMM showed a phenotype behaviour similar to the literature.

B. bruxellensis

B. bruxellensis grows on carbon sources other than glucose [154, 228], as mentioned in section 3.7.3. The *in silico* simulation results are available in Table 42,

Experimental data show that *B. bruxellensis* grows in media with D-fructose, D-glucose and D-mannose as single carbon sources, both in aerobiosis and anaerobiosis. According to the literature, some strains do not grow on a medium with glycerol, D-galactose or alpha,alpha-trehalose as the carbon source. Both aerobiosis and anaerobiosis simulations showed that the model could produce biomass only when alpha,alpha-trehalose was the carbon source. In the case of glycerol, the model simulation was infeasible under anaerobic conditions, while no biomass was produced in aerobic conditions. When the carbon source is D-fructose, D-glucose and D-mannose, the model exhibited growth in aerobic and anaerobic conditions. These results

Table 41: *P.damnosus* GSMM *in silico* simulation under different carbons sources using pFBA.

Metabolite	KEGG ID	Experimental data	Biomass h^{-1}	
			Aerobiosis	Anaerobiosis
alpha,alpha-Trehalose	C01083	+/-	0.112	0.112
Cellobiose	C00185	+	0.088	0.088
Dextrin	C00721	-	NA	NA
D-Fructose	C00095	+	0.040	0.040
D-Galactose	C00124	+/-	0.040	0.040
D-Glucose	C00031	+	0.040	0.040
D-Mannose	C00159	+	0.040	0.040
D-Ribose	C00121	-	NA	NA
D-Xylose	C00181	-	NA	NA
Glycerol	C00116	-	infeasible	infeasible
Lactose	C00243	-	NA	NA
L-Rhamnose	C00507	-	NA	NA
Starch	C00369	-	NA	NA
Sucrose	C00089	+/-	0.088	0.088

NOTE: (+), the organism grow in the medium; (-), the organism do not grow in the medium; (+/-), some strains grow in the medium, but others do not. NA: no exchange reaction in the model

are corroborated by the data reported in the literature. For the remaining carbon sources, the model did not show any exchange reactions, and therefore there was no biomass production as expected.

4.5.4 Different nitrogen sources

B. bruxellensis can grow in different nitrogen sources. Considering the same uptake rate of ammonia, the models were simulated with different nitrogen sources in aerobiosis and anaerobiosis [154, 228]. Due to the lack of information, a qualitative comparison between the *in silico* simulation results and experimental data was performed.

The different nitrogen sources are listed in section 3.7.4. The growth in some nitrogen sources depends on oxygen presence. Table 43 shows the *in silico* simulation results under aerobic and anaerobic conditions.

Table 42: *B. bruxellensis* GSMM pFBA *in silico* simulation under different carbons sources.

Metabolite	KEGG ID	Experimental data	Biomass h^{-1}	
			Aerobiosis	Anaerobiosis
alpha,alpha-Trehalose	C01083	+/-	0.240	0.156
D-Fructose	C00095	+	0.120	0.073
D-Galactose	C00124	+/-	NA	NA
D-Glucose	C00031	+	0.120	0.073
D-Mannose	C00159	+	0.120	0.073
D-Ribose	C00121	-	NA	NA
Glycerol	C00116	+/-	0.000	infeasible
Lactose	C00243	-	NA	NA
L-Rhamnose	C00507	-	NA	NA
Maltose	C00208	+/-	0.240	0.156
Raffinose	C00492	+/-	NA	NA
Starch	C00369	-	NA	NA

NOTE: (+), the organism grow in the medium; (-), the organism do not grow in the medium; (+/-), some strains grow in the medium, but others do not. NA: no exchange reaction in the model

In aerobiosis, *B. bruxellensis* can use as nitrogen source all metabolites shown in Table 10 (section 3.7.4). In turn, the simulations provided results following the experimental data since the *B. bruxellensis* GSMM produced biomass for all the different nitrogen sources used.

Under anaerobic conditions, *B. bruxellensis* does not grow when L-cystein and L-tryptophan are used as nitrogen sources. The simulations performed using these amino acids as nitrogen sources demonstrated that the *B. bruxellensis* GSMM could not produce biomass for either of them. In the case of L-methionine or L-proline as nitrogen sources, only a few strains can grow. The simulations' results report that the *B. bruxellensis* GSMM can grow when L-proline is used as nitrogen source but cannot produce biomass from L-methionine. For the remaining nitrogen sources, the model produced biomass, confirming the data obtained in the literature.

4.5.5 Minimal medium test

Tables 44 and 45 show that although some compounds do not seem to be essential in *in silico* simulations, all interfere in biomass production. For *B. bruxellensis* model, it can be conclued that *B. bruxellensis* GSMM is auxotrophic for three amino acids, both *in silico* and *in vivo*. *In silico*, the *P. damnosus* GSMM does not seem to be auxotrophic for some amino acids in a defined minimal medium.

Table 43: *B. bruxellensis* GSMM pFBA *in silico* simulation under different nitrogen sources in aerobiosis and anaerobiosis.

Metabolite	KEGG ID	Aerobiosis		Anaerobiosis	
		Experimental data	Biomass (h^{-1})	Experimental data	Biomass (h^{-1})
Ammonia	C00014	+	0.120	+	0.073
L-Arginine	C00062	+	0.134	+	0.816
L-Cystein	C00097	+	0.036	-	0.000
L-Glutamate	C00025	+	1.317	+	1.311
L-Methionine	C00073	+	0.036	+/-	0.000
L-Proline	C00148	+	0.132	+/- *1	0.080
L-Tryptophan	C00078	+	0.036	-	0.000
Nitrate	C00244	+ *2	0.153	+ *2	0.096
Nitrite	C00088	+	0.153	+	0.096

NOTE: (+), the organism grow in the medium; (-), the organism do not grow in the medium; +/-, some strains grow in the medium, but others do not.
*1- Crauwels et al. 2015 [229],Conterno et al. 2006 [230];
*2-A. R. Borneman et al. 2014 [231].

Overall, the GSMM simulation results matched the experimental data. Thus, the *P. damnosus* GSMM and *B. bruxellensis* GSMM were considered validated.

4.6 Community models

The interaction's prediction, the assembly of the *P. damnosus* and *B. bruxellensis* community model and the simulations results are described in this section.

4.6.1 Interactions predictions

After validating both GSMMs, SMETANA [214] was used to analyse the interactions between *P. damnosus* and *B. bruxellensis*. The predicted interactions are presented in Table 46.

All the predicted interactions had *B. bruxellensis* as donor and *P. damnosus* as the receiver. Most of the interactions are amino acid exchanges, making sense since *P. damnosus* is auxotrophic for many amino acids while *B. bruxellensis* can synthesize almost all the ones it needs. Vitamins like riboflavin and nicotinate and the cofactor S-adenosyl-L-methionine are other metabolites predicted to be exchanged between the models. In sum, there are predicted exchanges from *B. bruxellensis* GSMM to *P. damnosus* GSMM of metabolites that are important and necessary for its metabolism.

4.6.2 Community model reconstruction

The community model was reconstructed using ReFramed. All reactions and metabolites were assigned with a suffix referring to the model, as the latter were treated as compartments. Exchange reactions named "pool", representing the exchanges between extracellular medium with the community were also added. The community model presents 3045 reactions and 1979 metabolites.

4.6.3 Community model simulation

Due to the lack of information on the *P. damnosus* and *B. bruxellensis* community, only two simulations were performed under aerobic conditions with restricted oxygen uptake, considering lambic beer the conditions. Using the set of minimal media used in the validation process of each model (Table S3 of support material) as environmental conditions, the first simulation was performed in COBRApy using pFBA. In all simulations, the lower bound value of the amino acid exchange reactions was set to -0.2 mmol $h^{-1}g_{DW}^{-1}$, for oxygen was -0.1 mmol $h^{-1}g_{DW}^{-1}$, for glucose was -2 mmol $h^{-1}g_{DW}^{-1}$ and for the exchange reactions of the remaining metabolites represented in Table S3 of Support material was -10000 mmol $h^{-1}g_{DW}^{-1}$. These uptake values were based on the values used in the GSMM validation process in section 4.5. The metabolites consumed and their rate of consumption are shown in Table 47, while Table 48 shows the metabolites secreted and their production rates.

The simulation results showed a community growth rate of 0.0495 h^{-1}, a value close to the specific growth rate of *P. damnosus*. Volatile phenols, which are aromatic compounds produced by *B. bruxellensis*, were secreted. D-Lactate results from the lactic acid fermentation of *P. damnosus* and ethanol from the ethanol fermentation of *B. bruxellensis*.

A new simulation was performed as many amino acids are being excreted, and *B. bruxellensis* is only auxotrophic for three amino acids. From the minimal medium set (represented in Table S3 in Support material), the amino acids L-histidine, L-methionine, and L-tryptophan kept the lower bound value of their exchange reactions of -0.2 mmol $h^{-1}g_{DW}^{-1}$. The lower bounds of the exchange reactions of the remaining amino acids were set to zero. Tables 49 and 50 show the metabolites consumed and secreted, as well as the consumption and production rates, respectively.

In this simulation only L-alanine was secreted. The *B. bruxellensis* GSMM produced L-alanine and most of it was consumed by the *P. damnosus* GSMM. Although ethanol and D-lactate are not secreted, volatile phenol 4-ethylguaiacol and succinate are still secreted. The latter is produced by the *B. bruxellensis* and results from the reaction catalyzed by succinate-CoA ligase in the Krebs cycle. Also produced by the *B. bruxellensis*, the secreted carbon dioxide originates from isocitrate dehydrogenase in the Krebs cycle and the pyruvate dehydrogenase

action. The community growth rate was 0.39 h^{-1}, keeping close to the specific growth rate of the *P. damnosus*.

4.6.4 Interactions analysis

The interactions which occurred in the simulations are shown in Tables 51 and 52. Further analyzing the results, in both simulations, there were reciprocal interactions between the GSMMs, which SMETANA did not predict. Several amino acids were exchanged in the community in both simulations, mostly produced by *B. bruxellensis* GSMM and consumed by *P. damnosus* GSMM. It is noteworthy that in both simulations, *B. bruxellensis* GSMM consumed D-lactate produced by *P. damnosus* GSMM. The *B. bruxellensis* GSMM used D-lactate to produce pyruvate and 2-hydroxyglutarate in a reaction catalyzed by (R)-2-hydroxyglutarate-pyruvate transhydrogenase, a characterized enzyme in *S. cerevisiae* [259]. Nevertheless, the interactions obtained match those predicted by SMETANA.

4.6.5 Microorganism abundance

After community model simulation, the abundance of each microorganism was assumed using SteadyCom, which is implemented in Reframed. Steadycom was run three times:

1. Firstly, the lower bounds of all exchange reactions present in the community model was set to -10000 mmol $h^{-1}g_{DW}^{-1}$ (open drains);
2. Secondly, SteadyCom was performed with the minimum media set as environmental conditions;
3. And lastly, the environmental conditions were changed to a minimal medium with only three amino acids.

Tables 53, 54, and 55 describe the results obtained. SteadyComVA was also run to evaluate the abundance variance of each microorganism in the community.

When analyzing the results, all performances predict the existence of only one microorganism. In the first and second performances, community growth was verified, although SteadyComVA showed that the abundance of the *B. bruxellensis* GSMM model did not change. On the other hand, in the performance with a minimal medium with only three amino acids, the results report that there was no community growth. The SteadyCom limitations justify these results, as there are no obligatory cross-feeding interactions.

The GSMMs xml files can be found in the following link:
https://nextcloud.bio.di.uminho.pt/s/jH8STJQTN22qtDB

Table 44: Minimal medium test *P. damnosus* - Biomass values (h^{-1}).

Metabolite	KEGG ID	Biomass h^{-1}	
		Aerobiosis	Anaerobiosis
Adenine	C00147	0.040141	0.040199
Biotin	C00120	0.000000	0.000000
D-Glucose	C00031	infeasible	infeasible
$Fe_3{}^+$	C14819	0.000000	0.040269
Folinic acid	C03479	0.000000	0.000000
Glycine	C00037	0.040162	0.040269
Guanine	C00242	0.040056	0.040171
L-Alanine	C00041	0.000000	0.000000
L-Arginine	C00062	0.039941	0.039973
L-Aspartate	C00049	0.000000	0.000000
L-Cysteine	C00097	0.000000	0.000000
L-Glutamate	C00025	0.000000	0.000000
L-Histidine	C00135	0.000000	0.000000
L-Isoleucine	C00407	0.000000	0.000000
L-Leucine	C00123	0.000000	0.000000
L-Lysine	C00047	0.040053	0.040053
L-Methionine	C00073	0.040120	0.040177
L-Phenylalanine	C00079	0.000000	0.000000
L-Proline	C00148	0.040146	0.040215
L-Serine	C00065	0.040162	0.040269
L-Threonine	C00188	0.000000	0.000000
L-Tryptophan	C00078	0.000000	0.000000
L-Tyrosine	C00082	0.000000	0.000000
L-Valine	C00183	0.000000	0.000000
Nicotinate	C00253	0.000000	0.000000
Orthophosphate	C00009	0.000000	0.000000
Pantothenate	C00864	0.000000	0.000000
Pyridoxine	C00314	0.000000	0.000000
Riboflavin	C00255	0.000000	0.000000
Thymine	C00178	0.040162	0.040269
Uracil	C00106	0.000000	0.000000
Xanthine	C00385	0.040162	0.040269

Table 45: Minimal medium test *B. bruxellensis* - Biomass values (h^{-1}).

Metabolit	KEGG ID	Biomass h^{-1}	
		Aerobiosis	Anaerobiosis
(9Z)-Hexadecenoic acid	C08362	—	0.00000
(9Z)-Octadecenoic acid	C00712	—	0.00000
(9Z)-Tetradecenoic acid	C08322	—	0.00000
4-Aminobenzoate	C00568	0.12001	0.07293
Ammonia	C00014	0.03558	0.00000
Biotin	C00120	0.00000	0.00000
D-Glucose	C00031	infeasible	infeasible
Ergosterol	C01694	—	0.00000
Fe_2^+	C14818	0.00000	0.07295
Folate	C00504	0.11989	0.07286
Lanosterol	C01724	—	0.00000
L-Histidine	C00135	0.00000	0.00000
Linoleate	C01595	—	0.07286
L-Methionine	C00073	0.00000	0.00000
L-Tryptophan	C00078	0.00000	0.00000
Molybdate	C06232	0.00000	0.00000
myo-Inositol	C00137	0.00000	0.00000
Nicotinate	C00253	0.11993	0.00000
Orthophosphate	C00009	0.00000	0.00000
Pantothenate	C00864	0.12000	0.07293
Pyridoxal	C00250	0.00000	0.00000
Riboflavin	C00255	0.11976	0.07279
Sulfate	C00059	0.12001	0.07293
Thiamine	C00378	0.00000	0.00000
Zymosterol	C05437	—	0.00000

Table 46: Predicted interactions between *P. damnosus* and *B. bruxellensis*.

Metabolite	KEGG ID	Receiver	Donor
(R)-Pantothenate	C00864	*P. damnosus*	*B. bruxellensis*
4-Phospho-L-aspartate	C03082	*P. damnosus*	*B. bruxellensis*
alpha-D-Glucose	C00267	*P. damnosus*	*B. bruxellensis*
CO_2	C00011	*P. damnosus*	*B. bruxellensis*
D-Alanine	C00133	*P. damnosus*	*B. bruxellensis*
D-Frutose	C00095	*P. damnosus*	*B. bruxellensis*
D-Glucose	C00031	*P. damnosus*	*B. bruxellensis*
Diphosphate	C00013	*P. damnosus*	*B. bruxellensis*
Glycine	C00037	*P. damnosus*	*B. bruxellensis*
H^+	C00080	*P. damnosus*	*B. bruxellensis*
Hydrogen sulfide	C00283	*P. damnosus*	*B. bruxellensis*
L-Alanine	C00041	*P. damnosus*	*B. bruxellensis*
L-Asparagine	C00152	*P. damnosus*	*B. bruxellensis*
L-Aspartate	C00049	*P. damnosus*	*B. bruxellensis*
L-Cystein	C00097	*P. damnosus*	*B. bruxellensis*
L-Glutamate	C00025	*P. damnosus*	*B. bruxellensis*
L-Glutamate 5-semialdehyde	C01165	*P. damnosus*	*B. bruxellensis*
L-Glutamine	C00064	*P. damnosus*	*B. bruxellensis*
L-Homocysteine	C00155	*P. damnosus*	*B. bruxellensis*
L-Isoleucine	C00407	*P. damnosus*	*B. bruxellensis*
L-Leucine	C00123	*P. damnosus*	*B. bruxellensis*
L-Phenylalanine	C00079	*P. damnosus*	*B. bruxellensis*
L-Serine	C00065	*P. damnosus*	*B. bruxellensis*
L-Threonine	C00188	*P. damnosus*	*B. bruxellensis*
L-Tyrosine	C00082	*P. damnosus*	*B. bruxellensis*
L-Valine	C00183	*P. damnosus*	*B. bruxellensis*
N-Carbamoyl-L-aspartate	C00438	*P. damnosus*	*B. bruxellensis*
Nicotinate	C00253	*P. damnosus*	*B. bruxellensis*
Ornithine	C00077	*P. damnosus*	*B. bruxellensis*
Orthophosphate	C00009	*P. damnosus*	*B. bruxellensis*
Riboflavin	C00255	*P. damnosus*	*B. bruxellensis*
S-Adenosyl-L-methionine	C00019	*P. damnosus*	*B. bruxellensis*
Urea	C00086	*P. damnosus*	*B. bruxellensis*

Table 47: Community GSMM pFBA *in silico* consumption rates of metabolites present in the minimal medium set by the *P. damnosus-B. bruxellensis* community model.

Metabolite	**KEGG ID**	**Consumption rate** (mmol h^{-1} g_{DW}^{-1})
Adenine	C00147	0.0105
Ammonia	C00014	0.2873
Biotin	C00120	0.0011
D-Glucose	C00031	2.0000
Fe_3^+	C14819	10.9117
Folate	C00504	0.0006
Glycine	C00037	0.2000
Guanine	C00242	0.0062
L-Arginine	C00062	0.2000
L-Aspartate	C00049	0.2000
L-Glutamate	C00025	0.2000
L-Histidine	C00135	0.0073
L-Isoleucine	C00407	0.0228
L-Leucine	C00123	0.0322
L-Lysine	C00047	0.0256
L-Methionine	C00073	0.0137
L-Phenylalanine	C00079	0.0150
L-Proline	C00148	0.2000
L-Serine	C00065	0.2000
L-Threonine	C00188	0.2000
L-Tryptophan	C00078	0.0035
L-Tyrosine	C00082	0.2000
Molybdate	C06232	0.0001
myo-Inositol	C00137	0.0003
Nicotinate	C00253	0.0007
Orthophosphate	C00009	0.8570
Oxygen	C00007	0.1000
Pantothenate	C00864	0.0003
Pyridoxal	C00250	0.0002
Pyridoxine	C00314	0.0009
Riboflavin	C00255	0.0016
Thiamine	C00378	0.0009
Uracil	C00106	0.0182

Table 48: Community GSMM pFBA *in silico* secreted compounds by the *P. damnosus-B. bruxellensis* community model in the minimal medium set.

Metabolite	**KEGG ID**	**Production rate** (mmol h^{-1} g_{DW}^{-1})
CO_2	C00011	3.4860
D-Lactate	C00256	0.0863
4-Ethylphenol	C13637	0.1844
Biomass		0.0990
Diphosphate	C00013	0.3800
Ethanol	C00469	0.1801
Fe_2^+	C14818	10.9116
H^+	C00080	10.6974
L-Alanine	C00041	1.4052
L-Homoserine	C00263	0.0047
L-Valine	C00183	0.3491
4-Ethylguaiacol	62465 (PubChem ID)	0.0035
Succinate	C00042	1.1799
Urea	C00086	0.1844
Community Growth		0.0495

Table 49: *In silico* consumption rates of metabolites present in a minimal medium with only three amino acids by the *P. damnosus-B. bruxellensis* community model.

Metabolite	**KEGG ID**	**Consumption rate** (mmol h^{-1} g_{DW}^{-1})
Adenine	C00147	0.00849
Ammonia	C00014	1.31123
Biotin	C00120	0.00086
D-Glucose	C00031	2.00000
Fe_3^+	C14819	14.16228
Folate	C00504	0.00047
Guanine	C00242	0.00500
L-Histidine	C00135	0.00591
L-Methionine	C00073	0.01107
L-Tryptophan	C00078	0.02752
Molybdate	C06232	0.00006
myo-Inositol	C00137	0.00023
Nicotinate	C00253	0.00060
Orthophosphate	C00009	1.26406
Oxygen	C00007	0.10000
Pantothenate	C00864	0.00027
Pyridoxal	C00250	0.00013
Pyridoxine	C00314	0.00072
Riboflavin	C00255	0.00129
Thiamine	C00378	0.00075
Uracil	C00106	0.01471

Table 50: *In silico* secreted compounds by the *P. damnosus*-*B. bruxellensis* community model in a minimal medium with only three amino acids.

Metabolite	**KEGG ID**	**Production rate** (mmol h^{-1} g_{DW}^{-1})
CO_2	C00011	3.488
Fe_2^+	C14818	14.162
Succinate	C00042	0.708
Biomass		0.080
4-Ethylguaiacol	62465 (PubChem ID)	0.003
H^+	C00080	14.005
L-Alanine	C00041	0.972
Diphosphate	C00013	0.593
Community Growth		0.039

Table 51: Interactions of *P. damnosus*-*B. bruxellensis* community model in minimal medium set simulation.

Meatbolite	**KEGG ID**	**Donor**	**Receiver**
D-Lactate	C00256	*P. damnosus*	*B. bruxellensis*
L-Aspartate	C00049	*B. bruxellensis*	*P. damnosus*
L-Asparagine	C00152	*B. bruxellensis*	*P. damnosus*
L-Glutamine	C00064	*B. bruxellensis*	*P. damnosus*
Carbamate	C01563	*B. bruxellensis*	*P. damnosus*
4-Phospho-L-aspartate	C03082	*B. bruxellensis*	*P. damnosus*
CO_2	C00011	*B. bruxellensis*	*P. damnosus*
L-Ornithine	C00773	*B. bruxellensis*	*P. damnosus*
L-Cysteine	C00097	*B. bruxellensis*	*P. damnosus*
S-Adenosyl-L-methionine	C00019	*B. bruxellensis*	*P. damnosus*
L-Glutamate 5-semialdehyde	C01165	*B. bruxellensis*	*P. damnosus*

Table 52: Interactions of *P. damnosus*-*B. bruxellensis* community model resulting from the simulation in a minimal medium with only three amino acids.

Meatbolite	KEGG ID	Donor	Receiver
D-Lactate	C00256	*P. damnosus*	*B. bruxellensis*
Glycerol	C00116	*P. damnosus*	*B. bruxellensis*
L-Proline	C00148	*P. damnosus*	*B. bruxellensis*
Phosphatidate	C00416	*P. damnosus*	*B. bruxellensis*
HCO_3^-	C00288	*P. damnosus*	*B. bruxellensis*
L-Threonine	C00188	*B. bruxellensis*	*P. damnosus*
L-Leucine	C00123	*B. bruxellensis*	*P. damnosus*
L-Aspartate	C00049	*B. bruxellensis*	*P. damnosus*
L-Asparagine	C00152	*B. bruxellensis*	*P. damnosus*
L-Phenylalanine	C00079	*B. bruxellensis*	*P. damnosus*
L-Glutamine	C00064	*B. bruxellensis*	*P. damnosus*
L-Isoleucine	C00407	*B. bruxellensis*	*P. damnosus*
L-Alanine	C00041	*B. bruxellensis*	*P. damnosus*
4-Phospho-L-aspartate	C03082	*B. bruxellensis*	*P. damnosus*
D-Alanine	C00133	*B. bruxellensis*	*P. damnosus*
CO_2	C00011	*B. bruxellensis*	*P. damnosus*
L-Ornithine	C00065	*B. bruxellensis*	*P. damnosus*
L-Serine	C00077	*B. bruxellensis*	*P. damnosus*
L-Cysteine	C00097	*B. bruxellensis*	*P. damnosus*
L-Tyrosine	C00082	*B. bruxellensis*	*P. damnosus*
L-Arginine	C00062	*B. bruxellensis*	*P. damnosus*
S-Adenosyl-L-methionine	C00019	*B. bruxellensis*	*P. damnosus*
Glycine	C00037	*B. bruxellensis*	*P. damnosus*
L-Lysine	C00047	*B. bruxellensis*	*P. damnosus*
L-Valine	C00183	*B. bruxellensis*	*P. damnosus*

Table 53: SteadyCom results - open drains.

SteadyCom		
Community growth	29.35 h^{-1}	
B. bruxellensis GSMM	0	
P. damnosus GSMM	1	
SteadyComVA	Minimum	Maximum
B. bruxellensis GSMM	0	0
P. damnosus GSMM	1	1

Table 54: SteadyCom results - minimal medium set.

SteadyCom		
Community growth	0.11 h^{-1}	
B. bruxellensis GSMM	0	
P. damnosus GSMM	1	
SteadyComVA	Minimum	Maximum
B. bruxellensis GSMM	0	0
P. damnosus GSMM	1	1

Table 55: SteadyCom results - minimal medium with only three amino acids.

SteadyCom		
Community growth	0 h^{-1}	
B. bruxellensis GSMM	0	
P. damnosus GSMM	1	
SteadyComVA	Minimum	Maximum
B. bruxellensis GSMM	0	1
P. damnosus GSMM	0	1

5

CONCLUSION AND FUTURE PERSPECTIVES

Lambic beer is one of the oldest types of beer still produced nowadays, and the composition of its unique flavour is of enormous interest. As mentioned before, this flavour is due to the co-occurrence of different microorganisms which composition also varies throughout almost two years of fermentation. Part of the microbial community present in the last stages of this fermentation was studied in the thesis. The present work showed some insight about the metabolism of the lactic acid bacterium *P. damnosus* and the yeast *B. bruxellensis* as for the possible interactions between the two and flavour enhancers possibly present in Lambic beer.

A GSMM 809 reactions and 589 metabolites was obtained for *P. damnosus*, whereas *B. bruxellensis*' GSMM has 2095 reactions and 1249 metabolites. The validation showed phenotypic predictions in agreement with available data for both models. The *P. damnosus* GSMM simulations showed that the model grew in minimal medium, aerobically, and anaerobically. Specific growth rates in these conditions and the utilization of different carbon sources were consistent with literature data. The production of D-Lactate and L-Lactate, and the production of acetoin or diacetyl were confirmed. Auxotrophy was also analyzed, and the model seems not to be auxotrophic for certain metabolites in the minimal experimental medium.

The *B. bruxellensis* GSMM simulations showed the yeast's ability to grow in the reported minimal medium aerobically and anaerobically, and ethanol production. Simulations with FVA showed that the *B. bruxellensis* GSMM could produce acetate and glycerol from minimal medium, in agreement with experimental data. All simulations predicted the production of volatile phenols, aromatic compounds present in wines and beers that contribute to flavour. On the other hand, succinate production, which is usually used as an acidity regulator in beverages, was predicted. Thus, this model brings new insights into this yeast's role in the final stage of Lambic beer production that might be associated with the flavour's acidification.

Microbial community simulations revealed that *B. bruxellensis* is responsible for producing amino acids for *P. damnosus*, which seems to have a passive role in the community. In fact, in the last stage of Lambic beer fermentation, the number of LAB decreases gradually, which seems to be correctly predicted. Predictions showed that aromatic amino acids and aspartate are the most probable amino acids exchanged between the two organisms, therefore the most common in the medium, unveiling that these metabolites are somewhat responsible for the unique flavour of the Lambic beer. The community model simulation predicted the production

of compounds such as 4-ethylguaiacol, succinate, ethanol, and lactate that can also influence the Lambic beer flavour.

With this work, we were able to predict the role of two microorganisms in the last stage of Lambic beer fermentation. For future work, it would be interesting to study the other stages in microbial community composition, the role of each organism present in those communities, and its influence on the Lambic beer flavour composition. We used only three approaches of microbial community simulation in this work. It would be interesting to use specifically dynamic simulation approaches, such as dFBA, to understand the organism's behaviour over time and metabolite consumption and production profiles. Since the assembly files used for the *P. damnosus* GSMM reconstruction have been updated, the model should be adjusted. The SteadyCom results should also be improved.

BIBLIOGRAPHY

[1] Dai, Z., & Nielsen, J. (2015). Advancing metabolic engineering through systems biology of industrial microorganisms. *Current opinion in biotechnology*, *36*, 8–15.

[2] Biz, A., Proulx, S., Xu, Z., Siddartha, K., Indrayanti, A. M., & Mahadevan, R. (2019). Systems biology based metabolic engineering for non-natural chemicals. *Biotechnology advances*, *37*(6), 107379.

[3] Dias, O., Saraiva, J., Faria, C., Ramirez, M., Pinto, F., & Rocha, I. (2019). Ids372, a phenotypically reconciled model for the metabolism *Streptococcus pneumoniae* strain r6. *Frontiers in microbiology*, *10*, 1283.

[4] Rocha, I., Förster, J., & Nielsen, J. (2008). Design and application of genome-scale reconstructed metabolic models. *Microbial gene essentiality: Protocols and bioinformatics* (pp. 409–431). Springer.

[5] Dias, O., Rocha, M., Ferreira, E. C., & Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. *Nucleic acids research*, *43*(8), 3899–3910.

[6] Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018a). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic acids research*, *46*(15), 7542–7553.

[7] Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, *5*(1), 93.

[8] Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., & Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nature communications*, *2*(1), 1–7.

[9] Spitaels, F., Wieme, A. D., Janssens, M., Aerts, M., Van Landschoot, A., De Vuyst, L., & Vandamme, P. (2015). The microbial diversity of an industrially produced lambic beer shares members of a traditionally produced one and reveals a core microbiota for lambic beer fermentation. *Food Microbiology*, *49*, 23–32.

[10] De Roos, J., & De Vuyst, L. (2019). Microbial acidification, alcoholization, and aroma production during spontaneous lambic beer production. *Journal of the Science of Food and Agriculture*, *99*(1), 25–38.

[11] Rozpedowska, E., Hellborg, L., Ishchuk, O. P., Orhan, F., Galafassi, S., Merico, A., Woolfit, M., Compagno, C., & Piškur, J. (2011). Parallel evolution of the make–accumulate–consume strategy in *Saccharomyces* and *Dekkera* yeasts. *Nature communications*, *2*(1), 1–7.

[12] Barrett, C. L., Kim, T. Y., Kim, H. U., Palsson, B. Ø., & Lee, S. Y. (2006). Systems biology as a foundation for genome-scale synthetic biology. *Current opinion in biotechnology*, *17*(5), 488–492.

[13] Zhang, S.-Y., & Liu, S.-L. (2013). Bioinformatics. In S. Maloy & K. Hughes (Eds.), *Brenner's encyclopedia of genetics (second edition)* (Second Edition, pp. 338–340). Academic Press.

[14] Chen, B.-S., & Wu, C.-C. (2013). Systems biology as an integrated platform for bioinformatics, systems synthetic biology, and systems metabolic engineering. *Cells*, *2*(4), 635–688.

[15] Mo, M. L., Jamshidi, N., & Palsson, B. Ø. (2007). A genome-scale, constraint-based approach to systems biology of human metabolism. *Molecular Biosystems*, *3*(9), 598–603.

[16] Stephanopoulos, G. (1999). Metabolic fluxes and metabolic engineering. *Metabolic engineering*, *1*(1), 1–11.

[17] Blazeck, J., & Alper, H. (2010). Systems metabolic engineering: Genome-scale models and beyond. *Biotechnology Journal*, *5*(7), 647–659.

[18] Palsson, B. (2009). Metabolic systems biology. *FEBS letters*, *583*(24), 3900–3904.

[19] Kim, W. J., Kim, H. U., & Lee, S. Y. (2017). Current state and applications of microbial genome-scale metabolic models. *Current Opinion in Systems Biology*, *2*, 10–18.

[20] Khandelwal, R. A., Olivier, B. G., Röling, W. F., Teusink, B., & Bruggeman, F. J. (2013). Community flux balance analysis for microbial consortia at balanced growth. *PloS one*, *8*(5), e64567.

[21] Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., & Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Current opinion in biotechnology*, *23*(4), 617–623.

[22] Khodaee, S., Asgari, Y., Totonchi, M., & Karimi-Jafari, M. H. (2020). Imm1865: A new reconstruction of mouse genome-scale metabolic model. *Scientific reports*, *10*(1), 1–13.

[23] Durot, M., Bourguignon, P.-Y., & Schachter, V. (2008). Genome-scale models of bacterial metabolism: Reconstruction and applications. *FEMS microbiology reviews*, *33*(1), 164–190.

[24] Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., & Chen, M. (2013). Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. *Applied microbiology and biotechnology*, *97*(2), 519–539.

[25] Park, J. H., Lee, K. H., Kim, T. Y., & Lee, S. Y. (2007). Metabolic engineering of *Escherichia coli* for the production of l-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proceedings of the national academy of sciences*, *104*(19), 7797–7802.

[26] Edwards, J. S., & Palsson, B. O. (1999). Systems properties of the *Haemophilus influenzae*Rd metabolic genotype. *Journal of Biological Chemistry*, *274*(25), 17410–17416.

[27] Notebaart, R. A., Van Enckevort, F. H., Francke, C., Siezen, R. J., & Teusink, B. (2006). Accelerating the reconstruction of genome-scale metabolic networks. *BMC bioinformatics*, *7*(1), 1–10.

[28] Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., & Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, *7*(2), 129–143.

[29] Baart, G. J., & Martens, D. E. (2012). Genome-scale metabolic models: Reconstruction and analysis. *Neisseria meningitidis* (pp. 107–126). Springer.

[30] Edwards, J., & Palsson, B. (2000). The *Escherichia coli* mg1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, *97*(10), 5528–5533.

[31] Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of *Escherichia coli* k-12 (ijr904 gsm/gpr). *Genome biology*, *4*(9), R54.

[32] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., & Palsson, B. Ø. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Molecular systems biology*, *3*(1), 121.

[33] Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular systems biology*, *7*(1), 535.

[34] Biggs, M. B., Medlock, G. L., Kolling, G. L., & Papin, J. A. (2015). Metabolic network modeling of microbial communities. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *7*(5), 317–334.

[35] Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018b). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic acids research*, *46*(15), 7542–7553.

[36] Mendoza, S. N., Olivier, B. G., Molenaar, D., & Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome biology*, *20*(1), 1–20.

[37] Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A. et al. (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, *19*(4), 524–531.

[38] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). Kegg: New perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*(D1), D353–D361.

[39] Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., & Schomburg, D. (2016). Brenda in 2017: New perspectives and new tools in brenda. *Nucleic acids research*, gkw952.

[40] Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T. et al. (2019). Database resources of the national center for biotechnology information. *Nucleic acids research*, *47*(Database issue), D23.

[41] Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Katta, H. Y., Mojica, A., Chen, I.-M. A., Kyrpides, N. C., & Reddy, T. (2019). Genomes OnLine database (GOLD) v. 7: Updates and new features. *Nucleic acids research*, *47*(D1), D649–D659.

[42] Webb, E. C. et al. (1992). *Enzyme nomenclature 1992. recommendations of the nomenclature committee of the international union of biochemistry and molecular biology on the nomenclature and classification of enzymes.* Academic Press.

[43] Saier Jr, M. H., Tran, C. V., & Barabote, R. D. (2006). TCDB: The Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research*, *34*(suppl_1), D181–D186.

[44] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403–410.

[45] Karp, P. D., Latendresse, M., Paley, S. M., Krummenacker, M., Ong, Q. D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P. et al. (2016). Pathway Tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Briefings in bioinformatics*, *17*(5), 877–890.

[46] Karp, P. D., Paley, S., & Romero, P. (2002). The pathway tools software. *Bioinformatics*, *18*(suppl_1), S225–S232.

[47] Pinney, J. W., Shirley, M. W., McConkey, G. A., & Westhead, D. R. (2005). MetaSHARK: Software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic acids research*, *33*(4), 1399–1409.

[48] Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2010). Entrez Gene: Gene-centered information at NCBI. *Nucleic acids research*, *39*(suppl_1), D52–D57.

[49] Saier Jr, M. H., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., & Moreno-Hagelsieb, G. (2016). The transporter classification database (TCDB): Recent advances. *Nucleic acids research*, *44*(D1), D372–D379.

[50] Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P., & Karp, P. D. (2020). The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic acids research*, *48*(D1), D445–D453.

[51] Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O., & King, Z. (2020). BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic acids research*, *48*(D1), D402–D406.

[52] Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q. et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, *20*(4), 1085–1093.

[53] Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E. et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic acids research*, *40*(W1), W597–W603.

[54] Consortium, U. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic acids research*, *47*(D1), D506–D515.

[55] Capela, J., Lagoa, D., Rodrigues, R., Cunha, E., Cruz, F., Barbosa, A., Bastos, J., Lima, D., Ferreira, E. C., Rocha, M. et al. (2021). *merlin* v4. 0: An updated platform for the reconstruction of high-quality genome-scale metabolic models. *bioRxiv*.

[56] Feng, X., Xu, Y., Chen, Y., & Tang, Y. J. (2012). MicrobesFlux: A web platform for drafting metabolic models from the KEGG database. *BMC systems biology*, *6*(1), 1–9.

[57] Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S. C., Ester, M., Foster, L. J. et al. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, *26*(13), 1608–1615.

[58] Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C., & Nakai, K. (2007). WoLF PSORT: Protein localization predictor. *Nucleic acids research*, *35*(suppl_2), W585–W587.

[59] Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., Altermann, U., Angerer, P., Ansorge, S., Balasz, K. et al. (2014). LocTree3 prediction of localization. *Nucleic acids research*, *42*(W1), W350–W355.

[60] Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martınez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M. et al. (2013). EcoCyc: Fusing model organism databases with systems biology. *Nucleic acids research*, *41*(D1), D605–D612.

[61] Brooksbank, C., Cameron, G., & Thornton, J. (2005). The European Bioinformatics Institute's data resources: Towards systems biology. *Nucleic Acids Research*, *33*(suppl_1), D46–D53.

[62] Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., & Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, *28*(9), 977–982.

[63] Seaver, S. M., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E. et al. (2021). The ModelSEED biochemistry database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic acids research*, *49*(D1), D575–D588.

[64] Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C., & Rocha, M. (2010). OptFlux: An open-source software platform for *in silico* metabolic engineering. *BMC systems biology*, *4*(1), 1–12.

[65] Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V. et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nature protocols*, *14*(3), 639–702.

[66] Gudmundsson, S., & Thiele, I. (2010). Computationally efficient flux variability analysis. *BMC bioinformatics*, *11*(1), 489.

[67] Veras, H. C., Campos, C. G., Nascimento, I. F., Abdelnur, P. V., Almeida, J. R., & Parachin, N. S. (2019). Metabolic flux analysis for metabolome data validation of naturally xylose-fermenting yeasts. *BMC biotechnology*, *19*(1), 1–14.

[68] Raman, K., & Chandra, N. (2009). Flux balance analysis of biological systems: Applications and challenges. *Briefings in bioinformatics*, *10*(4), 435–449.

[69] Oberhardt, M. A., Chavali, A. K., & Papin, J. A. (2009). Flux balance analysis: Interrogating genome-scale metabolic networks. *Systems biology* (pp. 61–80). Springer.

[70] Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature biotechnology*, *28*(3), 245–248.

[71] Jenior, M. L., Moutinho Jr, T. J., Dougherty, B. V., & Papin, J. A. (2020). Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLOS Computational Biology*, *16*(4), e1007099.

[72] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D. et al. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, *6*(1), 390.

[73] Segre, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, *99*(23), 15112–15117.

[74] Oyetunde, T., Czajka, J., Wu, G., Lo, C., & Tang, Y. (2017). Metabolite patterns reveal regulatory responses to genetic perturbations. *arXiv preprint arXiv:1701.01744*.

[75] Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the national academy of sciences*, *102*(21), 7695–7700.

[76] Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome biology*, *20*(1), 121.

[77] Dias, O., Rocha, M., Ferreira, E. C., & Rocha, I. (2018). Reconstructing high-quality large-scale metabolic models with *merlin*. *Metabolic network reconstruction and modeling* (pp. 1–36). Springer.

[78] Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., Mendoza, S. N., Carrier, G., Dameron, O., Guillaudeux, N. et al. (2018). Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models. *PLoS computational biology*, *14*(5), e1006146.

[79] Hanemaaijer, M., Olivier, B. G., Röling, W. F., Bruggeman, F. J., & Teusink, B. (2017). Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS One*, *12*(3), e0173183.

[80] Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., & Kerkhoven, E. J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLOS Computational Biology*, *14*(10), 1–17.

[81] Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., Schaff, J. C., Smith, L. P., & Wilkinson, D. J. (2015). The Systems Biology Markup language (SBML): Language specification for level 3 version 1 core. *Journal of integrative bioinformatics*, *12*(2), 382–549.

[82] Hucka, M., Finney, A., Hoops, S., Keating, S., & Le Novere, N. (2007). Systems biology markup language (SBML) level 2: Structures and facilities for model definitions. *Nature Precedings*, 1–1.

[83] Olivier, B. G., & Bergmann, F. T. (2018). SBML level 3 package: Flux balance constraints version 2. *Journal of integrative bioinformatics*, *15*(1).

[84] Hucka, M., & Smith, L. P. (2016). SBML Level 3 package: Groups, Version 1 Release 1. *Journal of integrative bioinformatics*, *13*(3), 8–29.

[85] Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P. et al. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature biotechnology*, *23*(12), 1509–1515.

[86] Latendresse, M. (2014). Efficiently gap-filling reaction networks. *BMC bioinformatics*, *15*(1), 1–8.

[87] Thiele, I., Vlassis, N., & Fleming, R. M. (2014). FastGapFill: Efficient gap filling in metabolic networks. *Bioinformatics*, *30*(17), 2529–2531.

[88] Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular biology and evolution*, *34*(8), 2115–2122.

[89] Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763.

[90] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M. et al. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC genomics*, *9*(1), 1–15.

[91] Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic acids research*, *36*(suppl_2), W5–W9.

[92] Daniel Machado, E. M. B. L. (2019). ReFramed documentation — ReFramed 1.1.0 documentation https://reframed.readthedocs.io/en/latest/ [(Accessed on 09/12/2020)].

[93] Pereira, V., Cruz, F., & Rocha, M. (2021). MEWpy: A computational strain optimization workbench in Python. *Bioinformatics*.

[94] Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: Constraints-based reconstruction and analysis for python. *BMC systems biology*, *7*(1), 74.

[95] Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, *84*(6), 647–657.

[96] Terzer, M., & Stelling, J. (2008). Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, *24*(19), 2229–2235.

[97] Kitano, H., Funahashi, A., Matsuoka, Y., & Oda, K. (2005). Using process diagrams for the graphical representation of biological networks. *Nature biotechnology*, *23*(8), 961–966.

[98] Chan, S. H. J., Simons, M. N., & Maranas, C. D. (2017). SteadyCom: Predicting microbial abundances while ensuring community stability. *PLoS computational biology*, *13*(5), e1005539.

[99] De Roos, J., Van der Veken, D., & De Vuyst, L. (2019). The interior surfaces of wooden barrels are an additional microbial inoculation source for lambic beer production. *Applied and environmental microbiology*, *85*(1).

[100] Spitaels, F., Wieme, A. D., Janssens, M., Aerts, M., Daniel, H.-M., Van Landschoot, A., De Vuyst, L., & Vandamme, P. (2014). The microbial diversity of traditional spontaneously fermented lambic beer. *PloS one*, *9*(4), e95384.

[101] Van Wyk, S., & Silva, F. V. (2016). Impedance technology reduces the enumeration time of *Brettanomyces* yeast during beer fermentation. *Biotechnology journal*, *11*(12), 1667–1672.

[102] De Roos, J., & De Vuyst, L. (2018). Acetic acid bacteria in fermented foods and beverages. *Current opinion in biotechnology*, *49*, 115–119.

[103] Steensels, J., Daenen, L., Malcorps, P., Derdelinckx, G., Verachtert, H., & Verstrepen, K. J. (2015). *Brettanomyces* yeasts—from spoilage organisms to valuable contributors to industrial fermentations. *International journal of food microbiology*, *206*, 24–38.

[104] Van Oevelen, D., Spaepen, M., Timmermans, P., & Verachtert, H. (1977). Microbiological aspects of spontaneous wort fermentation in the production of lambic and gueuze. *Journal of the Institute of Brewing*, *83*(6), 356–360.

[105] GARVIE, E. I. (1974). Nomenclatural problems of the pediococci. request for an opinion. *International Journal of Systematic and Evolutionary Microbiology*, *24*(2), 301–306.

[106] Commission, J. et al. (1976). Opinion 52. Conservation of the generic name *Pediococcus* Claussen with the type species *Pediococcus damnosus* Claussen. *International Journal of Systematic Bacteriology*, *26*, 292.

[107] Kufferath, H., & Van Laer, M. (1921). Études sur les levures du lambic. *Bull Soc Chim Belgique*, *30*, 270–276.

[108] Thompson Witrick, K., Duncan, S. E., Hurley, K. E., & O'Keefe, S. F. (2017). Acid and volatiles of commercially-available lambic beers. *Beverages*, *3*(4), 51.

[109] Kochláňová, T., Kij, D., Kopecká, J., Kubizniaková, P., & Matoulková, D. (2016). Non-*Saccharomyces* yeasts and their importance in the brewing industry part i-*Brettanomyces* (*Dekkera*). *KVASNÝ PRMYSL*, *62*(7-8), 198–205.

[110] Chapot-Chartier, M.-P., & Kulakauskas, S. (2014). Cell wall structure and function in lactic acid bacteria. *Microbial cell factories*, *13*(S1), S9.

[111] Mokoena, M. P. (2017). Lactic acid bacteria and their bacteriocins: Classification, biosynthesis and applications against uropathogens: A mini-review. *Molecules*, *22*(8), 1255.

[112] Sun, Z., Yu, J., Dan, T., Zhang, W., & Zhang, H. (2014). Phylogenesis and evolution of lactic acid bacteria. *Lactic acid bacteria* (pp. 1–101). Springer.

[113] Wu, C., Huang, J., & Zhou, R. (2017). Genomics of lactic acid bacteria: Current status and potential applications. *Critical Reviews in Microbiology*, *43*(4), 393–404.

[114] Oliveira, A. P., Nielsen, J., & Förster, J. (2005). Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC microbiology*, *5*(1), 39.

[115] Taxonomy browser (*Lactobacillus*) [(Accessed on 08/24/2020)]. (n.d.).

[116] Taxonomy browser (*Lactococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[117] Taxonomy browser (*Leuconostoc*) [(Accessed on 08/24/2020)]. (n.d.).

[118] Taxonomy browser (*Pediococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[119] Taxonomy browser (*Streptococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[120] Taxonomy browser (*Aerococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[121] Taxonomy browser (*Alloiococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[122] Taxonomy browser (*Carnobacterium*) [(Accessed on 08/24/2020)]. (n.d.).

[123] Taxonomy browser (*Dolosigranulum*) [(Accessed on 08/24/2020)]. (n.d.).

[124] Taxonomy browser (*Enterococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[125] Taxonomy browser (*Oenococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[126] Taxonomy browser (*Tetragenococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[127] Taxonomy browser (*Vagococcus*) [(Accessed on 08/24/2020)]. (n.d.).

[128] Taxonomy browser (*Weissella*) [(Accessed on 08/24/2020)]. (n.d.).

[129] Papagianni, M., & Anastasiadou, S. (2009). Pediocins: The bacteriocins of pediococci. sources, production, properties and applications. *Microbial cell factories*, *8*(1), 1–16.

[130] Taxonomy browser (*Pediococcus*) [(Accessed on 09/01/2020)]. (n.d.).

[131] Vandamme, P., De Bruyne, K., Pot, B., Holzapfel, W., & Wood, B. (2014). Phylogenetics and systematics. *Lactic Acid Bacteria: Biodiversity and Taxonomy. John Wiley & Sons, Ltd., New York*, 31–44.

[132] Franz, C., Endo, A., Abriouel, H., Reenen, C. A. V., Gálvez, A., & Dicks, L. M. (2014). The genus *Pediococcus*. *Lactic acid Bacteria: biodiversity and taxonomy*, 359–376.

[133] Wassenaar, T. M., & Lukjancenko, O. (2014). Comparative genomics of *Lactobacillus* and other lab. *Lactic Acid Bacteria: Biodiversity and Taxonomy*, 55–69.

[134] Behr, J., Geissler, A. J., Schmid, J., Zehe, A., & Vogel, R. F. (2016). The identification of novel diagnostic marker genes for the detection of beer spoiling *Pediococcus damnosus* strains using the BlAst diagnostic gene findER. *PLoS One*, *11*(3), e0152747.

[135] Kajala, I., Bergsveinson, J., Friesen, V., Redekop, A., Juvonen, R., Storgårds, E., & Ziola, B. (2018). *Lactobacillus backii* and *Pediococcus damnosus* isolated from 170-year-old beer recovered from a shipwreck lack the metabolic activities required to grow in modern lager beer. *FEMS microbiology ecology*, *94*(1), fix152.

[136] Snauwaert, I., Stragier, P., De Vuyst, L., & Vandamme, P. (2015). Comparative genome analysis of *Pediococcus damnosus* lmg 28219, a strain well-adapted to the beer environment. *Bmc Genomics*, *16*(1), 1–12.

[137] Holzapfel, W. H., Franz, C., Ludwig, W., Back, W., & Dicks, L. M. (2006). The genera *Pediococcus* and *Tetragenococcus*. *Prokaryotes*, *4*, 229–266.

[138] Carr, F. J., Chill, D., & Maida, N. (2002). The lactic acid bacteria: A literature survey. *Critical reviews in microbiology*, *28*(4), 281–370.

[139] Wade, M., Strickland, M. T., Osborne, J. P., & Edwards, C. G. (2019). Role of *Pediococcus* in winemaking. *Australian journal of grape and wine research*, *25*(1), 7–24.

[140] De Roos, J., Verce, M., Weckx, S., & De Vuyst, L. (2020). Temporal shotgun metagenomics revealed the potential metabolic capabilities of specific microorganisms during lambic beer production. *Frontiers in Microbiology*, *11*, 1692.

[141] Jensen, E. M., & Seeley, H. W. (1954). The nutrition and physiology of the genus *Pediococcus*. *Journal of bacteriology*, *67*(4), 484.

[142] Solberg, O., & Clausen, O. (1973). Vitamin requirements of certain pediococci isolated from brewery products. *Journal of the Institute of Brewing*, *79*(3), 231–237.

[143] Zhou, Y., Cui, Y., & Qu, X. (2019). Exopolysaccharides of lactic acid bacteria: Structure, bioactivity and associations: A review. *Carbohydrate polymers*, *207*, 317–332.

[144] Giacometti, A., Cirioni, O., Barchiesi, F., Fortuna, M., & Scalise, G. (1999). *In-vitro* activity of cationic peptides alone and in combination with clinically used antimicrobial agents against *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy*, *44*(5), 641–645.

[145] Brumfitt, W., Salton, M. R., & Hamilton-Miller, J. M. (2002). Nisin, alone and combined with peptidoglycan-modulating antibiotics: Activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci. *Journal of antimicrobial chemotherapy*, *50*(5), 731–734.

[146] Wu, J., Hu, S., & Cao, L. (2007). Therapeutic effect of nisin z on subclinical mastitis in lactating cows. *Antimicrobial agents and chemotherapy*, *51*(9), 3131–3135.

[147] Delcour, J., Ferain, T., Deghorain, M., Palumbo, E., & Hols, P. (1999). The biosynthesis and functionality of the cell-wall of lactic acid bacteria. *Lactic acid bacteria: genetics, metabolism and applications*, 159–184.

[148] Fujii, T., Nakashima, K., & Hayashi, N. (2005). Random amplified polymorphic DNA-PCR based cloning of markers to identify the beer-spoilage strains of *Lactobacillus brevis*, *Pediococcus damnosus*, *Lactobacillus collinoides* and *Lactobacillus coryniformis*. *Journal of applied microbiology*, *98*(5), 1209–1220.

[149] White, P. (1968). A comparison of the cell walls of *Pediococcus cerevisiae* and of a substrain that requires methicillin for growth. *Microbiology*, *50*(1), 107–120.

[150] Sanalibaba, P., & Çakmak, G. A. (2016). Exopolysaccharides production by lactic acid bacteria. *Appl. Microbiol. Open Access*, *2*(1000115).

[151] Dueñas-Chasco, M. T., Rodrıguez-Carvajal, M. A., Mateo, P. T., Franco-Rodrıguez, G., Espartero, J., Irastorza-Iribas, A., & Gil-Serrano, A. M. (1997). Structural analysis of the exopolysaccharide produced by *Pediococcus damnosus* 2.6. *Carbohydrate research*, *303*(4), 453–458.

[152] Lambo-Fodje, A., Leeman, M., Wahlund, K.-G., Nyman, M., Öste, R., & Larsson, H. (2007). Molar mass and rheological characterisation of an exopolysaccharide from *Pediococcus damnosus* 2.6. *Carbohydrate polymers*, *68*(3), 577–586.

[153] Liu, S.-Q. (2002). Malolactic fermentation in wine–beyond deacidification. *Journal of applied microbiology*, *92*(4), 589–601.

[154] Smith, B. D., & Divol, B. (2016). *Brettanomyces bruxellensis*, a survivalist prepared for the wine apocalypse and other beverages. *Food microbiology*, *59*, 161–175.

[155] Avramova, M., Cibrario, A., Peltier, E., Coton, M., Coton, E., Schacherer, J., Spano, G., Capozzi, V., Blaiotta, G., Salin, F. et al. (2018). *Brettanomyces bruxellensis* population survey reveals a diploid-triploid complex structured according to substrate of isolation and geographical distribution. *Scientific reports*, *8*(1), 1–13.

[156] de Barros Pita, W., Teles, G. H., Peña-Moreno, I. C., da Silva, J. M., Ribeiro, K. C., & de Morais Junior, M. A. (2019). The biotechnological potential of the yeast *Dekkera bruxellensis*. *World Journal of Microbiology and Biotechnology*, *35*(7), 103.

[157] Taxonomy browser (*Brettanomyces*) [(Accessed on 09/05/2020)]. (n.d.).

[158] Taxonomy browser (*Brettanomyces anomalus*) [(Accessed on 09/05/2020)]. (n.d.).

[159] Taxonomy browser (*Brettanomyces bruxellensis*) [(Accessed on 09/05/2020)]. (n.d.).

[160] Taxonomy browser (*Brettanomyces custersianus*) [(Accessed on 09/05/2020)]. (n.d.).

[161] Taxonomy browser (*Brettanomyces naardenensis*) [(Accessed on 09/05/2020)]. (n.d.).

[162] Cai, J., Roberts, I. N., & COLLINS, M. D. (1996). Phylogenetic relationships among members of the ascomycetous yeast genera *Brettanomyces*, *Debaryomyces*, *Dekkera*, and *Kluyveromyces* deduced by small-subunit rrna gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, *46*(2), 542–549.

[163] Roach, M. J., & Borneman, A. R. (2020). New genome assemblies reveal patterns of domestication and adaptation across *Brettanomyces* (*Dekkera*) species. *BMC genomics*, *21*(1), 1–14.

[164] Varela, C., Bartel, C., Onetto, C., & Borneman, A. (2020). Targeted gene deletion in *Brettanomyces bruxellensis* with an expression-free crispr-cas9 system. *Applied Microbiology and Biotechnology*, *104*(16), 7105–7115.

[165] Haase, M. A., Kominek, J., Opulente, D. A., Shen, X.-X., LaBella, A. L., Zhou, X., DeVirgilio, J., Hulfachor, A. B., Kurtzman, C. P., Rokas, A. et al. (2021). Repeated horizontal gene transfer of *GAL*actose metabolism genes violates dollo's law of irreversible loss. *Genetics*, *217*(2), iyaa012.

[166] Curtin, C. D., Borneman, A. R., Chambers, P. J., & Pretorius, I. S. (2012). De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *Dekkera bruxellensis* awri1499. *PLOS one*, *7*(3), e33840.

[167] Piškur, J., Ling, Z., Marcet-Houben, M., Ishchuk, O. P., Aerts, A., LaButti, K., Copeland, A., Lindquist, E., Barry, K., Compagno, C. et al. (2012). The genome of wine yeast *Dekkera bruxellensis* provides a tool to explore its food-related properties. *International journal of food microbiology*, *157*(2), 202–209.

[168] Genome list - genome - ncbi [(Accessed on 07/01/2021)]. (n.d.).

[169] Oelofse, A., Lonvaud-Funel, A., & Du Toit, M. (2009). Molecular identification of *Brettanomyces bruxellensis* strains isolated from red wines and volatile phenol production. *Food microbiology*, *26*(4), 377–385.

[170] Kheir, J., Salameh, D., Strehaiano, P., Brandam, C., & Lteif, R. (2013). Impact of volatile phenols and their precursors on wine quality and control measures of brettanomyces/dekkera yeasts. *European Food Research and Technology*, *237*(5), 655–671.

[171] Dias, L., Pereira-da-Silva, S., Tavares, M., Malfeito-Ferreira, M., & Loureiro, V. (2003). Factors affecting the production of 4-ethylphenol by the yeast *Dekkera bruxellensis* in enological conditions. *Food Microbiology*, *20*(4), 377–384.

[172] Garcia-Rubio, R., de Oliveira, H. C., Rivera, J., & Trevijano-Contador, N. (2020). The fungal cell wall: *Candida*, *Cryptococcus*, and *Aspergillus* species. *Frontiers in microbiology*, *10*, 2993.

[173] Orlean, P. (2012). Architecture and biosynthesis of the *Saccharomyces cerevisiae* cell wall. *Genetics*, *192*(3), 775–818.

[174] Klis, F. M., Boorsma, A., & De Groot, P. W. (2006). Cell wall construction in *Saccharomyces cerevisiae*. *Yeast*, *23*(3), 185–202.

[175] Lipke, P. N., & Ovalle, R. (1998). Cell wall architecture in yeast: New structure and new challenges. *Journal of bacteriology*, *180*(15), 3735–3740.

[176] Brown, H. E., Esher, S. K., & Alspaugh, J. A. (2019). Chitin: A "hidden figure" in the fungal cell wall. *The Fungal Cell Wall*, 83–111.

[177] Aimanianda, V., Clavaud, C., Simenel, C., Fontaine, T., Delepierre, M., & Latgé, J.-P. (2009). Cell wall $b\beta$-(1, 6)-glucan of *Saccharomyces cerevisiae*: Structural characterization and in situ synthesis. *Journal of Biological Chemistry*, *284*(20), 13401–13412.

[178] Buffie, C. G., Bucci, V., Stein, R. R., McKenney, P. T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A. et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to clostridium difficile. *Nature*, *517*(7533), 205–208.

[179] Bosi, E., Bacci, G., Mengoni, A., & Fondi, M. (2017). Perspectives and challenges in microbial communities metabolic modeling. *Frontiers in genetics*, *8*, 88.

[180] Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A., & Stahl, D. A. (2007). Metabolic modeling of a mutualistic microbial community. *Molecular systems biology*, *3*(1), 92.

[181] Hellweger, F. L., Clegg, R. J., Clark, J. R., Plugge, C. M., & Kreft, J.-U. (2016). Advancing microbial sciences by individual-based modelling. *Nature Reviews Microbiology*, *14*(7), 461–471.

[182] Klitgord, N., & Segre, D. (2010). The importance of compartmentalization in metabolic flux models: Yeast as an ecosystem of organelles. *Genome informatics 2009: Genome informatics series vol. 22* (pp. 41–55). World Scientific.

[183] Klitgord, N., & Segrè, D. (2010). Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol*, *6*(11), e1001002.

[184] Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø., & Jamshidi, N. (2010). Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular systems biology*, *6*(1), 422.

[185] Shoaie, S., Karlsson, F., Mardinoglu, A., Nookaew, I., Bordel, S., & Nielsen, J. (2013). Understanding the interactions between bacteria in the human gut through metabolic modeling. *Scientific reports*, *3*(1), 1–10.

[186] Zomorrodi, A. R., & Maranas, C. D. (2012). OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol*, *8*(2), e1002363.

[187] El-Semman, I. E., Karlsson, F. H., Shoaie, S., Nookaew, I., Soliman, T. H., & Nielsen, J. (2014). Genome-scale metabolic reconstructions of *Bifidobacterium adolescentis* l2-32 and *Faecalibacterium prausnitzii* a2-165 and their interaction. *BMC systems biology*, *8*(1), 1–11.

[188] Mahadevan, R., Edwards, J. S., & Doyle III, F. J. (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical journal*, *83*(3), 1331–1340.

[189] Hanly, T. J., & Henson, M. A. (2011). Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnology and bioengineering*, *108*(2), 376–385.

[190] Hanly, T. J., & Henson, M. A. (2014). Dynamic model-based analysis of furfural and hmf detoxification by pure and mixed batch cultures of *S. cerevisiae* and *S. stipitis*. *Biotechnology and bioengineering*, *111*(2), 272–284.

[191] Biggs, M. B., & Papin, J. A. (2013). Novel multiscale modeling tool applied to *Pseudomonas aeruginosa* biofilm formation. *PLoS One*, *8*(10), e78011.

[192] Harcombe, W. R., Riehl, W. J., Dukovski, I., Granger, B. R., Betts, A., Lang, A. H., Bonilla, G., Kar, A., Leiby, N., Mehta, P. et al. (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell reports*, *7*(4), 1104–1115.

[193] Chiu, H.-C., Levy, R., & Borenstein, E. (2014). Emergent biosynthetic capacity in simple microbial communities. *PLoS Comput Biol*, *10*(7), e1003695.

[194] Tzamali, E., Poirazi, P., Tollis, I. G., & Reczko, M. (2011). A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities. *BMC systems biology*, *5*(1), 1–15.

[195] Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., & Lovley, D. R. (2011). Genome-scale dynamic modeling of the competition between *Rhodoferax* and *Geobacter* in anoxic subsurface environments. *The ISME journal*, *5*(2), 305–316.

[196] Klitgord, N., & Segrè, D. (2011). Ecosystems biology of microbial metabolism. *Current opinion in biotechnology*, *22*(4), 541–546.

[197] Greenblum, S., Turnbaugh, P. J., & Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences*, *109*(2), 594–599.

[198] Tobalina, L., Bargiela, R., Pey, J., Herbst, F.-A., Lores, I., Rojo, D., Barbas, C., Peláez, A. I., Sánchez, J., von Bergen, M. et al. (2015). Context-specific metabolic network reconstruction of a naphthalene-degrading bacterial community guided by metaproteomic data. *Bioinformatics*, *31*(11), 1771–1779.

[199] Taffs, R., Aston, J. E., Brileya, K., Jay, Z., Klatt, C. G., McGlynn, S., Mallette, N., Montross, S., Gerlach, R., Inskeep, W. P. et al. (2009). *In silico* approaches to study mass and energy flows in microbial consortia: A syntrophic case study. *BMC systems biology*, *3*(1), 1–16.

[200] Borenstein, E., Kupiec, M., Feldman, M. W., & Ruppin, E. (2008). Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences*, *105*(38), 14482–14487.

[201] Levy, R., & Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, *110*(31), 12804–12809.

[202] Handorf, T., Ebenhöh, O., & Heinrich, R. (2005). Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *Journal of molecular evolution*, *61*(4), 498–512.

[203] Christian, N., Handorf, T., & Ebenhöh, O. (2007). Metabolic synergy: Increasing biosynthetic capabilities by network cooperation. *Genome informatics 2007: Genome informatics series vol. 18* (pp. 320–329). World Scientific.

[204] Steinway, S. N., Biggs, M. B., Loughran Jr, T. P., Papin, J. A., & Albert, R. (2015). Inference of network dynamics and metabolic interactions in the gut microbiome. *PLoS computational biology*, *11*(6), e1004338.

[205] Bartell, J. A., Yen, P., Varga, J. J., Goldberg, J. B., & Papin, J. A. (2014). Comparative metabolic systems analysis of pathogenic *Burkholderia*. *Journal of bacteriology*, *196*(2), 210.

[206] Vinay-Lara, E., Hamilton, J. J., Stahl, B., Broadbent, J. R., Reed, J. L., & Steele, J. L. (2014). Genome–scale reconstruction of metabolic networks of *Lactobacillus casei* atcc 334 and 12a. *PloS one*, *9*(11), e110785.

[207] Koch, S., Kohrs, F., Lahmann, P., Bissinger, T., Wendschuh, S., Benndorf, D., Reichl, U., & Klamt, S. (2019). RedCom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLoS computational biology*, *15*(2), e1006759.

[208] Hjersted, J., & Henson, M. (2009). Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. *IET systems biology*, *3*(3), 167–179.

[209] Henson, M. A., & Hanly, T. J. (2014). Dynamic flux balance analysis for synthetic microbial communities. *IET systems biology*, *8*(5), 214–229.

[210] Gomez, J. A., Höffner, K., & Barton, P. I. (2014). DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis. *BMC bioinformatics*, *15*(1), 409.

[211] Popp, D., & Centler, F. (2020). $m\mu$BialSim: Constraint-based dynamic simulation of complex microbiomes. *Frontiers in Bioengineering and Biotechnology*, *8*.

[212] Diener, C., Gibbons, S. M., & Resendis-Antonio, O. (2020). MICOM: Metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *MSystems*, *5*(1).

[213] Koch, S., Benndorf, D., Fronk, K., Reichl, U., & Klamt, S. (2016). Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnology for biofuels*, *9*(1), 1–16.

[214] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., & Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, *112*(20), 6449–6454.

[215] Cruz, F., Capela, J., Ferreira, E. C., Rocha, M., & Dias, O. (2021). *BioISO*: An objective-oriented application for assisting the curation of genome-scale metabolic models. *bioRxiv*.

[216] Lagoa, D., Faria, J. L., Liu, F., Cunha, E., Henry, C., & Dias, O. (2021). Transyt, the transport systems tracker. *bioRxiv*.

[217] Teusink, B., Wiersma, A., Molenaar, D., Francke, C., De Vos, W. M., Siezen, R. J., & Smid, E. J. (2006). Analysis of growth of *Lactobacillus plantarum* wcfs1 on a complex medium using a genome-scale metabolic model. *Journal of Biological Chemistry*, *281*(52), 40041–40048.

[218] Mishra, P., Park, G.-Y., Lakshmanan, M., Lee, H.-S., Lee, H., Chang, M. W., Ching, C. B., Ahn, J., & Lee, D.-Y. (2016). Genome-scale metabolic modeling and in silico analysis of lipid accumulating yeast *Candida tropicalis* for dicarboxylic acid production. *Biotechnology and bioengineering*, *113*(9), 1993–2004.

[219] Xu, N., Liu, L., Zou, W., Liu, J., Hua, Q., & Chen, J. (2013). Reconstruction and analysis of the genome-scale metabolic network of *Candida glabrata*. *Molecular BioSystems*, *9*(2), 205–216.

[220] Dias, O., Pereira, R., Gombert, A. K., Ferreira, E. C., & Rocha, I. (2014). Iod907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnology journal*, *9*(6), 776–790.

[221] Mo, M. L., Palsson, B. Ø., & Herrgård, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC systems biology*, *3*(1), 1–17.

[222] Santos, S., & Rocha, I. (2016). Estimation of biomass composition from genomic and transcriptomic information. *Journal of integrative bioinformatics*, *13*(2), 1–14.

[223] Xavier, J. C., Patil, K. R., & Rocha, I. (2017). Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metabolic engineering*, *39*, 200–208.

[224] Xavier, J. C., Patil, K. R., & Rocha, I. (2018). Metabolic models and gene essentiality data reveal essential and conserved metabolism in prokaryotes. *PLoS computational biology*, *14*(11), e1006556.

[225] Beverly, M. B., Basile, F., & Voorhees, K. J. (1997). Fatty acid analysis of beer spoiling microorganisms using pyrolysis mass spectrometry. *Journal of the American Society of Brewing Chemists*, *55*(2), 79–82.

[226] Von Cosmos, N. H., & Edwards, C. G. (2016). Use of nutritional requirements for *Brettanomyces bruxellensis* to limit infections in wine. *Fermentation*, *2*(3), 17.

[227] Hammes, W. P., & Hertel, C. (2006). The genera *Lactobacillus* and *Carnobacterium*. *Prokaryotes*, *4*, 320–403.

[228] Smith, M. T. (1998). *Dekkera* van der walt. *The yeasts* (pp. 174–177). Elsevier.

[229] Crauwels, S., Van Assche, A., de Jonge, R., Borneman, A., Verreth, C., De Samblanx, G., Marchal, K., Van de Peer, Y., Willems, K., Verstrepen, K. et al. (2015). Comparative phenomics and targeted use of genomics reveals variation in carbon and nitrogen assimilation among different *Brettanomyces bruxellensis* strains. *Applied microbiology and biotechnology*, *99*(21), 9123–9134.

[230] Conterno, L., Joseph, C. L., Arvik, T. J., Henick-Kling, T., & Bisson, L. F. (2006). Genetic and physiological characterization of *Brettanomyces bruxellensis* strains isolated from wines. *American Journal of Enology and Viticulture*, *57*(2), 139–147.

[231] Borneman, A. R., Zeppel, R., Chambers, P. J., & Curtin, C. D. (2014). Insights into the *Dekkera bruxellensis* genomic landscape: Comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLoS genetics*, *10*(2), e1004161.

[232] de Souza Vandenberghe, L. P., Karp, S. G., Pagnoncelli, M. G. B., von Linsingen Tavares, M., Junior, N. L., Diestra, K. V., Viesser, J. A., & Soccol, C. R. (2020). Classification of enzymes and catalytic properties. *Biomass, biofuels, biochemicals* (pp. 11–30). Elsevier.

[233] McDonald, A. G., Boyce, S., & Tipton, K. F. (2015). Enzyme classification and nomenclature. *eLS*, 1–11.

[234] of Biochemistry, I.-I. U., & Biology, M. (2018). Translocases (ec 7): A new ec class [(Accessed on 07/15/2021)].

[235] Godoy, L., Vera-Wolf, P., Martinez, C., Ugalde, J. A., & Ganga, M. A. (2016). Comparative transcriptome assembly and genome-guided profiling for *Brettanomyces bruxellensis* lamap2480 during p-coumaric acid stress. *Scientific reports*, *6*(1), 1–13.

[236] Capusoni, C., Arioli, S., Zambelli, P., Moktaduzzaman, M., Mora, D., & Compagno, C. (2016). Effects of oxygen availability on acetic acid tolerance and intracellular ph in *Dekkera bruxellensis*. *Applied and environmental microbiology*, *82*(15), 4673–4681.

[237] Ryan, D. G., Frezza, C., & O'Neill, L. A. (2021). Tca cycle signalling and the evolution of eukaryotes. *Current opinion in biotechnology*, *68*, 72–88.

[238] Busch, W., & Saier, M. H. (2002). The transporter classification (tc) system, 2002. *Critical reviews in biochemistry and molecular biology*, *37*(5), 287–337.

[239] Wolfersberger, M. G. (1994). Uniporters, symporters and antiporters. *The Journal of experimental biology*, *196*(1), 5–6.

[240] Locher, K. P. (2009). Structure and mechanism of atp-binding cassette transporters. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1514), 239–245.

[241] Chang, A. B., Lin, R., Studley, W. K., Tran, C. V., & Saier, M. H., Jr. (2004). Phylogeny as a guide to structure and function of membrane transport proteins. *Molecular membrane biology*, *21*(3), 171–181.

[242] Teixeira, P. (2014). *Lactobacillus brevis*. In BATT, Carl, A.; TORTORELLO, Mary-Lou (eds.) - Encyclopedia of Food Microbiology. Reino Unido: Academic Press.

[243] Caligiani, A., & Lolli, V. (2018). Cyclic fatty acids in food: An under investigated class of fatty acids. *Biochemistry and Health Benefits of Fatty Acids*, *2018*.

[244] Lolli, V., Zanardi, E., Moloney, A. P., & Caligiani, A. (2020). An overview on cyclic fatty acids as biomarkers of quality and authenticity in the meat sector. *Foods*, *9*(12), 1756.

[245] Jones, S. E., Whitehead, K., Saulnier, D., Thomas, C. M., Versalovic, J., & Britton, R. A. (2011). Cyclopropane fatty acid synthase mutants of probiotic human-derived *Lactobacillus reuteri* are defective in tnf inhibition. *Gut microbes*, *2*(2), 69–79.

[246] Buck, B., Azcarate-Peril, M., & Klaenhammer, T. (2009). Role of autoinducer-2 on the adhesion ability of *Lactobacillus acidophilus*. *Journal of applied microbiology*, *107*(1), 269–279.

[247] Henry, C. S., Zinner, J. F., Cohoon, M. P., & Stevens, R. L. (2009). I bsu1103: A new genome-scale metabolic model of *Bacillus subtilis* based on seed annotations. *Genome biology*, *10*(6), 1–15.

[248] Razes, N., Garcia-Jares, C., Larue, F., & Lonvaud-Funel, A. (1992). Differentiation between fermenting and spoilage yeasts in wine by total free fatty acid analysis. *Journal of the Science of Food and Agriculture*, *59*(3), 351–357.

[249] Galafassi, S., Toscano, M., Vigentini, I., Zambelli, P., Simonetti, P., Foschino, R., & Compagno, C. (2015). Cold exposure affects carbohydrates and lipid metabolism, and induces hog1p phosphorylation in *Dekkera bruxellensis* strain cbs 2499. *Antonie van Leeuwenhoek*, *107*(5), 1145–1153.

[250] Oosthuizen, A., Kock, J., Viljoen, B., Muller, H., & Lategan, P. (1987). The value of long-chain fatty acid composition in the identification of some brewery yeasts. *Journal of the Institute of Brewing*, *93*(3), 174–176.

[251] Conterno, L., Aprea, E., Franceschi, P., Viola, R., & Vrhovsek, U. (2013). Overview of *Dekkera bruxellensis* behaviour in an ethanol-rich environment using untargeted and targeted metabolomic approaches. *Food research international*, *51*(2), 670–678.

[252] Giovani, G., Rosi, I., & Bertuccioli, M. (2012). Quantification and characterization of cell wall polysaccharides released by non-saccharomyces yeast strains during alcoholic fermentation. *International journal of food microbiology*, *160*(2), 113–118.

[253] Perli, T., van der Vorm, D. N., Wassink, M., van den Broek, M., Pronk, J. T., & Daran, J.-M. (2021). Engineering heterologous molybdenum-cofactor-biosynthesis and nitrate-assimilation pathways enables nitrate utilization by *Saccharomyces cerevisiae*. *Metabolic Engineering*, *65*, 11–29.

[254] BILLON-GRAND, G. (1987). Minor ubiquinones of the yeast coenzyme q system: Importance in the taxonomy of the yeasts. *The Journal of General and Applied Microbiology*, *33*(5), 381–390.

[255] Shane, B. (1989). Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. *Vitamins & Hormones*, *45*, 263–335.

[256] Yan, X., Gu, S., Cui, X., Shi, Y., Wen, S., Chen, H., & Ge, J. (2019). Antimicrobial, anti-adhesive and anti-biofilm potential of biosurfactants isolated from *Pediococcus acidilactici* and *Lactobacillus plantarum* against *Staphylococcus aureus* cmcc26003. *Microbial pathogenesis*, *127*, 12–20.

[257] Chanos, P., & Mygind, T. (2016). Co-culture-inducible bacteriocin production in lactic acid bacteria. *Applied microbiology and biotechnology*, *100*(10), 4297–4308.

[258] Marbois, B., Xie, L. X., Choi, S., Hirano, K., Hyman, K., & Clarke, C. F. (2010). Para-aminobenzoic acid is a precursor in coenzyme q6 biosynthesis in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *285*(36), 27827–27838.

[259] Becker-Kettern, J., Paczia, N., Conrotte, J.-F., Kay, D. P., Guignard, C., Jung, P. P., & Linster, C. L. (2016). Saccharomyces cerevisiae forms d-2-hydroxyglutarate and couples its degradation to d-lactate formation via a cytosolic transhydrogenase*. *Journal of Biological Chemistry*, *291*(12), 6036–6058.

A

SUPPORT MATERIAL

In this additional chapter, all support data was attached.

Table S4: *P. damnosus* GSMM pathways

Pathway name	Number of reactions
2-Oxocarboxylic acid metabolism	5
Alanine, aspartate and glutamate metabolism	13
Amino sugar and nucleotide sugar metabolism	19
Aminoacyl-tRNA biosynthesis	21
Arginine and proline metabolism	6
Arginine biosynthesis	6
Biomass Pathway	13
Biosynthesis of amino acids	53
Biosynthesis of antibiotics	80
Biosynthesis of secondary metabolites	97
Butanoate metabolism	5
C5-Branched dibasic acid metabolism	2

Continued on next page

Table S4 – continued from previous page

Pathway name	Number of reactions
Carbapenem biosynthesis	3
Carbon fixation pathways in prokaryotes	14
Carbon metabolism	50
Citrate cycle (TCA cycle)	4
Cyanoamino acid metabolism	4
Cysteine and methionine metabolism	19
D-Alanine metabolism	2
D-Glutamine and D-glutamate metabolism	4
Drains pathway	87
Fatty acid biosynthesis	51
Fatty acid metabolism	44
Folate biosynthesis	12
Fructose and mannose metabolism	10
Galactose metabolism	15
Glutathione metabolism	10
Glycerolipid metabolism	14
Glycerophospholipid metabolism	12

Continued on next page

Table S4 – continued from previous page

Pathway name	Number of reactions
Glycine, serine and threonine metabolism	13
Glycolysis / Gluconeogenesis	19
Glyoxylate and dicarboxylate metabolism	10
Inositol phosphate metabolism	2
Lysine biosynthesis	14
Metabolic pathways	322
Metabolism of xenobiotics by cytochrome P450	4
Methane metabolism	9
Microbial metabolism in diverse environments	78
Monobactam biosynthesis	5
Nicotinate and nicotinamide metabolism	7
Nitrogen metabolism	3
Non enzymatic	1
One carbon pool by folate	20
Oxidative phosphorylation	1
Pantothenate and CoA biosynthesis	8
Pentose and glucuronate interconversions	4

Continued on next page

Table S4 – continued from previous page

Pathway name	Number of reactions
Pentose phosphate pathway	17
Peptidoglycan biosynthesis	11
Porphyrin and chlorophyll metabolism	3
Propanoate metabolism	3
Purine metabolism	41
Pyrimidine metabolism	54
Pyruvate metabolism	15
Riboflavin metabolism	2
Selenocompound metabolism	1
Spontaneous	13
Starch and sucrose metabolism	11
Sulfur metabolism	6
Synthesis and degradation of ketone bodies	2
Terpenoid backbone biosynthesis	10
Thiamine metabolism	7
Transporters pathway	152
Tryptophan metabolism	1

Continued on next page

Table S4 – continued from previous page

Pathway name	Number of reactions
Ubiquinone and other terpenoid-quinone biosynthesis	4
Valine, leucine and isoleucine biosynthesis	2
Valine, leucine and isoleucine degradation	2
Vitamin B6 metabolism	5

Table S5: *B. bruxellensis* GSMM pathways

Pathway name	Number of reactions
2-Oxocarboxylic acid metabolism	68
Alanine, aspartate and glutamate metabolism	29
alpha-Linolenic acid metabolism	2
Amino sugar and nucleotide sugar metabolism	23
Aminoacyl-tRNA biosynthesis	20
Arginine and proline metabolism	39
Arginine biosynthesis	22
beta-Alanine metabolism	9
Biomass Pathway	12
Biosynthesis of amino acids	118
Biosynthesis of secondary metabolites	309

Continued on next page

Table S5 – continued from previous page

Pathway name	Number of reactions
Biosynthesis of unsaturated fatty acids	26
Butanoate metabolism	16
Carbapenem biosynthesis	6
Carbon metabolism	100
Citrate cycle (TCA cycle)	29
Cysteine and methionine metabolism	32
D-Alanine metabolism	1
D-Glutamine and D-glutamate metabolism	5
Drains pathway	213
Ether lipid metabolism	2
Fatty acid biosynthesis	129
Fatty acid degradation	59
Fatty acid elongation	25
Fatty acid metabolism	143
Folate biosynthesis	25
Fructose and mannose metabolism	19
Galactose metabolism	4

Continued on next page

Table S5 – continued from previous page

Pathway name	Number of reactions
Glutathione metabolism	17
Glycerolipid metabolism	18
Glycerophospholipid metabolism	44
Glycine, serine and threonine metabolism	36
Glycolysis / Gluconeogenesis	31
Glyoxylate and dicarboxylate metabolism	27
Histidine metabolism	2
Inositol phosphate metabolism	16
Linoleic acid metabolism	1
Lipoic acid metabolism	3
Lysine biosynthesis	16
Lysine degradation	14
Metabolic pathways	863
Methane metabolism	22
Microbial metabolism in diverse environments	153
Monobactam biosynthesis	2
Nicotinate and nicotinamide metabolism	24

Continued on next page

Table S5 – continued from previous page

Pathway name	Number of reactions
Nitrogen metabolism	14
Non enzymatic	4
One carbon pool by folate	31
Oxidative phosphorylation	13
Pantothenate and CoA biosynthesis	25
Penicillin and cephalosporin biosynthesis	1
Pentose and glucuronate interconversions	7
Pentose phosphate pathway	16
Phenylalanine metabolism	1
Phenylalanine, tyrosine and tryptophan biosynthesis	22
Phenylpropanoid biosynthesis	4
Porphyrin and chlorophyll metabolism	10
Propanoate metabolism	8
Purine metabolism	80
Pyrimidine metabolism	49
Pyruvate metabolism	44
Retinol metabolism	2

Continued on next page

Table S5 – continued from previous page

Pathway name	Number of reactions
Riboflavin metabolism	15
Selenocompound metabolism	5
Sesquiterpenoid and triterpenoid biosynthesis	6
Sphingolipid metabolism	20
Spontaneous	11
Starch and sucrose metabolism	21
Steroid biosynthesis	21
Sulfur metabolism	8
Terpenoid backbone biosynthesis	15
Thiamine metabolism	10
Transporters pathway	735
Tryptophan metabolism	17
Tyrosine metabolism	6
Ubiquinone and other terpenoid-quinone biosynthesis	13
Valine, leucine and isoleucine biosynthesis	24
Valine, leucine and isoleucine degradation	11
Vitamin B6 metabolism	9

Table S1: Corrections made in genome annotation of *P. damnosus*.

Gene	E.C. number(s)	Correct annotation
ADU72_0297	2.7.1.144	2.7.1.56
ADU72_0454	3.6.1.27	3.1.3.4
ADU72_0686	1.-.-.-	1.1.1.21
ADU72_0879	3.2.1.85	3.2.1.86
ADU72_0909	3.2.-.-	3.2.2.1
ADU72_1025	1.8.1.7	1.8.1.12
ADU72_1082	1.-.-.-	1.1.1.140
ADU72_1252	2.7.1.-	2.7.1.192
ADU72_1303	1.6.5.5	2.3.1.41
ADU72_1373	1.2.3.3	2.2.1.6
ADU72_1439	5.1.3.21	5.1.3.3
ADU72_1441	3.2.1.20	3.2.1.10
ADU72_1830	4.4.1.13	2.5.1.48
ADU72_1864	2.7.8.-	2.7.8.20
ADU72_1870	2.4.1.-	2.4.1.208
BSQ38_01560	2.7.1.49, 2.7.4.7	2.5.1.3, 2.7.1.49, 2.7.4.7
BSQ38_02925	3.2.-.-	3.2.2.3
BSQ38_06370	1.6.99.3	1.11.1.1
BSQ38_06860	2.4.1.216	2.4.1.8
BSQ38_08415	4.4.1.13	4.4.1.1
WP_046870966.1		2.7.8.-
WP_046870975.1	5.1.1.21	2.6.1.13
WP_046870990.1	1.1.1.-	1.1.1.65
WP_046871029.1	3.2.1.20	3.2.1.93
WP_046871283.1	2.7.1.76	2.7.1.76, 2.7.1.74
WP_046872192.1	6.4.1.2, 6.3.4.14	6.4.1.1
WP_046872327.1	4.1.2.-	4.1.2.9
WP_075168953.1	2.7.-.-	4.1.1.1
WP_080945549.1		3.1.1.31

Table S2: Corrections made in genome annotation of *B. bruxellensis*.

Gene	E.C. number(s)	Correct annotation
BN3033_LOCUS1058	1.1.1.168	1.1.1.283
BN3033_LOCUS1081	2.5.1.-	2.5.1.141
BN3033_LOCUS1166	2.7.1.67	2.7.1.154
BN3033_LOCUS12		4.1.1.36
BN3033_LOCUS1289	2.3.1.-	2.3.1.30, 2.3.1.31
BN3033_LOCUS1423	2.4.1.15	3.1.3.12
BN3033_LOCUS15	1.14.13.-, 3.5.-.-	1.14.99.60
BN3033_LOCUS154	1.1.1.4	1.1.1.9, 1.1.1.-, 1.1.1.14
BN3033_LOCUS2062	2.3.1.-	2.3.1.20
BN3033_LOCUS2147	2.5.1.47, 2.5.1.49	5.1.1.1
BN3033_LOCUS2205	1.18.1.6	1.18.1.2
BN3033_LOCUS2259	2.5.1.16	2.5.1.22
BN3033_LOCUS2281	3.1.4.-	
BN3033_LOCUS2310	1.1.1.307	1.1.1.156
BN3033_LOCUS2567	3.5.1.-	3.5.1.-, 3.5.1.23
BN3033_LOCUS2739	4.1.99.12	3.5.4.25, 4.1.99.12
BN3033_LOCUS2825	1.1.1.3	1.1.1.3, 2.7.2.4
BN3033_LOCUS2865	1.1.1.41	1.1.1.286
BN3033_LOCUS3052	2.3.3.14	2.3.3.-, 2.3.3.14, 2.3.3.10
BN3033_LOCUS3123		2.1.2.10
BN3033_LOCUS322	2.5.1.-	2.5.1.84
BN3033_LOCUS3251	1.1.1.168	1.1.1.21
BN3033_LOCUS3395		3.6.1.13
BN3033_LOCUS3582	3.5.1.-	3.5.1.-, 3.5.1.23
BN3033_LOCUS3704	2.3.1.9	2.3.1.16, 3.1.2.1, 3.1.2.2, 3.1.2.-, 2.3.1.9
BN3033_LOCUS3748	3.1.3.64	3.1.3.95, 3.1.3.64
BN3033_LOCUS3753	1.-.-.-	1.14.18.5
BN3033_LOCUS3823	2.7.-.-	2.7.1.31
BN3033_LOCUS3853	2.10.1.1	2.10.1.1, 2.7.7.75
BN3033_LOCUS3934	3.1.1.-	3.1.1.32, 3.1.1.4
BN3033_LOCUS3959		1.5.3.17
BN3033_LOCUS4054	4.2.1.36	4.2.1.114
BN3033_LOCUS496	1.2.1.-	1.2.1.95
BN3033_LOCUS613	1.1.1.41	1.1.1.85
BN3033_LOCUS684	2.6.1.44	2.6.1.51, 2.6.1.44
BN3033_LOCUS785	1.2.1.41	1.2.1.41, 2.7.2.11
BN3033_LOCUS905	3.1.3.41	3.1.3.3, 3.1.3.74

Table S3: Minimal defined medium of *P. damnosus* and *B. bruxellensis*.

Pediococcus damnosus			**Brettanomyces bruxellensis**		
Compound	**Formula**	**KEGG ID**	**Compound**	**Formula**	**KEGG ID**
Adenine	$C_5H_5N_5$	C00147	4-Aminobenzoate	$C_7H_7NO_2$	C00568
alpha-D-Glucose	$C_6H_{12}O_6$	C00267	alpha-D-Glucose	$C_6H_{12}O_6$	C00267
Biotin	$C_{10}H_{16}N_2O_3S$	C00120	Ammonia	NH_3	C00014
Fe_3^+	Fe	C14819	Biotin	$C_{10}H_{16}N_2O_3S$	C00120
Folinic acid	$C_{20}H_{23}N_7O_7$	C03479	Fe_2^+	Fe	C14818
Glycine	$C_2H_5NO_2$	C00037	Folate	$C_{19}H_{19}N_7O_6$	C00504
Guanine	$C_5H_5N_5O$	C00242	L-Histidine	$C_6H_9N_3O_2$	C00135
L-Alanine	$C_3H_7NO_2$	C00041	Linoleate	$C_{18}H_{32}O_2$	C01595
L-Arginine	$C_6H_{14}N_4O_2$	C00062	L-Methionine	$C_5H_{11}NO_2S$	C00073
L-Aspartate	$C_4H_7NO_4$	C00049	L-Tryptophan	$C_{11}H_{12}N_2O_2$	C00078
L-Cysteine	$C_3H_7NO_2S$	C00097	Molybdate	H_2MoO_4	C06232
L-Glutamate	$C_5H_9NO_4$	C00025	myo-Inositol	$C_6H_{12}O_6$	C00137
L-Histidine	$C_6H_9N_3O_2$	C00135	Nicotinate	$C_6H_5NO_2$	C00253
L-Isoleucine	$C_6H_{13}NO_2$	C00407	Orthophosphate	H_3PO_4	C00009
L-Leucine	$C_6H_{13}NO_2$	C00123	Pantothenate	$C_9H_{17}NO_5$	C00864
L-Lysine	$C_6H_{14}N_2O_2$	C00047	Pyridoxal	$C_8H_9NO_3$	C00250
L-Methionine	$C_5H_{11}NO_2S$	C00073	Riboflavin	$C_{17}H_{20}N_4O_6$	C00255
L-Phenylalanine	$C_9H_{11}NO_2$	C00079	Sulfate	H_2SO_4	C00059
L-Proline	$C_5H_9NO_2$	C00148	Thiamine	$C_{12}H_{17}N_{4O}S$	C00378
L-Serine	$C_3H_7NO_3$	C00065			
L-Threonine	$C_4H_9NO_3$	C00188			
L-Tryptophan	$C_{11}H_{12}N_2O_2$	C00078			
L-Tyrosine	$C_9H_{11}NO_3$	C00082			
L-Valine	$C_5H_{11}NO_2$	C00183			
Nicotinate	$C_6H_5NO_2$	C00253			
Orthophosphate	H_3PO_4	C00009			
Pantothenate	$C_9H_{17}NO_5$	C00864			
Pyridoxine	$C_8H_{11}NO_3$	C00314			
Riboflavin	$C_{17}H_{20}N_4O_6$	C00255			
Thymine	$C_5H_6N_2O_2$	C00178			
Uracil	$C_4H_4N_2O_2$	C00106			
Xanthine	$C_5H_4N_4O_2$	C00385			