UNIVERSITÉ LIBRE DE BRUXELLES ULB VUB VRIJE UNIVERSITEIT BRUSSEL

Genome-scale metabolic modeling of candidate functional starter cultures for cocoa bean fermentation

ir. Rudy Pelicaen

Faculteit Wetenschappen en Bio-ingenieurswetenschappen
Onderzoeksgroep Industriële Microbiologie en Voedingsbiotechnologie
Academiejaar 2019-2020

Proefschrift voorgedragen tot het bekomen van de graad van
Doctor in de Bio-ingenieurswetenschappen

Faculté des Sciences
Unité de Chronobiologie Théorique
Année académique 2019-2020

Thèse soumise en vue de l'obtention du grade de
Docteur en Sciences

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (Ph.D.) of Bioengineering Sciences (VUB) and Sciences (ULB)

Promotors: Prof. Dr. Stefan Weckx (VUB)
and Prof. Dr. Didier Gonze (ULB)

Published by the Research Group of Industrial Microbiology and Food biotechnology (IMDO, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium) and the Unit of Theoretical Chronobiology (Université libre de Bruxelles, Boulevard du Triomphe 155, B-1050 Brussels, Belgium).

R. Pelicaen - Genome-scale metabolic modeling of candidate functional starter cultures for cocoa bean fermentation - 2020

Ph.D. thesis Vrije Universiteit Brussel, Brussels, Belgium and Université libre de Bruxelles, Brussels, Belgium - With summary in Dutch and French - 172 p.

ISBN: 9782500013933

Acknowledgements

The PhD thesis presented here is the result of a joint PhD between the Vrije Universiteit Brussel (VUB) and the Université Libre de Bruxelles (ULB), and was accomplished with the guidance and support of different people.

I'd like to thank my promotors, Prof. Dr. Stefan Weckx and Prof. Dr. Didier Gonze, for giving me the freedom to develop my research interests and letting me present the results at scientific meetings. I thank Prof. Dr. ir. Luc De Vuyst and the members of the jury, Prof. Dr. Dominque Maes, Prof. Dr. Tom Lenaerts, Prof. Dr. Peter Tompa, Prof. Dr. Laurence Van Melderen, Prof. Dr. ir. Kristel Bernaerts, and Prof. Dr. Christoph Wittmann, for critical reading of the PhD thesis.

I'm grateful to Prof. Dr. Bas Teusink, Prof. Dr. Frank Bruggeman, and Prof. Dr. Rob Van Spanning for the help and guidance during my stay at the Systems Bioinformatics group at the Vrije Universiteit Amsterdam. I'd like to thank the whole team at the Systems Bioinformatics group to involve me in the day-to-day lab activities. I also want to thank my VUB colleagues at IMDO (VUB) and ULB colleagues at IB^2. We had many fruitful discussions, laughs, and even musical explorations. I'm sure we'll stay in touch in the future. Finally, I want to acknowledge the Research Foundation Flanders (FWO-Vlaanderen) for granting me a PhD Fellowship strategic basic research.

I want to thank my family, for hearing me expand on my research, staying ubiquitously interested, and always supporting me to bring my PhD to a proper end. I'm most grateful to my fiancé Delia Mensitieri. Delia, you have supported me from the very beginning and I will always thank you for it. Together, we even took special interest in the development of our own fermentation techniques; and we found out that we may learn a great deal more from practice than from theory alone. I also want to thank your family, Delia, as they also have supported me a lot during my PhD. Finally, I thank my friends, for questioning the meaningfulness of making metabolic models of bacteria to obtain better chocolate.

Contents

Introduction

Chocolate is a widely consumed cocoa-derived food product that is appreciated by consumers for its organoleptic properties. However, few people know about the complex manufacturing process that is needed to produce chocolate. The cocoa bean fermentation process is an essential first step in the manufacturing process to obtain cured cocoa beans, the raw material for the production of cocoa-derived products. Being first in the manufacturing process related to a billion-dollar industry, the cocoa bean fermentation process has a huge economic importance for the food industry. Therefore, it is both surprising and intriguing that the cocoa bean fermentation process is until now still a spontaneous fermentation process. Candidate functional starter cultures have been isolated from spontaneous cocoa bean fermentation processes. These are single microbial strains with specific functional properties that are thought to be beneficial for the cocoa bean fermentation process and that should enable more controlled cocoa bean fermentation processes. Yet, current candidate functional starter cultures are insufficiently characterized at the metabolic level to allow their optimal use in this complex fermentation process. Therefore, it is important to increase the knowledge on the metabolism of the candidate functional starter cultures to obtain a better control over the fermentation process.

With the advent of next-generation sequencing, it has now become possible to sequence bacterial genomes in a matter of hours. The bottleneck in information retrieval has therefore shifted to the *in silico* analysis of the genome and its annotation. A step further to increase knowledge regarding the metabolism of a strain is to reconstruct, from its genome, a genome-scale metabolic network that can be used to perform computer simulations of the metabolic flux distribution that characterizes the strain. In this PhD thesis, an *in silico* strategy was undertaken to increase the knowledge on the metabolism of candidate functional starter cultures for controlled cocoa bean fermentation processes. In Chapter 1, the state-of-the art of genome annotation, genome-scale metabolic reconstruction and modeling, and the cocoa bean fermentation process is presented. Chapter 2 represents the aims and the objectives. In Chapter 3, the genome-scale metabolic reconstruction of *Acetobacter pasteurianus* 386B is exemplified. Chapter 4 builds further by reconstructing genome-scale metabolic models of *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B. The Chapters 5 and 6 are dedicated to (dynamic) genome-scale metabolic modeling of *A. pasteurianus* 386B and *A. ghanensis* LMG 23848^T. In Chapter 7 the whole workflow is applied for *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80.

List of abbreviations

AAB	Acetic acid bacteria
dFBA	Dynamic flux balance analysis
EC	Enzyme Commission number
EMP	Embden-Meyerhof-Parnas pathway
FBA	Flux balance analysis
FVA	Flux variability analysis
GEM	Genome-scale metabolic model
GPR	Gene-protein-reaction association
LAB	Lactic acid bacteria
NDH	NADH dehydrogenase
ODE	Ordinary differential equation
pFBA	Parsimonious flux balance analysis
PGDB	Pathway/Genome Database
PMF	Proton motive force
PPP	Pentose phosphate pathway
PQQ	Pyrroloquinoline quinone
TCA	Tricarboxylic acid cycle
UOX	Ubiquinol oxidase

Chapter 1

Genome-scale metabolic modeling of microorganisms and the cocoa bean fermentation process

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

1 Genome-scale metabolic reconstruction of microorganisms

1.1 Systems biology

From molecules over cells to whole organisms, biological entities interact with each other in various ways. The identification and enumeration of these biological entities is the subject of many biological studies. However, the real challenge in biology lies in the understanding of how living organisms function (Bruggeman and Westerhoff, 2007). This demands to uncover and understand the biological functions that arise through the - possibly time-dependent - interactions between the biological entities. To obtain a better insight into these interactions, a network-based approach has been proposed, whereby the biological entities are collected into a biological network with links between entities corresponding to their interactions (Zhu *et al.*, 2007).

Because of limited knowledge of the network parts and functions that are involved in the interactions, two complementary approaches of systems biology-based studies have been proposed. Top-down approaches aim to infer the functions of the network parts based on the overall behavior of the system, whereas bottom-up systems biology aims to understand the systemic functions from well-characterized network parts (Figure 1) (Bruggeman and Westerhoff, 2007).

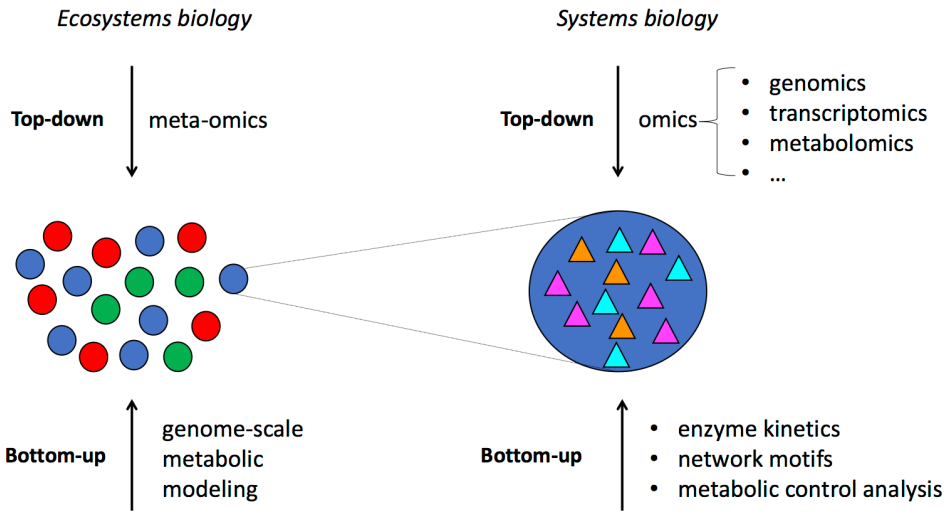


Figure 1. Top-down and bottom-up approaches in (eco)systems biology. In ecosystems biology, the different microorganisms are denoted by a different color. In systems biology, the different biomolecules are denoted by a different color. Based on Bruggeman and Westerhoff, 2007; Röling and van Bodegom, 2014; and Blasche *et al.*, 2017.

Single-celled microorganisms are biological entities ubiquitously present in the natural environment, where they are part of complex microbial ecosystems. A limited number of microorganisms can be cultivated in isolation, providing a way to study their physiology

independently (Blount, 2015; Gill, 2017). Systems biology addresses the gap between the biomolecules that make up the cell and its physiology (Bruggeman and Westerhoff, 2007). Ecosystems biology goes further into analyzing the space- and time-dependent interactions between microorganisms in ecosystems (Klitgord and Segrè, 2011; Blasche *et al.*, 2017). Technological advancements in life sciences, collectively called the omics technologies, have led to the possibility of identifying and enumerating the biomolecules of microorganisms in a comprehensive way, and that in isolation or in mixed ecosystems, as found in natural environments, called meta-omics (Bruggeman and Westerhoff, 2007; Abram, 2015). The analysis of huge multivariate data sets obtained in these top-down studies are challenging to analyze, and methods to obtain a quantitative understanding of the microbial community structure and functioning from these data alone are currently limited. To find a way out, a bottom-up system biological approach is preferred, at least to complement the current data deluge obtained by top-down meta-omics studies (Figure 1) (Röling and van Bodegom, 2014). Indeed, data-rich top-down studies are in urgent need for integration with theory, providing biological models that are amenable to *in silico* simulations and demonstrations (Westerhoff and Palsson, 2004; Danchin, 2009).

Genome-scale metabolic models (GEMs) provide representations of the metabolism of (micro)organisms and are therefore a useful tool to simulate the metabolic potential of particular strains. The reconstruction of such a model requires a top-down systems biology approach, since the model includes biochemical reactions associated to enzyme-encoding genes, and this for as many genes as possible. Also, reconstructed GEMs may be subsequently used in a bottom-up ecosystems biology approach, since they provide a collection of strain-specific metabolic functionalities, and reliable microbial species annotation still represents a bottleneck in top-down studies (Figure 2; Blasche *et al.*, 2017).

1.2 Genome annotation

1.2.1 The gene catalog

Living organisms continuously adapt to changing environments. To keep track of the changes and adaptations required for survival, they rely on a genetic program (Danchin, 2009). The central dogma of molecular biology states that genetic information is encoded in genes that are defined areas on a physical body, DNA, a polymeric macromolecule composed of nucleotides as building blocks. The total content of DNA in a cell is called the genome. The genetic information that this genome contains is transferred *via* transcription, forming the intermediary information transfer molecule messenger RNA (mRNA), and translation, forming the proteins conferring the structural and biochemical functions needed for the cell to survive and reproduce. This translation process makes use of the genetic code, describing the transfer of a triplet encoded in a four nucleotide-containing language (codons) to a 20 proteinogenic amino acid-containing language (proteins). The genetic code, however, is degenerate, meaning that multiple codons can encode the same amino acid.

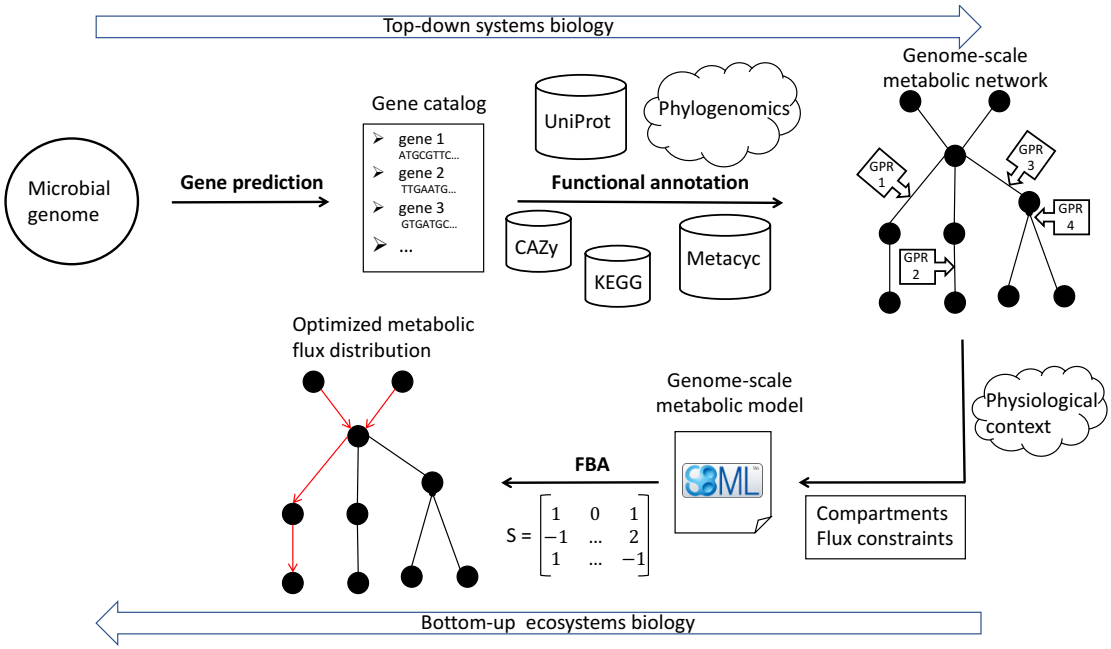


Figure 2. Representation of the workflow of genome-scale metabolic reconstruction and modeling. A top-down systems biology approach is used to reconstruct a genome-scale metabolic network of a particular microbial strain. The genome-scale metabolic model obtained is used to perform simulations of the metabolism, to gain ecological insights, using a bottom-up ecosystems biology approach. FBA, flux balance analysis; GPR, gene-protein-reaction association. Based on Thiele and Palsson, 2010, and Röling and van Bodegom, 2014.

Genome sequencing provides a way to discover nucleotide per nucleotide the encoded genetic information. The genetic code allows to decode the genome, since each amino acid sequence (determining a protein) is associated to a corresponding sequence of DNA, the coding sequence (CDS). Genome annotation deals with the deciphering of the genetic information content, and in the case of protein-encoding genes, to try to infer the probable function of those proteins once they are expressed. In this way, a relation can be made between biological functions and biological sequences of information.

In bacteria, protein-encoding gene prediction amounts to the identification of all coding sequences in the bacterial genome (Richardson and Watson, 2013). *Ab initio* gene prediction tools require the genome sequence solely and are based on statistical learning of nucleotide compositions and succession to predict gene candidates (Delcher *et al.*, 2007; Hyatt *et al.*, 2010; Lomsadze *et al.*, 2018). Nowadays, bacterial genome annotation is mostly performed in an automated way by means of genome annotation pipelines (Meyer, 2003; Van Domselaar *et al.*, 2005; Seemann, 2014; Huntemann *et al.*, 2015; Koskinen *et al.*, 2015; Thakur and Guttman, 2016; Vallenet *et al.*, 2017). These pipelines consist of a number of bioinformatics tools, chained in a way that covers the entire annotation process. Genome annotation pipelines mostly combine *ab initio* gene prediction with sequence similarity searches to identify homologous proteins. The result of gene prediction is a protein-encoding gene catalog, containing the

predicted total protein content (the proteome) encoded in the genome of a particular bacterial strain (Meyer, 2003; Tatusova *et al.*, 2016).

1.2.2 Functional annotation

Once the gene catalog is obtained, the crucial task comes to assign biological functions to the predicted proteins. The functional annotation process assigns functional meta-data to the predefined genes in the catalog. The gene ontology (GO) is a formal ontology organized in a graph to represent biological concepts related to molecular functions, cellular components, and biological processes. It has been proposed as a way to perform consistent functional annotation (Burge *et al.*, 2012). It provides a shared vocabulary toward unifying biological databases and evidence codes, indicating the pieces of evidence (*e.g.*, experimental, genomic, phylogenetic, etc.) that support a certain annotation. The ontology has the advantage of being computer-readable, thus annotation transfer, at different levels of granularity, is independent of exact matching of functional description text (Hastings, 2017). The use of this ontology has also led to the development of functional annotation tools, such as Blast2GO and GO FEAT (Conesa *et al.*, 2005; Araujo *et al.*, 2018). However, its use is not yet as widespread and current functional annotations still very much rely on single-line descriptive texts, in which the biological function of the protein is explained. While being more straightforward, this has the disadvantage of decreased annotation consistency, including problems related to terminology, syntax, spelling, and abbreviations used (Richardson and Watson, 2013).

Biological functions are complex. At this stage of the functional annotation process, the common motto is "from sequence to structure to function" (Lee *et al.*, 2007; Shenoy and Jayaram, 2010). This is actually a misconception, as in its strong version it implicitly leads to intelligent design. It is therefore more appropriate to assign the biological function first and acknowledge that functions can be associated to already existing protein structures (Danchin and Fang, 2016; Danchin *et al.*, 2018).

In practice, however, inferring and, if possible, experimentally demonstrating these biological functions, or their systemic effects, is a slow and labor-intensive process. It must be emphasized that any known biological function today was once unknown, and has been the result of conceptual thinking, modeling, and dedicated experimentation. Most of these biological functions are currently stored in specialized biological databases, originating from the painstaking work of manual curators extracting the information from scientific literature. However, both experimental efforts and manual curation cannot keep up with the enormous amounts of sequencing data generated. Therefore, to functionally annotate a query amino acid sequence (further referred to as protein sequence), one must infer its most probable biological function by comparing it to sequences with a known function. This inductive reasoning is concluded by transferring the biological function to the query protein (Danchin *et al.*, 2018).

There are two main pitfalls regarding functional annotation. One is the relying on biological functions that have not been curated and thus could be wrong; the other is the forced transfer

of a function to the query protein that seems to be the closest based on annotation pipelines, although it is in fact not its actual function (Koonin and Galperin, 2003; Richardson and Watson, 2013). The former relates directly to error propagation in biological databases, which is a well-known problem but difficult to mitigate (Brenner, 1999). The latter is a more subtle disadvantage of the inductive process, as in this case potential discovery of a new protein function is disfavored, preferring to assign a known function to the query protein. The inductive process of functional annotation is currently used, since deductive and abductive reasoning do not have practical applications yet. Thus, the process of gathering as much evidence as possible to support the functional inference must be improved. These pieces of evidence may be found in the neighborhoods of the query protein, given that appropriate measures are taken to assure the quality of the tapped information source (Danchin *et al.*, 2018).

1.2.3 Analysis of neighborhoods

Any relationship that may link genes of the gene catalog to genes in the same or in other species provide the possibility of exploring neighborhoods (Danchin *et al.*, 2018). The types of neighborhoods considered in the functional annotation process performed in this PhD thesis consist of the gene neighborhood, the phylogenetic neighborhood, and the metabolic neighborhood. The information about potential neighbors is provided by acquired knowledge, present in publicly available biological databases. These databases can be general or more specialized regarding the biological information they contain (Table 1).

Table 1. Overview of the discussed biological databases with their attributes. GEM, genome-scale metabolic model. Attribute presence is denoted with a binary code: 0, absence; 1, presence.

Databases	Sequences	Biochemical reactions	Enzyme characteristics	Metabolites	Metabolic pathways	GEMs	Curated information	Ontology
NCBI RefSeq	1	0	0	0	0	0	1	1
the SEED	1	0	0	0	0	0	1	1
UniProt	1	0	0	0	0	0	1	0
eggNOG	1	0	0	0	0	0	1	1
CAZy	1	1	1	0	0	0	1	1
KEGG	1	1	0	1	1	0	1	1
BIGG	0	1	0	1	0	1	1	1
ModelSEED	0	1	0	1	0	1	1	1
MetaCyc	0	1	0	1	1	1	1	1
GO	0	0	0	0	0	0	1	1
IntEnz	0	1	1	0	0	0	1	1
MetaNetX	0	1	0	1	0	1	1	1
Rhea	0	1	0	0	0	0	1	0
ChEBI	0	0	0	1	0	0	1	1

1.2.3.1 Gene neighborhood

The neighborhood of a gene can relate to its neighboring genes in sequence space (Rost, 1999; Danchin *et al.*, 2018). The most used metrics to explore gene neighborhoods are sequence similarity and identity, which are reported by different sequence alignment tools. A query sequence is aligned to a set of target sequences stored in a database. For nucleotide sequence data, the International Nucleotide Sequence Database Collaboration (INSDC, Karsch-Mizrachi *et al.*, 2018) ensures data archiving, accessibility, and organization of daily data exchange between its collaborators, namely the DNA Databank of Japan (DDBJ, Mashima et al., 2017), the European Nucleotide Archive (ENA, Toribio et al., 2017) and the Genbank databank hosted at the National Center for Biological Information (NCBI, Benson et al., 2017). For protein sequence data, the UniProt database stores curated (SwissProt) and uncurated (TrEMBL) protein sequences. The curation is exerted at the functional annotation level, including references to experimental data if available (The UniProt Consortium, 2019). Guidelines are available to select optimal threshold values for annotation transfer based on sequence alignment, such as a cutoff value of 30 % sequence identity, although this fixed value is debatable (Rost, 1999; Addou *et al.*, 2009).

1.2.3.2 Phylogenetic/phylogenomic neighborhood

At the genome level, comparative genomics may be used to assess conservation of the gene order in different genomes, termed gene synteny (Richardson and Watson, 2013). Especially for prokaryotes, for which operons are important transcription units, this is a powerful method to infer gene functions. Evolutionary mechanisms may split and fuse genes, leading to the emergence of novel functions (Henry *et al.*, 2016). Knowledge regarding the function of a protein and recognition that it originated from a fusion process from genes in another genome, which are still independently translated, may lead to the identification of the function of these genes. These proteins have been called 'Rosetta stone proteins' (Marcotte *et al*., 1999). If no independent gene counterparts are known, the protein is called a multidomain protein and may thus be multifunctional (Henry *et al.*, 2016). The lack of recognition of the multidomain architecture means that part of the functionality of the protein is lost.

The phylogenetic and phylogenomic neighborhoods allow to explore the possible evolutionary history of genes and genomes, respectively, thereby providing hints to biological functions. Evolution is characterized by a gradual change in the sequence of genes sharing a common origin. These genes are homologous. For a gene family considered, with an identified common ancestor gene, these changes can be depicted by a phylogenetic tree, called a gene tree, representing the most likely evolutionary path that gave rise to the genes found. If additionally, a species tree is available of the different species from which the genes are originating, the gene and species tree can be reconciled, indicating possible speciation and gene duplication events, from which ortholog and paralog (in- and out-paralogs) relationships between genes can be inferred (Wu *et al.*, 2014a). Two proteins are considered to be in-paralogs when they originated from a gene duplication event within the same species and orthologs when they

originated from a speciation event. Alternatively, if for a pair of genes the gene duplication event has occurred earlier in time than the speciation event, these are called out-paralogs (Figure 3) (Koonin, 2005; Richardson and Watson, 2013). Naturally, the evolutionary path can be complex, combining many speciation and duplication events, gene losses, horizontal gene transfer, and non-orthologous gene displacements (Koonin, 2005). Thus, the identification of orthologs and paralogs in gene families is a difficult task within the domain of evolutionary genomics (Koonin, 2005; Emms and Kelly, 2015). However, they may have important functional implications. According to evolutionary theory, a successful gene function in an organism will be selected for and passed on to its offspring. Thus, orthologous proteins are thought to have the same or at least similar functions in their respective organisms. This notion has been successfully applied to propose new functions for proteins that have the same phylogenetic profile, whereby the profile depicts the presence of the protein in different organisms (Pellegrini *et al*., 1999). Similarly, the concept of orthologous groups (or orthogroups) has been used to cluster orthologous proteins into groups depending on the speciation event considered, and to link those groups to putative biological functions (Tatusov *et al*., 2000, 2003; Emms and Kelly, 2015; Huerta-Cepas *et al*., 2016, 2019). Also, a gene duplication creates new information and therefore possibly new biological functions. Thus, paralogous proteins may have different functions, which may be related to the function of their ancestor protein (Koonin, 2005).

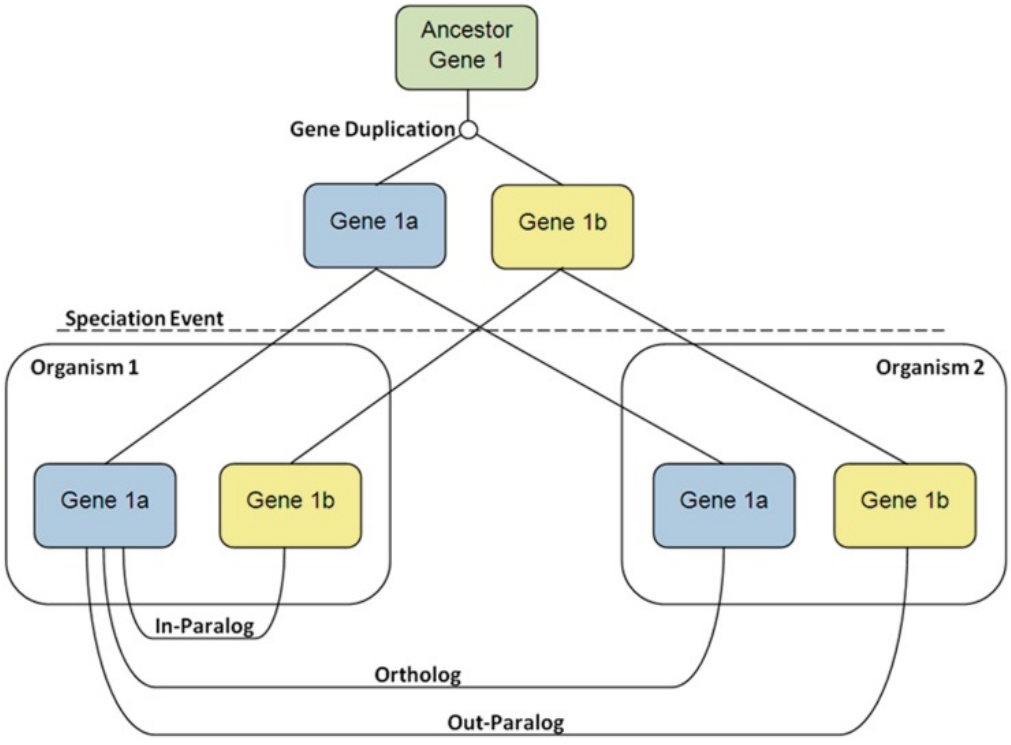


Figure 3. Difference between orthologous and paralogous genes. An additional distinction is made for the paralogous genes, these may be in-paralogs (the duplication event has occurred later in time than the speciation event) and out-paralogs (the duplication event has occurred earlier in time than the speciation event) (Figure from Richardson & Watson, 2013).

Different biological databases group information concerning neighborhoods in distinct biological contexts, and are hence important tools for genome annotation. The KEGG database groups protein sequences, based on their sequence similarity and phylogenetic relatedness, into KEGG orthology (KO) groups (Kanehisa *et al*., 2017). The idea behind is that all proteins in a specific KO group share the same biological function. In addition, the function putatively shared by the KO group is mapped onto metabolic pathways, which are embedded in the database. KEGG includes a predictor tool, BlastKOALA, that performs KO assignment based on sequence alignments to a non-redundant dataset of pangenome sequences, at a specific taxonomic selection level related to the query species. The KOALA algorithm subsequently assigns the query protein to its respective KO group (Kanehisa *et al*., 2016). The SEED database contains a number of grouped proteins, called FIGfams, each group sharing a specific biological function (Aziz *et al*., 2012). It has been used to predict functions of previously unseen proteins via the RAST pipeline, and the resulting annotated sequences can be further linked to biochemical reactions via ModelSEED (Aziz *et al*., 2008, 2012; Henry *et al*., 2010). Databases may also store knowledge regarding specific biological functions. For example, the CAZy database of carbohydrate-active enzymes has developed an ontology to group enzymes into families based on protein sequence similarity and the modular structure of the enzymes. (Lombard *et al*., 2014). For annotation purposes, the information stored in CAZy is leveraged by the dbCAN database, which stores hidden Markov models of signature domains of the different CAZy families (Yin *et al*., 2012).

1.2.3.3 Metabolic neighborhood

Finally, the metabolic neighborhood relates to genes encoding enzymes of the same metabolic pathway (Danchin *et al*., 2018). These genes may be co-regulated, which may be identifiable by genomic rearrangements, leading to gene clusters and possibly operons. Also, metabolic pathways shared among organisms may encode orthologous enzymes with the same phylogenetic profile. The collection of metabolic pathways of an organism defines its metabolic potential. Since the organization and regulation of metabolic pathways in cells is complex, knowledge bases are needed that integrate information concerning the interactions among the different biomolecules. An example is MetaCyc, and the curated BioCyc databases, from which it is derived (Karp *et al*., 2002; Krieger, 2004; Caspi *et al*., 2014, 2016). This database contains curated metabolic pathways of organisms across the tree of life and information about the relationship between the enzymatic reactions and the enzyme-coding genes, and includes a vast number of crosslinks to other databases and experimental data.

High-quality functional annotation requires integration of different neighborhoods of the query proteins explored. As the number of neighborhoods that can be explored by a curator is limited, depending on the number of genes under study and the research question posed, and the resulting annotations can only be as accurate as the reference material used, it is important to define *a priori* the goal of the annotation process (Richardson and Watson, 2013; Danchin *et al*., 2018). In this way, the trade-off between functional annotation accuracy and coverage can be better managed.

1.3 Genome-scale metabolic network reconstruction

To obtain a systemic view of the metabolism of a (micro)organism, accurate functional annotation of metabolic enzymes and transporters is needed. Biomolecules, especially small molecules (metabolites) involved in the core metabolism, can be interconnected through biochemical reactions catalyzed by enzymes encoded by the genome, thus forming a genome-scale metabolic network (Palsson, 2006). This specific type of biological network is amenable to simulate the metabolic potential of microorganisms once the reaction reversibility criteria are applied. Thermodynamics and *in vitro* enzyme characterization should be used to fine-tune reaction directionalities (Feist *et al*., 2009; Thiele and Palsson, 2010). To create a genome-scale metabolic model (GEM), the compartmentalization of the reactions as well as the metabolites transported between compartments need to be defined. It is important to state that many parameters of GEMs are strain-specific and thus the reconstruction process requires strain-specific integration of phenotypical and physiological knowledge (Feist *et al*., 2009; Thiele and Palsson, 2010).

In practice, once the gene catalog has been functionally annotated, the predicted functions can be linked to biochemical reactions in the microbial cell. For enzymes, the link between the functional annotation and a (set of) biochemical reactions is made using the Enzyme Commission (EC) number (Figure 4). The Integrated relational Enzyme database (IntEnz) provides a rich interface between the EC classification and other biological databases (Fleischmann, 2004). The Rhea database of annotated reactions provides the stoichiometry of all enzymatic reactions described in IntEnz using metabolic entities from the Chemical Entities of Biological Interest (ChEBI) database (Hastings *et al*., 2016; Morgat *et al*., 2017).

A

EC 1	EC 2	EC 3	EC 4	EC 5	EC 6	EC 7
Oxidoreductase	Transferases	Hydrolases	Lyases	Isomerases	Ligases	Translocases

B

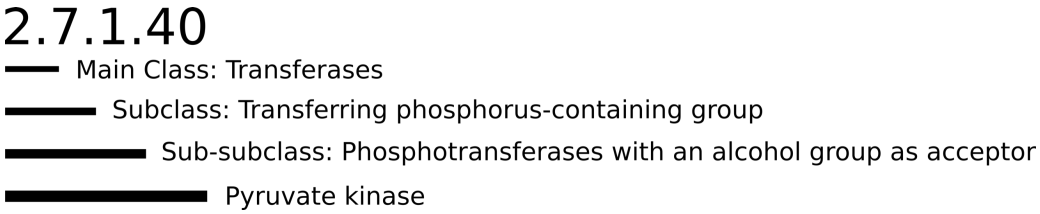


Figure 4. The Enzyme Commission (EC) system of enzymatic reaction classification. The EC system is a hierarchical system of main classes, subclasses, and sub-subclasses. There are seven main classes (A). The classification of the enzyme pyruvate kinase is given as an example (B).

Currently, different GEM reconstruction tools are available, aiming at automating the metabolic reconstruction process. KBase is an online platform for metabolic model

reconstruction, which automates the whole process from raw genomic sequence reads to ready-to-use GEMs (Arkin *et al*., 2018). Similarly, Pathway Tools reconstructs a metabolic network from a functionally annotated genome, taking the MetaCyc database as a reference (Karp *et al*., 2010, 2016). The Pathologic tool embedded in Pathway Tools links the annotated genes to reactions using an algorithm that incorporates curated rules and heuristics (Dale *et al*., 2010; Karp *et al*., 2011). Importantly, these tools represent draft metabolic network reconstructions, which need to be manually curated before biologically meaningful metabolic simulations can be performed. To partly address this problem, the BIGG database combines all biochemical reactions of curated GEMs into one large reaction repository, that can be used for new metabolic reconstructions (King *et al*., 2016).

The existence of many different resources of information related to biochemical reactions has led to a dispersion of the different nomenclatures. MNXref is a biochemical reaction database that proposes a reconciliation of the reaction and metabolite identifiers from other databases, such as BIGG, ModelSEED, MetaCyc and Rhea, into a common namespace (Bernard *et al*., 2014). The MetaNetX database contains genome-scale metabolic network reconstructions that use MNXref as an identifier system. These metabolic network reconstructions have mostly been 'translated' from GEMs occurring in other resources (Moretti *et al*., 2016).

1.4 The Systems Biology Markup Language data format

A reconstructed metabolic network and additional gene-protein-reaction (GPR) associations represent a rich source of information (Thiele and Palsson, 2010). The GPR associations represent the complex relationship between the genome, proteome, and reactome of a microbial strain. To enable to store all the information related to the GEM in a standardized and efficient way, the Systems Biology Markup Language (SBML) data format has been developed (Hucka *et al*., 2003, 2018). SBML is an XML (the eXtensible Markup Language)-based file format that represents all different structural elements of GEMs and the different parameters that relate to their use. Cellular compartments, biochemical reactions, cellular metabolites, and protein-encoding genes are stored as separate entries. The associations between the entries are defined by additional entries. Although the most accurate representation of biochemical activities would be to represent the enzymes as separate entries, in practice current GEMs lump the enzyme-encoding genes into a Boolean GPR rule, of which the logic outcome defines the reaction presence (Figure 5, Jensen *et al*., 2011; Aung *et al*., 2013; Monk *et al*., 2017).

This practice is easier to implement, but may lead to problems in expressing complex GPR associations. For example, multi-enzyme complexes, such as the pyruvate dehydrogenase complex, are composed of different enzymes, each having their own catalytic activity. Lumping the enzyme-encoding genes into a Boolean GPR rule forces to encode the net stoichiometric reaction catalyzed by the pyruvate dehydrogenase complex. The reactions encoded in the model can be specific or generic, depending on the available knowledge.

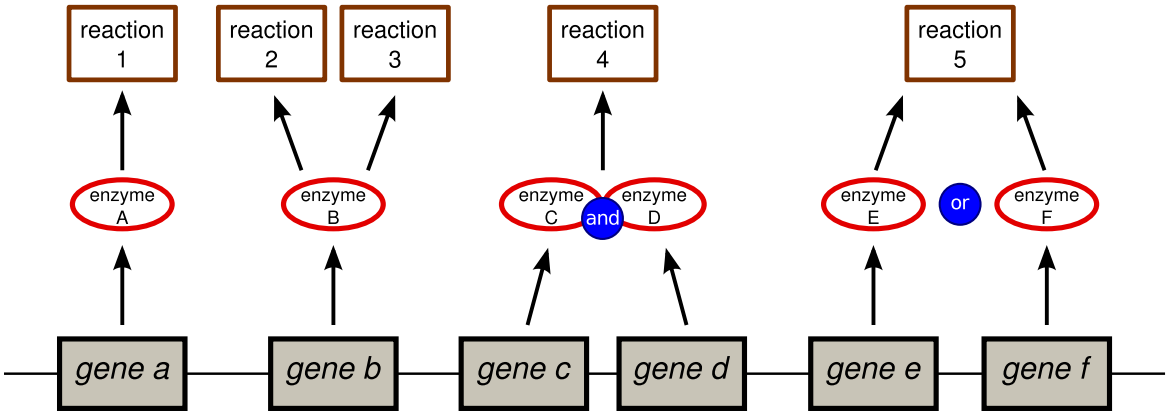


Figure 5. Boolean logic rules are used to describe gene-protein-reaction (GPR) associations. Both genes and reactions are defined in the genome-scale metabolic model (GEM), expressed enzymes form an imaginary translation step that is not incorporated into the GEM. Boolean GPR rules can be used to describe reactions' presence of varying complexity. Reaction 1 is catalyzed by a single gene product, *gene a*. Reactions 2 and 3 are both catalyzed by the gene product of *gene b*. Reaction 4 is catalyzed by an enzyme complex, containing two enzyme subunits being the gene products of *gene c* and *gene d*. Reaction 5 can be catalyzed both by enzyme E and enzyme F, these are isozymes. Figure from Jensen *et al*., 2011.

For specific reactions, specific metabolites are part of the reaction definition, whereas for generic reactions, a group of metabolites (*e.g.*, a fatty acid, an electron acceptor) is defined. The former is preferred, but may require manual creation of stoichiometric reactions, if these reactions are not available in biochemical reaction databases. The latter allows to fill metabolic network gaps to make the model feasible for simulations when insufficient species-specific information is available.

Use of the SBML data format also allows to accurately compare the entries of different models and compare the topological features of the underlying genome-scale metabolic networks. This is a necessity to be able to assess the difference in performance of genome-scale metabolic model reconstruction tools for different microorganisms (Mendoza *et al*., 2019).

2 Microbial population dynamics

2.1 The microbial population growth curve

When an inoculum of microbial cells is added to a suitable growth medium, these cells start to multiply, reflected in a certain specific growth rate μ, by consuming the available substrates in the medium (Peleg and Corradini, 2011; Jun *et al*., 2018). Initially, the microbial cell population grows slowly, since the cells must adapt to the new conditions, which is termed the lag phase. Once adapted, the growth rate increases, ultimately reaching a maximum. During a certain time frame when nutrients are still plentiful and no other factors inhibit the growth of the microbial cell population, the population will continue to grow at a constant maximum specific growth rate (μ_{max}). This is termed the exponential growth phase. When resources

decline and conditions become less favorable, a decrease in the specific growth rate of the population occurs. The population size reaches a plateau, and the cells undergo physiological changes. This growth phase is called the stationary phase. Ultimately, if no refreshment of the growth medium is applied, the number of living cells in the population will decline, leading to the death phase of the microbial cell population (Figure 6A). Importantly, the lag and stationary phases correspond both to a population growth rate that is less than maximal; however, the physiological state of the cells in these growth phases is different (Bertrand, 2019).

2.2 Population growth models

Different phenomenological models have been proposed to describe the microbial population growth curve. These models differ in the number and nature of their parameters. For example, the growth models of Verhulst (1845) and Baranyi and Roberts (1994) include parameters that are intuitive for microbiologists, *e.g.*, the maximum specific growth rate of the population, the lag phase, and the asymptotic carrying capacity. Other models, such as the Gompertz model (1825), include less intuitive parameters, but fit equally well the experimental data obtained from a growing microbial population. Phenomenological models can be used to describe and predict microbial population growth curves, but none of them offer any mechanistic insight into the microbial growth process (Peleg and Corradini, 2011).

The specific growth rate is an important microbial growth parameter, and one that has fascinated generations of microbiologists (Jun *et al*., 2018). It is dependent on different abiotic factors, such as the substrates and their concentrations (S), temperature (T), pH, and water activity (a_w), of which the influence can be described by so-called secondary growth models (Peleg and Corradini, 2011). To describe the microbial population (X) dynamics, the following general Equation (1) applies:

$$\frac{dX}{dt} = \mu(S, X, T, pH, a_w, \dots) * X \qquad (1)$$

One of the most used growth models is the Monod growth model, describing the non-linear relationship between the specific growth rate and a single growth-limiting substrate (Figure 6B) (Monod, 1949). A maximal, constant specific growth rate parameter is characteristic of an exponentially growing culture. Its corresponding physiological state is named the steady-state growth condition, or, the less restrictive, balanced growth condition (Neidhardt *et al*., 1990; Fishov *et al*., 1995; Jun *et al*., 2018). This physiological state implies a steady-state of the intracellular metabolite concentrations in the microbial cell population, leading to a proportionally equal production of all biomass components (de Jong *et al*., 2017). Thermodynamic principles have been proposed to explain the (pseudo-) steady-state of the intracellular metabolite concentrations, stating that the osmotic limit cells face lead to metabolite dynamics that can be safely neglected at timescales larger than minutes (Heijnen, 2010). Thus, balanced growth and the maximum specific growth rate are two faces of the same coin (Schaechter, 2015). As a consequence, balanced growth may provide mechanistic insight into the maximal specific growth rate of a microbial population. So, if a model could be made

that captures balanced growth, predictions of the maximum specific growth rate of a microbial cell population would offer biological insight into how this maximum specific growth rate is attained. This is exactly what GEMs were invented for. Importantly, GEMs only capture the role of the metabolic network leading to a specific growth rate, and do not include the influence of abiotic factors on the specific growth rate (Nilsson *et al*., 2017). Additionally, it follows from the theory that the GEMs can only predict the specific growth rate when the growing microbial cell culture is in balanced growth (Gottstein *et al*., 2016).

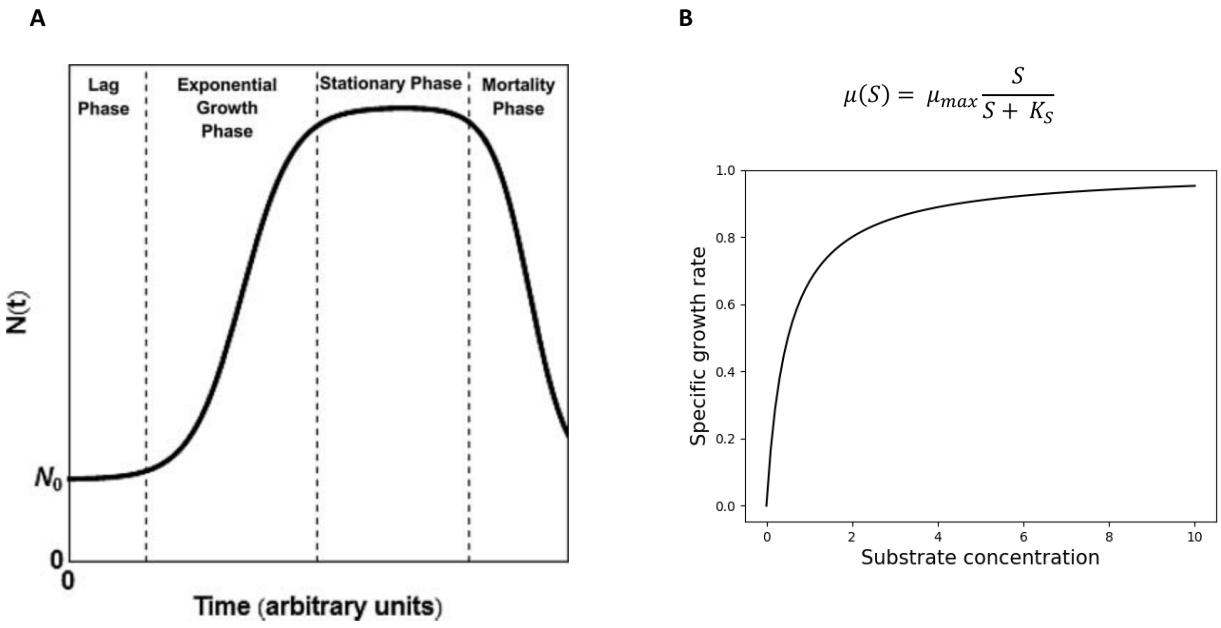
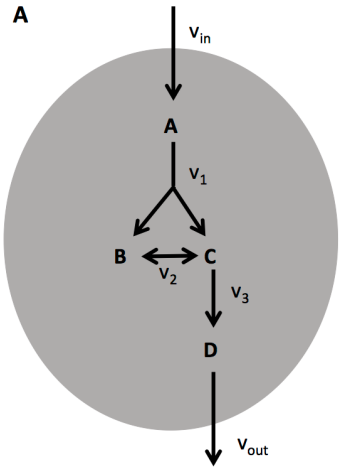


Figure 6. A microbial population growth curve (A; figure from Peleg and Corradini, 2011). Illustration of the Monod growth model that defines the specific growth rate (μ) as a function of the substrate concentration (S) (B; own work, $\mu_{max} = 1$ h^{-1}, $K_S = 0.5$ mM).

3 Constraint-based modeling

A genome-scale metabolic network reconstruction is a representation of the overall metabolism of a microbial cell population (Figure 7A). The nodes of the network correspond to the metabolites and the links correspond to the stoichiometric conversions of the metabolites, performed by the respective enzymes. The graphical representation of the network is amenable to different types of analyses, such as topological analysis or pathway identification, and can be used to obtain ecological insights (Mahadevan and Palsson, 2005; Faust *et al*., 2011; Carr and Borenstein, 2012). However, to perform metabolic simulations, it is more practical to convert the metabolic network into a mathematical format, namely a stoichiometric matrix (Figure 7B), that can be better handled (Orth *et al*., 2010; Thiele and Palsson, 2010).



B

Stoichiometric matrix (**S**)

	in	1	2F	2R	3	out
A	1	-1	0	0	0	0
B	0	1	-1	1	0	0
C	0	1	1	-1	-1	0
D	0	0	0	0	1	-1

C

Reaction flux vector (**v**)

v_{in}
v_1
v_{2F}
v_{2R}
v_3
v_{out}

Figure 7. Representation of a compartmentalized (grey area) metabolic network, with v_i the reaction fluxes (v_{in} and v_{out} represent the exchange fluxes with the environment); A, B, C, and D represent intracellular metabolites (A), the corresponding stoichiometric matrix (**S**) (B), and the reaction flux vector (**v**) (C).

The balanced growth condition that leads to the exponential growth phase, is associated with a steady-state (dynamic equilibrium) of the intracellular metabolite concentrations, given by the following mass balance:

$$d\boldsymbol{m}/dt = \boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{0} \qquad (2)$$

Equation (2) indicates that the mass balance of all intracellular metabolites (**m**) is equal to the matrix product of the stoichiometric matrix (**S**) and the flux vector (**v**). In this equation, the dilution of metabolites due to growth was neglected (De Jong *et al*., 2017). The flux vector (Figure 7C) contains all reaction flux values v_i associated to the links in the metabolic network and has the unit mmol/g_{CDW}/h. The flux thus expresses the amount of substance that flows through one gram of dry weight of the growing culture. Setting the expression equal to the null vector expresses the steady-state condition. Graphically, it corresponds to an equality of all production fluxes and all consumption fluxes with respect to one metabolite node in the metabolic network.

Applying the steady-state principle to a metabolic network leads to a set of possible metabolic flux distributions, which represent the solution space of the network. Exploratory modeling samples this solution space to obtain potential metabolic flux distributions. Examples of such methods are elementary flux mode analysis and flux sampling (Trinh *et al*., 2009; Machado *et al*., 2012; Megchelenbrink *et al*., 2014). Further constraints can be applied to reduce this solution space, essentially expressing the feasible solution space in a particular environment. For example, by measuring a set of reaction fluxes (v_i) that range from a minimum to a maximum value, these values can be added as a constraint to the model:

$$v_{i,min} \leq v_i \leq v_{i,max} \qquad (3)$$

This strategy is used in metabolic flux analysis (MFA), whereby reaction fluxes can be inferred from consumption of isotopically-labeled substrates (Antoniewicz, 2013). For example, ^{13}C-MFA results in specific labelling patterns of product metabolites, which depend on the particular flux distribution of the microorganism under study (Quek *et al*., 2009; Kohlstedt *et al*., 2010).

In contrast to MFA, flux balance analysis (FBA) does not directly rely on the measurement of reaction fluxes, but takes an additional assumption. This assumption finds its origin in the principle of optimality by natural selection, and assumes that the genetic program of cells is optimized to reach a certain cellular objective. Many different cellular objectives have been examined and there is a lot of debate about which one, if there is one, is applicable to which organism in which condition (Schuetz *et al*., 2007; Feist and Palsson, 2016; Zhao *et al*., 2016). The most widely used objective is the biomass objective function, which assumes that microorganisms optimize their metabolic flux distribution toward the optimization of the production of biomass compounds, which is taken as an approximation for the specific growth rate of the microbial cell population (Feist and Palsson, 2010). The optimization problem that FBA solves takes the form of a linear program:

$$\begin{gathered} Maximize\ Z\ (e.g., with\ Z = v_{biomass}) \\ \text{subject to} \\ \boldsymbol{S} \times \boldsymbol{v} = 0 \\ v_{i,min} \leq v_i \leq v_{i,max} \end{gathered} \tag{4}$$

The operations performed by the linear program in the flux solution space are shown in Figure 8. The flux, $v_{biomass}$, represents the flux of the biomass reaction. This reaction simulates biomass formation by including all biomass precursor metabolites and assigning stoichiometric weights to each of them, mostly based on experimental data concerning the cellular composition (Feist and Palsson, 2010; Thiele and Palsson, 2010). Thus, biomass production acts as a sink of matter to simulate the growing culture.

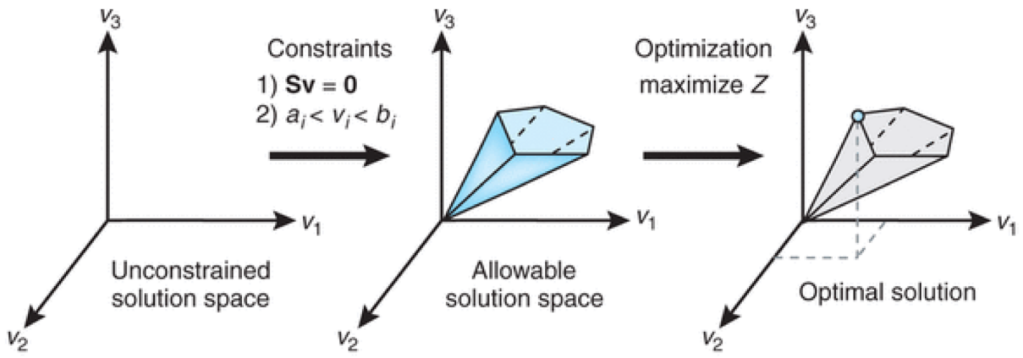


Figure 8. Constraining the flux solution space with flux balance analysis, with **S** the stoichiometrix matrix, **v** the flux vector, and v_i the reaction fluxes (Figure from Orth *et al*., 2010).

Even though solving the linear program will lead to a unique maximal biomass production flux value, the corresponding metabolic flux distribution may not be unique. This reflects the existence of the optimal solution space, and it can be explored using flux variability analysis (FVA; Gudmundsson and Thiele, 2010). This FBA-based method reports the minimum and maximum flux values of each reaction in the model, using the predicted maximal specific growth rate by FBA as an additional constraint (Maarleveld *et al*., 2013). In parsimonious FBA (pFBA), the same set of constraints as in FVA are applied but now the objective becomes to minimize the total amount of flux in the model, thus returning the most parsimonious metabolic flux distribution corresponding with a maximal biomass production (Lewis *et al.*, 2010). Finally, another method called CoPE-FBA has been developed to better understand in terms of metabolic network topology how optimal metabolic flux distributions predicted by FBA arise out of the combination of elementary flux modes, applied constraints, and the optimization objective (Maarleveld *et al*., 2015).

The parameters related to constraint-based modeling, *i.e.*, flux constraints and the objective function, are part of the GEM and thus should be defined in its corresponding SBML file. In this way, the GEM can be contextualized, referring to which kind of biological system it is supposed to represent (Gu *et al*., 2019). Sensitivity analysis of mathematical models can be used to determine the impact of changes in independent variables onto the dependent variables. For GEMs, the influence of the variables (reaction fluxes) on the objective function value can be investigated by computing reduced costs. Alternatively, the influence of varying constraints on the objective function value can be investigated with shadow prices (Palsson, 2006). The use of shadow prices has been applied to define a phenotype phase plane, demarcating regions of distinct metabolic pathway utilization on two growth-limiting substrates (Edwards *et al*., 2002). Specifically related to the steady-state constraint, sensitivity analysis allows to explore how deviations from this constraint (flux imbalance) can lead to the identification of growth-limiting metabolites (Reznik *et al*., 2003).

3.1 Overflow metabolism

The biomass yield, growth yield, or cell yield coefficient $Y_{X/S}$, is defined as the amount of biomass produced by the microbial cell population (ΔX) with respect to the amount of substrate consumed (ΔS):

$$Y_{X/S} = \frac{\Delta X}{|\Delta S|} \tag{5}$$

Taking the derivative of both biomass production and substrate consumption in Equation 5 allows to express the biomass yield in terms of the specific growth rate (μ) and the specific substrate consumption rate (substrate consumption flux, v_S) of the microbial cell population:

$$Y_{X/S} = \frac{dX/dt}{|dS/dt|} = \frac{\mu * X(t)}{|v_S| * X(t)} = \frac{\mu}{|v_S|} = \frac{v_{biomass}}{|v_S|} \tag{6}$$

Flux balance analysis, using the biomass reaction as the objective function, maximizes the specific growth rate with respect to a given constraint on the substrate consumption flux. From Equation 6 follows that the optimization of the specific growth rate, given a limited substrate consumption flux, inherently optimizes the biomass yield (Schuster *et al*., 2008; Santos *et al*., 2011). Thus, the amount of substrate consumed by the model is used in the most efficient way to maximize the production flux of the biomass precursor metabolites leading to a maximal specific growth rate. However, there is experimental evidence that microorganisms increase their growth rate by lowering their biomass yield. A prime example is *Saccharomyces cerevisiae* grown in a chemostat with glucose as the growth-limiting substrate under aerobic conditions (Fraenkel, 2011). At low dilution rates, the cell population will respire the available glucose and the biomass yield will stay constant. However, as the dilution rate is increased, at some point the cells will start to ferment the available glucose and form ethanol, with a concomitant decrease in the biomass yield. This is an example of overflow metabolism and for *S. cerevisiae* it has been named the Crabtree effect (Fraenkel, 2011; Pfeiffer and Morley, 2014).

Multiple arguments have been proposed to explain the occurrence of overflow metabolism and the use of lower energy-yielding pathways in general. Thermodynamic bottlenecks in high-yield pathways may lead to increased protein usage to obtain a higher pathway flux, thus forcing the cell to reallocate much of its protein content toward these pathways, or choose a low-yield strategy instead (Pfeiffer *et al*., 2001; Molenaar *et al*., 2009; Flamholz *et al*., 2013; Stettner and Segrè, 2013; Pfeiffer and Morley, 2014). Thus, microorganisms seem to face a trade-off regarding optimization of substrate usage (Zhao *et al*., 2016). An increase of the substrate uptake rate may lead to a decrease in the biomass yield due to the secretion of the metabolites formed. However, an increase of the biomass yield is typically seen when microorganisms face growth conditions under which it is difficult to catabolize substrates (Ibarra *et al*., 2002; Teusink *et al*., 2009).

3.2 Dynamic flux balance analysis

Flux balance analysis allows to predict the intracellular metabolic flux distribution of a microbial cell population based on a set of metabolic flux constraints, most importantly the substrate consumption constraints (Orth *et al*., 2010). However, it cannot account for the changes in the extracellular metabolite concentrations, which may influence the metabolic activities. For example, microbial growth in a batch reactor may lead to different outcomes, including the preferential consumption of substrates (catabolic repression), or the accumulation of the metabolites produced, which may inhibit growth (product inhibition), or may be consumed in a later stage of the population growth (make-accumulate-consume strategy) (Monod, 1949; Van Impe *et al*., 2005; Hagman *et al*., 2013). Dynamic flux balance analysis (dFBA) is an FBA variant that can be used to take the microbial population and extracellular metabolite dynamics into account. A pseudo-steady state assumption is made by considering that medium dynamics occur much more slowly, at time scales larger than minutes, than the intracellular metabolite dynamics, which are rapid, typically with turnover times in the order of seconds (Heijnen, 2010; Gottstein *et al*., 2016; Reimers and Reimers, 2016). In the static

optimization approach of dFBA, a specific time interval is defined, whereby a linear program is solved at the beginning of each interval and the fluxes predicted by the GEM are integrated over the time interval (Mahadevan *et al*., 2002). The time interval defined depicts the range in which metabolic transitions due to metabolic or genetic control are negligible. Thus, dynamic changes in the microbial population and extracellular metabolites are simulated as occurring by a series of snapshots of the predicted metabolic flux distribution, as a function of environmental growth constraints. These constraints need to be added by the modeler and typically concern the substrate consumption fluxes (v_i), for which the Michaelis-Menten equation of enzyme kinetics expresses the dependency of the flux on the substrate concentration (S_i) (Hanly and Henson, 2011; Gottstein *et al*., 2016):

$$v_{i,min} \leq v_i \leq v_{i,max} \frac{S_i}{S_i + K_{S_i}} \tag{7}$$

The Michaelis-Menten term in the above equation defines the upper bound of the substrate consumption flux (determined by $v_{i,max}$; K_{Si} is the substrate concentration at which the consumption flux value equals $v_{i,max}/2$). The specific growth rate (μ_{FBA}) and exchange fluxes (v_M) are predicted by the GEM by solving the linear program of Equation 4. The dynamics of the environmental state variables (X, microbial cell population; M_j, extracellular metabolites) are then updated according to a dynamic model (for a batch fermentation process):

$$\frac{dX}{dt} = \mu_{FBA} * X \tag{8}$$

$$\frac{dM_j}{dt} = v_{M_j} * X \tag{9}$$

Dynamic flux balance analysis has been applied to explain the diauxic shift of *Escherichia coli* when growing on glucose, whereby the accumulated acetic acid is consumed in a second growth phase (Mahadevan *et al*., 2002; Succurro *et al*., 2019). DFBAlab is a specific tool for dynamic flux balance analysis, addressing the challenges of obtaining unique exchange fluxes by performing lexicographic optimization and infeasibility of the linear program by detecting a basis change in the optimal solution space (Höffner *et al*., 2013; Gomez *et al*., 2014).

4 Applications of genome-scale metabolic models

In what follows an overview of applications of GEMs is presented. Crucially, the success of the application depends on how well these models approach the biological reality, entailing (multiple rounds of) manual curation. For species of lactic acid bacteria (LAB), acetic acid bacteria (AAB) and yeasts, all present in the cocoa bean fermentation process, an overview of currently available GEMs is shown in Table 2.

4.1 Optimization of biomass and product yield

GEMs are useful tools to establish the exact nutritional requirements of a particular microbial strain. By combining single-omission growth experiments regarding substrates and FBA simulations, one can predict the nutritional requirements of a strain and have a lead toward discovery if the experimental and *in silico* results do not correspond. Once a minimal medium is obtained, further addition of compounds that are predicted to increase the biomass yield can be added to the medium to obtain a chemically defined medium with improved biomass yield (Teusink *et al*., 2005; Wegkamp *et al*., 2010). The comparison of different *Yarrowia lypolytica* GEMs related to specific growth rate predictions in various culture conditions allowed to identify and curate discrepancies in the models related to the biomass equation and reaction directionality (Xu *et al*., 2020). For metabolic engineering purposes, using the production flux of a certain metabolic compound as the objective of the simulation, gene modification targets may be predicted to improve the productivity of the strain. This is performed by exploiting the GPR associations of the model, thus predicting which reaction fluxes do influence, positively or negatively, the objective (Yang *et al*., 2018). Other *in silico* strain design methods include co-factor modification analysis, whereby the influence of the co-factor specificity of the reactions in the model is investigated toward improved product and biomass yield (Lakshmanan *et al*., 2013).

Table 2. Overview of microorganisms of lactic acid bacteria (LAB), acetic acid bacteria (AAB), and yeast species for which a genome-scale metabolic model is available.

Species group	Strain	Model name	Reference
LAB			
Enterococcus faecalis	V583	None	Veith *et al*., 2015
Lactobacillus paracasei	LC2W	iJL846	Xu *et al*., 2015
Lactobacillus plantarum	WCFS1	None	Teusink *et al*., 2006
Lactobacillus reuteri	JCM1112	None	Saulnier *et al*., 2011
Lactococcus lactis subsp. *cremoris*	MG1363	None	Flahaut *et al*., 2013
Leuconostoc mesenteroides subsp. *mesenteroides*	ATCC 8293	iLME620	Koduru *et al*., 2017
Oenococcus oeni	PSU-1	iSM454	Mendoza *et al*., 2017
Streptococcus thermophilus	LMG 18311	None	Pastink *et al*., 2009
AAB			
Gluconobacter oxydans	621H	iXW433	Wu *et al*., 2014b
Komagataeibacter nataicola	RZS01	iHZ771	Zhang *et al*., 2017
Yeasts			
Saccharomyces cerevisiae	S288c	Yeast 7	Aung *et al*., 2013
Pichia pastoris	X-33	iMT1026	Tomàs-Gamisans *et al*., 2018
Scheffersomyces stipitis	CBS 6054	iSS884	Caspeta *et al*., 2012
Candida glabrata		iNX804	Xu *et al*., 2013
Candida tropicalis		iCT646	Mishra *et al*., 2016
Schizosaccharomyces pombe	DSM-70576	SpoMBEL1693	Sohn *et al*., 2012
Kluyveromyces lactis		iOD907	Dias *et al*., 2014
Yarrowia lipolytica	Po1g	iYali4	Kerkhoven *et al*., 2016

4.2 Context-specific genome-scale metabolic models

Context-specific GEMs are derived from their generic counterparts by adapting the topology or constraints of a metabolic network in different biological contexts (Gu *et al*., 2019). Different types of omics data, determined in a specific biological context, may be integrated into a GEM to infer the possible influence on the metabolic fluxes. Constraint-based modeling has been used to predict essential genes in microbial pathogens, which make them interesting drug targets. For example, essential genes of *Acinetobacter baumannii*, one of the notorious ESKAPE pathogens, were predicted with a GEM by integrating gene expression data obtained after colistin exposure (Presta *et al*., 2017). Comparing gene essentiality predictions of *Mycobacterium tuberculosis* H37Rv GEMs allowed to validate an updated version of the model (Kavvas *et al*., 2018). GEMs of the different lifecycle stages of the malaria parasite *Plasmodium falciparum* have been developed to identify stage-specific drug targets (Abdel-Haleem *et al*., 2018). In another study with the same organism, metabolomics data were integrated to parametrize the reaction flux constraints using a method called thermodynamics-based flux analysis. By doing so, new essential genes could be predicted (Chiappino-Pepe *et al*., 2017).

4.3 Metabolic function discovery

Genome-scale metabolic models can be used to discover new biological functions (Gu *et al*., 2019). By comparing gene essentiality simulations with experimental data of knockout strains, genes may be identified that are predicted to be essential but in fact are not. These correspond to enzyme functions for which another, possibly promiscuous enzyme exists that can compensate for the absence of the first. By using sequence homology searches in the query genome, candidate isozymes may be identified that can be further validated *in vitro* (Guzmán *et al*., 2015; Oberhardt *et al*., 2016). To compare the metabolic potential between related microbial species, a new concept of the pan-reactome was developed, complementing the notion of a pan-genome. In this way, the diversity of secondary metabolites of *Penicillium* species was investigated and compared to their phylogenetic relatedness (Prigent *et al*., 2018).

4.4 Metabolic interactions between organisms

In nature, microorganisms occur in communities. GEMs have been proven successful as a new tool to study the metabolic interactions between the microbial community members (Kim *et al*., 2017; Gu *et al*., 2019). Leveraging the available knowledge in GEMs may therefore prove useful to study microbial community properties *in silico*, test hypotheses, and design new experiments. Different approaches have been applied to build and simulate metabolic activities of microbial communities. The first microbial community model was established for an *in vitro* ecosystem consisting of two mutualistic species, namely the sulfate-reducing *Desulfovibrio vulgaris* and the methanogenic archaea *Methanococcus maripaludis* (Stolyar *et al*., 2007). Using FBA, the authors could investigate biomass and metabolic product yields in the single GEMs *versus* the community model, taking into account the influence of the biomass fractions,

as well as the influence of the transfer of different electron donors between the community members (Stolyar *et al*., 2007). A subsequent study took a different approach and enumerated all elementary flux modes in a community. This exploratory modeling approach allowed to investigate *in silico* the complex ecosystem of phototrophic mats in Yellowstone National Park (Taffs *et al*., 2009). The compartmentalization of the three microbial guilds (oxygenic phototrophs, filamentous anoxygenic phototrophs, and sulfate-reducing bacteria) and their metabolic interactions were defined based on previous knowledge. A day and night simulation was performed to investigate the biomass yield of the elementary flux modes and the influence of the flux ratio at an important branch point in the metabolic network, the RuBisCo enzyme of the photoautotrophic microorganisms (Taffs *et al*., 2009). The theoretical foundations of FBA for a microbial community were established by extending the balanced growth conditions to an obligatory mutualistic community (Khandelwal *et al*., 2013). In essence, for a microbial community to engage in balanced growth, the specific growth rate of the entire community must be constant, leading to a steady-state of all intra- and extracellular metabolites. However, at this point, the steady-state expression becomes dependent on the product of the community biomass fractions and the community growth rate, leading to a non-linear optimization problem. Linearization, by defining the biomass fraction as a parameter, makes the simulations computationally less extensive (Khandelwal *et al*., 2013).

5 Systems biology and genome-scale metabolic modeling in food fermentation processes

Microorganisms form the basis of food fermentation processes (Marco *et al*., 2017). Different traditional fermented foods are characterized by microbial communities consisting of diverse bacterial and fungal species. The interactions between the community members contribute to the robustness of the ecosystem to varying environmental conditions. Various culture-dependent and –independent methods are used to characterize the members of food fermentation ecosystems (Tamang *et al*., 2016). Culture-independent methods have greatly evolved with the advent of high-throughput sequencing. Amplicon sequencing of (partial) phylogenetic marker genes or metagenomic shotgun sequencing of the entire environmental DNA pool has been used to make an inventory of all microbial species present in a food fermentation ecosystem (Illeghems *et al*., 2012; Ercolini, 2013; De Filippis *et al*., 2017; Verce *et al*., 2019; Pothakos *et al*., 2020). Furthermore, top-down studies allow to estimate the correlations between metabolite profiles of fermented foods produced and the presence of specific microorganisms in the microbial ecosystem (Bertuzzi *et al*., 2018). Through careful monitoring of the environmental conditions during fermentation, associations can be made between these factors and the prevailing microbial species. The identification of ecological interactions between the community members provides further insight into how these communities are formed, and lead research toward characterizing these interactions in more mechanistic terms (Mounier *et al*., 2008). The metabolic potential of the microbial communities, for example in terms of aroma formation through metabolic interactions between community members, has been investigated by mining for metabolic functions in metagenomics data sets (Wolfe *et al*., 2014; Verce *et al*., 2019; Pothakos *et al*., 2020).

Many community members of food fermentation ecosystems have been cultured and identified (Wolfe and Dutton, 2015; Rezac *et al*., 2018). However, their physiological characterization has not always followed. This is important because some of these microorganisms are used as starter cultures in industrial food fermentation processes, thus requiring a higher demand of physiological control (Hansen, 2002). Therefore, the use of GEMs as knowledge bases for starter cultures has recently drawn the attention of food microbiologists (Rau and Zeidan, 2018). The combination of constraint-based modeling and omics data integration may improve the knowledge regarding industrially relevant traits, such as carbohydrate utilization, amino acid auxotrophies, (other) nutrient requirements, and the production of metabolites. Different microbial communities have been subjected to microbial community modeling to obtain more insight into how these communities are formed and propagate. In dairy products, new insights into the protocooperative relationship between *Streptococcus thermophilus* and *Lactobacillus delbrueckii* in yoghurt production seem promising toward the development of a microbial community model (Herve-Jimenez *et al*., 2009; Rau and Zeidan, 2018). A novel stable consortium of *S. thermophilus* and *Lactobacillus rhamnosus* has been developed using insights obtained with community FBA to enable propagation of probiotic bacterial strains in fermented foods (Kort *et al*., 2015). In general, food fermentation ecosystems seem to be perfectly suitable to combine modeling and experimental procedures. This is due to their relative simplicity and reproducibility, making them tractable ecosystems that could be used as representative models of microbial ecosystems (Wolfe and Dutton, 2015; Blasche *et al*., 2017).

6 The cocoa bean fermentation process

6.1 The rise of cocoa

Cocoa is the product of the cocoa tree, *Theobroma cacao* L. (Motamayor *et al*., 2008; De Vuyst and Weckx, 2016b). This evergreen tree grows in a shaded, tropical environment. Therefore, it can only be cultivated within a narrow belt of 20° latitude (the cocoa belt) around the equator (Colombo *et al*., 2012; Ozturk and Young, 2017). The origins of the cocoa tree are probably to be found in the Amazon basin (Zarrillo *et al*., 2018). The oldest currently reported usage of cocoa could be traced back to Ecuador and dates from approximately 5300 years ago (Zarrillo *et al*., 2018). The cocoa pods, the fruits of the cocoa tree, contain the cocoa beans. These beans are the seeds of the cocoa tree and are embedded in a carbohydrate-rich, mucilaginous pulp. Each cocoa bean contains an embryo and two cotyledons, which range from whitish to deep purple in color, depending on the anthocyanin content of the cocoa variety (Lima *et al*., 2011; Colombo *et al*., 2012). In Pre-Colombian times, cocoa trees were heavily cultivated in Mesoamerica by the Aztec civilization (Coe and Coe, 2013), and cocoa was consumed as a frothed drink made up of ground cocoa beans, water, and chili (Ozturk and Young, 2017). The spiciness would counter the bitterness and astringency present in the cocoa beans, which is due to their relatively high content of polyphenols and alkaloids (Colombo *et al*., 2012). With the advent of the Spanish conquistadores, Hernán Cortés may have introduced cocoa in Europe in the beginning of the 16th century, after observing its consumption and ritual usage during his expeditions (Schwan and Wheals, 2004; Ozturk and Young, 2017). In the 17th century, in

Europe, the cocoa drink was softened by adding sugar and vanilla (Coe and Coe, 2013). The invention of the cocoa press, enabling the separation of cocoa liquor or mass and cocoa butter by Coenraad van Houten in 1828 depicts the initial industrialization of the chocolate bar-making process (Ozturk and Young, 2017).

Nowadays, the cocoa market is a billion-dollar industry, being the most profitable food source (ICCO, 2017). There are three main cocoa crop varieties, namely Forastero, Criollo, and Trinitario. However, these do not accurately reflect the genetic diversity of *T. cacao* (Motamayor *et al.*, 2008). The main cocoa-producing countries are located in West Africa (mainly Ivory Coast and Ghana), Asia (mainly Indonesia), and Latin America (mainly Brazil and Ecuador) (Ozturk and Young, 2017). The cocoa value chain has long been plagued by fraudulent practices, leading to a decrease in revenues for cocoa farmers and ultimately a decrease in cocoa quality. The current restructuring of this global chain involves more attention to transparent trade and the origin of the cocoa consumed. This has led to the emerging bean-to-bar concept, applying the concept of terroir in the chocolate industry (Engeseth and Ac Pangan, 2018). Ultimately, programs can be set up to improve tracing of cocoa and contribute to good farming practices with respect to the cocoa plantations and the livelihood of cocoa farmers, for example Cacao-Trace of the Puratos group (https://www.cacaotrace.com).

6.2 From bean to bar

The current process of chocolate production and derived cocoa products involves multiple space- and time-dependent steps (Afoakwa, 2016; De Vuyst and Weckx, 2016b; Gutiérrez, 2017). At the cocoa farm, the harvest of the ripe cocoa pods, the fermentation of the cocoa pulp-bean mass, and the drying of the fermented cocoa beans is performed. Cured cocoa beans are then transported to a cocoa processing site (mostly to a country outside of the cocoa belt). Here, the cocoa beans are further processed to produce cocoa-derived products. The artisanal way of producing chocolate is to perform whole-bean roasting, followed by winnowing to remove the husks (bean shells). The cocoa nibs (bean cotyledons) are then ground, producing cocoa liquor or mass (Afoakwa, 2016; Gutiérrez, 2017). The cocoa liquor obtained is refined to reduce the particle size below human tongue detection, and subsequently conched. This last flavor-determining step agitates and aerates the liquor to further develop its flavor and to eliminate the remaining volatile organic acids. Conching is also necessary to stabilize the chocolate suspension, coating all solid cocoa particles with cocoa butter (Gutiérrez, 2017; Engeseth and Ac Pangan, 2018).

Cocoa bean fermentation is necessary to degrade the cocoa pulp, aiding the drying process and preventing bean spoilage, and to prevent the cocoa embryo from germination, thereby inducing the formation of various cocoa aroma precursor and colour compounds (Lima *et al.*, 2011; De Vuyst and Weckx, 2016a,b; de Melo Pereira *et al.*, 2016; Rawel *et al.*, 2019). The fermentation process depicts the totality of microbial activities in the cocoa pulp and induced biochemical changes in the cotyledons of the cocoa beans (Lima *et al.*, 2011; De Vuyst and Weckx, 2016a,b). Although the global demand for high-quality cocoa beans keeps increasing, the

necessary fermentation step still is a spontaneous one (De Vuyst and Weckx, 2016a,b; Kongor *et al*., 2016; Ozturk and Young, 2017). The cocoa bean fermentation process (Figure 9) is initiated by scooping the cocoa pulp-bean mass out of the cocoa pods. The most common fermentation methods include heaps and wooden boxes, but also baskets, trays, and platforms are in use, depending on the cocoa-producing region and local farm practices (De Vuyst and Weckx, 2016a,b; Gutiérrez, 2017; Figueroa-Hernández *et al*., 2019). The cocoa pulp-bean mass is left to ferment for 2 to 10 days, depending on the cocoa variety and local farm practices (De Vuyst and Weckx, 2016a,b).

The cocoa pulp-bean mass fermentation is performed by microorganisms that are present in the environment (Schwan and Wheals, 2004; De Vuyst and Weckx, 2016a,b). A multitude of bacterial and fungal species have been identified culture-dependently, and many more have been detected culture-independently. Notable groups of microorganisms include yeasts, LAB, AAB, enterobacteria, *Bacillus* species, and filamentous fungi (Camu et al., 2007; Papalexandratou *et al*., 2011b, 2013; Illeghems *et al*., 2012; Ozturk and Young, 2017). They start to consume the carbohydrates in the cocoa pulp-bean mass, degrading its content and decreasing its viscosity by degrading pectin and hemicellulose in the cocoa pulp, which allows air ingress (De Vuyst and Weckx, 2016a,b). The fermentation is characterized by an alcoholic production phase under quasi anaerobic conditions, followed by an acetic production phase under increasing aerobic conditions. Initially, high concentrations of mainly sucrose, glucose, fructose, and/or citric acid in the cocoa pulp are consumed by fermenting yeast species, forming mainly ethanol, and LAB, forming mainly lactic acid (Pettipher, 1986; Schwan and Wheals, 2004; De Vuyst and Weckx, 2016a,b). Forced aeration by turning the fermenting cocoa pulp-bean mass, or by making use of a step-wise system of boxes, is advised to favour the growth of AAB. Their important role is to oxidize the ethanol formed by the yeasts into acetic acid. Accumulation of acetic acid in the cocoa pulp-bean mass leads to diffusion of this compound into the cocoa beans, making them increasingly acid and resulting in the killing of the seed embryo. Additionally, the oxidation of ethanol is a highly exothermic reaction, leading to an increase in temperature of the cocoa pulp-bean mass, exceeding 45°C.

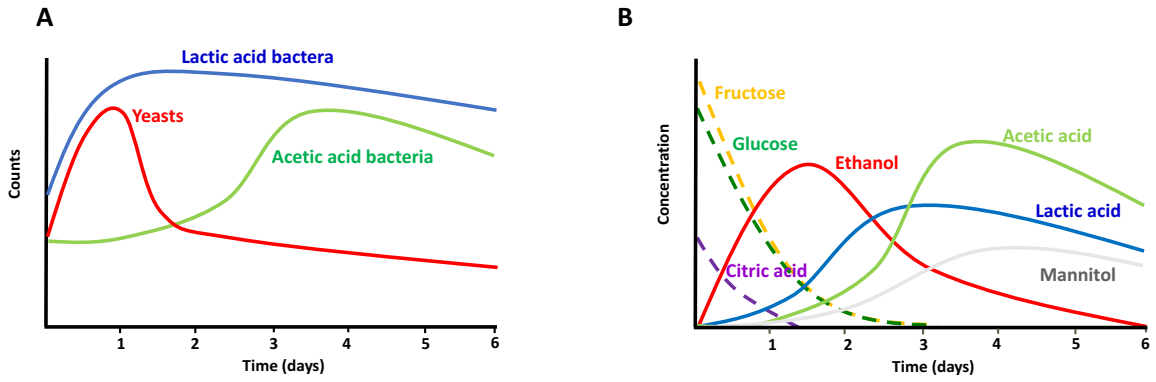


Figure 9. Overview of the characteristic microbial ecosystem (A) and metabolite dynamics (B) in the fermenting cocoa pulp-bean mass during a cocoa bean fermentation process (figure from De Vuyst and Weckx, 2016a).

In addition, the combined action of a lowering internal cocoa bean pH and rising fermentation temperature leads to the breaking of the cell barriers in the cocoa bean cotyledons. These are made up of two cell types, namely storage cells and pigment cells. Storage cells form the reserve material for the cocoa embryo, holding starch granules, fat globules (the cocoa butter), and protein vacuoles (Lima *et al*., 2011). The latter are home to storage proteins, made up of an albumin (52 %) and a vicilin-like globulin (43 %) fraction (Voigt *et al*., 1993; Rawel *et al*., 2019). Pigment cells contain alkaloid and polyphenol molecules, embedded in a large vacuole and necessary for cocoa bean stress resistance (Lima *et al*., 2011; Aprotosoaie *et al*., 2016; Wang *et al*., 2016). Theobromine, a methylxanthine, is the major alkaloid in cocoa beans, followed by caffeine. The polyphenolic compounds are mainly composed of flavonoids, represented by three main subgroups, namely catechins (37 %), anthocyanins (4 %), and proanthocyanidins (58 %) (Aprotosoaie *et al*., 2016; Rawel *et al*., 2019). Flavonoid intake has been associated to positive health effects, increasing antioxidant potential, lowering blood pressure, reducing LDL-cholesterol levels, and modulating inflammatory responses (Tsao, 2010; Aprotosoaie *et al*., 2016; Panche *et al*., 2016). The breaking of storage cells leads to diffusion of polyphenols out of the cotyledons. They are oxidized by polyphenol oxidase to reactive quinones, which polymerize with other polyphenols or form complexes with proteins (tannins). These processes lead to browning of the beans as well as reducing the astringency and bitterness (Lima *et al*., 2011; De Vuyst and Weckx, 2016a,b; Fayeulle *et al*., 2018; Barišić *et al*., 2019). Glycosidases hydrolyze anthocyanins, thereby releasing reducing sugars such as arabinose and galactose, whereas invertase hydrolyses sucrose into glucose and fructose (Lima *et al*., 2011; De Taeye *et al*., 2016). During cocoa bean fermentation, endogenous aspartic endoprotease and carboxypeptidase are activated due to their optimal activity at lower pH. The former protease cleaves storage proteins at the C-terminal end of hydrophobic residues, forming oligopeptides, of which subsequently hydrophobic amino acids are freed by the latter enzyme (Voigt and Biehl, 1995; Barišić *et al*., 2019; Rawel *et al*., 2019). The peptide and polyphenol profiles obtained have been proposed as biomarkers to monitor the status of the fermentation process (Kumari *et al*., 2016, 2018; D'Souza *et al*., 2017). A dynamic view emerges whereby the metabolite dynamics in the cocoa beans are characterized by an increase of oligopeptides and reducing sugars, and a decrease of polyphenols (Mayorga-Gross *et al*., 2016; Santander Muñoz *et al*., 2019).

After fermentation, drying is necessary to lower the water activity of the fermented beans to prevent spoilage of the beans (Aprotosoaie *et al*., 2016; De Vuyst and Weckx, 2016a,b). The typical chocolate flavor is obtained by roasting the dried cocoa beans. Here, cocoa flavor precursor molecules chemically react, forming aroma compounds that contribute to chocolate flavor development. Hydrophobic amino acids form Strecker aldehydes, *e.g.*, 2-methylbutanal (from isoleucine) and 3-methylbutanal (from leucine) (Barišić *et al*., 2019). Other cocoa flavor precursor molecules react through Maillard reactions and Strecker degradations to form a mixture of aldehydes (*e.g.*, 5-methyl-2-phenyl-2-hexanal), ketones (*e.g.*, acetophenone), and pyrazines (*e.g.*, tetramethylpyrazine and trimethylpyrazine), conferring the typical chocolate notes (Aprotosoaie *et al*., 2016).

6.3 Yeast species involved in cocoa bean fermentation

A wide biodiversity of yeast species has been reported in different cocoa bean fermentation processes around the world (Schwan and Wheals, 2004; De Vuyst and Weckx, 2016a). These include mostly species of the genera *Saccharomyces*, *Hanseniaspora*, *Candida*, and *Pichia* (Daniel *et al*., 2009; Illeghems *et al*., 2012; Meersman *et al*., 2013). These yeasts have been associated with sucrose hydrolysis and pectinolysis *via* invertase and pectinolytic enzymes, respectively, in the initial stage of the fermentation process (Schwan and Wheals, 2004; De Vuyst and Weckx, 2016a,b; Meersman *et al*., 2017). Species of *Saccharomyces* seem to prevail longer in the fermentation process than other yeast species. This could be due to their higher ethanol and temperature tolerances. *Hanseniaspora* species are more abundant in the initial stage of the fermentation process, which has been attributed to their resistance to lower pH values (Nielsen *et al*., 2005; Daniel *et al*., 2009; Meersman *et al*., 2013). Ultimately, and consistently, yeast species die off at the later stages of the fermentation process, most likely because temperatures exceeding 45°C hinder their growth (Daniel *et al*., 2009; Lima *et al*., 2011). The presence of yeast species may also be favorable for cocoa flavor development, since yeasts produce a large variety of fruity aroma compounds, mainly higher alcohols and esters. The contribution of these aroma compounds to the final chocolate flavor is unclear, since several steps in the cocoa value chain, such as drying, roasting, and conching, result in potential evaporation of aroma compounds and/or further reactions. However, analysis of chocolate samples indicated that long-chain ethyl and acetate esters remain present (Meersman *et al*., 2016).

6.4 Lactic acid bacteria involved in cocoa bean fermentation

Lactic acid bacteria are Gram-positive bacteria of the order *Lactobacillales* (Axelsson, 2004; Gänzle, 2015; Pot *et al*., 2019; Zheng *et al*., 2020). They are among the most versatile microorganisms known in human-associated ecosystems, such as fermented foods and the human microbiome. They obtained this wide-range occurrence due to niche specialization, in co-occurrence with genome size reduction. In this way, they abandoned the metabolic efficiency and versatility associated to other bacilli toward a lifestyle adapted to substrate availability in excess (Makarova *et al*., 2006; Makarova and Koonin, 2007). By increasing their specific substrate consumption rates, streamlining the metabolic pathways used for growth, thereby minimizing the amount of biomass building blocks that need to be synthesized *de novo*, and focusing solely on substrate-level phosphorylation as a way to gain energy, they refined their competitiveness. Rapid utilization of available carbohydrates results in accumulation of organic acids, in particular lactic acid and/or acetic acid, that decrease the environmental pH, thereby inhibiting the growth of their competitors, including spoilage and pathogenic bacteria (van Niel and Hahn-Hägerdal, 1999; Gänzle, 2015).

In the cocoa bean fermentation process, a high (relative) abundancy of LAB is consistently present during fermentation of the cocoa pulp-bean mass (Schwan and Wheals, 2004; De Vuyst and Weckx, 2016a). Species of the genera *Lactobacillus*, *Leuconostoc*, *Lactococcus*,

Enterococcus, and *Weissella* have been found in spontaneous fermentation processes (Bortolini *et al*., 2016; Ouattara *et al*., 2017; Figueroa-Hernández *et al*., 2019). More restricted LAB diversity has been reported where *Lactobacillus plantarum* or *Lactobacillus fermentum*, or both, dominated the fermentation process (Camu *et al*., 2007; Papalexandratou *et al*., 2011b; de Melo Pereira *et al*., 2012; Meersman *et al*., 2013; Miescher Schwenninger *et al*., 2016; Ouattara *et al*., 2017). Moreover, a typical succession of the dominance of *L. plantarum* followed by *L. fermentum* has been proposed based on experimental data obtained in different cocoa bean fermentation processes and which was explained due to differences in fermentation capacity and acid, pH, and temperature tolerance (De Vuyst and Weckx, 2016a,b).

LAB have been classified based on the main catabolic hexose fermentation pathways active. The differential use of fermentation pathways leading to altered metabolite production profiles is used for their classification (Axelsson, 2004). Homofermentative LAB preferentially consume glucose and produce lactic acid *via* the Emden-Meyerhof-Parnas pathway (glycolysis), resulting in a yield of two moles of ATP per mole of glucose. Heterofermentative LAB ferment glucose by the phosphoketolase pathway, resulting in the production of lactic acid and ethanol. When alternative electron acceptors are available, acetic acid can be produced instead of ethanol (Zaunmüller *et al*., 2006; Gänzle, 2015). The latter pathway is characterized by a lower yield of ATP per mole of glucose, because of obligatory formation of ethanol to obtain a redox balance of NAD; however, when acetic acid is produced, also two moles of ATP can be produced per mole of glucose consumed (Zaunmüller *et al*., 2006; Gänzle, 2015). Thus, heterofermentative LAB can use, if available in their environment, alternative electron acceptors such as fructose, citrate, malate, pyruvate, and oxygen to obtain redox balancing and increase their ATP yield on glucose (Starrenburg and Hugenholtz, 1991; Wisselink *et al*., 2002; Zaunmüller *et al*., 2006; Gänzle, 2015). In that regard, an interesting specialization is characterized by so-called fructophilic LAB that grow very poor on glucose alone, whereas their growth increases markedly when fructose is co-consumed. Genome analysis indicated that they miss the *adhE* gene, encoding a bifunctional alcohol/acetaldehyde dehydrogenase (Endo *et al*., 2018). Notably, *Fructobacillus* spp. have been identified in the initial stages of cocoa bean fermentation, when fructose is still plentifully present in the cocoa pulp (Papalexandratou *et al*., 2011a,b; Ouattara *et al*., 2017; Figueroa-Hernández *et al*., 2019).

6.5 Acetic acid bacteria involved in cocoa bean fermentation

Acetic acid bacteria are obligately aerobic bacteria that incompletely oxidize a range of alcohols and reducing sugars by means of a truncated respiratory chain (Matsushita *et al*., 1994; Cleenwerck and De Vos, 2008; Mamlouk and Gullo, 2013). They are associated with carbohydrate-rich environments, such as flowers, fruits, and fermented foods. In the cocoa bean fermentation process, mostly species of the genera *Gluconobacter*, typically *Gluconobacter oxydans*, and *Acetobacter*, typically *Acetobacter pasteurianus*, have been identified. The former species is found in the initial stage of the fermentation process and has been associated to the oxidation of glucose to gluconic acid, while the latter is responsible for the oxidation of

ethanol into acetic acid (Papalexandratou *et al*., 2011a,b, 2013; De Vuyst and Weckx, 2016a; Miescher Schwenninger *et al*., 2016).

To efficiently capture electrons from substrates in their environment, AAB contain a set of membrane-bound dehydrogenases in the periplasm (Adachi and Yakushi, 2016). Some of these are quinoproteins, containing a pyrroloquinoline quinone (PQQ) molecule as prosthetic group (Ameyama *et al*., 1981). PQQ is independently synthesized from a precursor peptide, named PqqA (Puehringer *et al*., 2008; Latham *et al*., 2015). The Type III PQQ-dependent ethanol dehydrogenase is a unique quinohemoprotein composed of three subunits and it contains four heme *c* binding sites (Toyama *et al*., 2004). The PQQ and heme *c* prosthetic groups form an electron cascade that reduces a membrane-soluble ubiquinone (Q) molecule (Frébortova *et al*., 1998; Toyama *et al*., 2004; Matsushita *et al*., 2008; Yakushi and Matsushita, 2010). The reduced ubiquinol (QH_2) molecule is subsequently re-oxidized by a terminal ubiquinol oxidase, using oxygen as terminal electron acceptor, which conserves energy in a proton motive force (PMF) (Yakushi and Matsushita, 2010; Matsutani *et al*., 2014). Phylogenetic studies have uncovered the preferred usage of ubiquinone-9 by *Acetobacter* species, compared to other acetic acid bacteria (Cleenwerck and De Vos, 2008). The electrochemical energy of the PMF is subsequently converted into chemical energy (ATP) by a proton-translocating ATP synthase (Matsushita *et al*., 1994; Mitchell, 2011).

Phenotypic and genomic analyses have revealed important metabolic differences between *G. oxydans* and *A. pasteurianus*. Both species contain an incomplete EMP pathway due to a missing phosphofructokinase (Deppenmeier and Ehrenreich, 2009; Illeghems *et al*., 2013). In addition to the pentose phosphate pathway (PPP) present in both AAB species, *G. oxydans* contains additionally the Entner-Doudoroff pathway for glycolysis. Finally, *A. pasteurianus* contains a complete but modified tricarboxylic acid (TCA) cycle, characterized by a succinyl-CoA:acetate CoA transferase, whereas the TCA cycle is incomplete for *G. oxydans* (Prust *et al*., 2005; Mullins *et al*., 2008; Illeghems *et al*., 2013).

The growth of *G. oxydans* on glucose is characterized by two phases of growth (Hanke *et al*., 2013). In a first phase of growth, glucose is mostly oxidized to gluconate in the periplasm, which accumulates in the medium. In a second phase of growth, the gluconate accumulated is re-oxidized, mostly to 2-ketogluconate in the periplasm and partly in the cytoplasm by the oxidative PPP, operating in a partially cyclic manner (Hanke *et al*., 2013). Notably, a proton-translocating $NAD(P)^+$ transhydrogenase is highly expressed in the second growth phase. Since the oxidative PPP is known to be a significant production source of NADPH, it has been proposed that the enzyme would be used to convert the NADPH produced to NADH, thereby maintaining an appropriate $NAD^+/NADP^+$ balance and fueling the respiratory chain (Deppenmeier and Ehrenreich, 2009; Rauch *et al*., 2010; Hanke *et al*., 2013).

A similar result has been found for *A. pasteurianus* strains growing in a cocoa pulp simulation medium with lactic acid and ethanol as main substrates (Adler *et al*., 2014; Moens *et al*., 2014). Lactic acid is oxidized in the cytoplasm, with a significant overflow of acetoin, whereas ethanol

is almost completely oxidized to acetic acid in the periplasm. Both accumulated products are subsequently completely re-oxidized into carbon dioxide by means of the TCA cycle (Adler *et al*., 2014; Moens *et al*., 2014).

In terms of carbon assimilation to biomass precursors, AAB present a somewhat puzzling case, since information regarding the exact growth requirements of these bacteria, and species of *Gluconobacter* and *Acetobacter* in particular, is lacking. In *E. coli* and *S. cerevisiae*, the glyoxylate pathway is employed for carbon assimilation from ethanol as substrate (Kunze *et al*., 2006). *Acetobacter aceti* also contains the glyoxylate pathway, and the presence of this pathway is associated to an increased growth rate and biomass yield on ethanol (Sakurai *et al*., 2011, 2012, 2013).

6.6 Starter culture development

The use of starter cultures in food fermentation processes has a long history (Hansen, 2002; Leroy and De Vuyst, 2004). Its primary use is to accelerate the fermentation process compared to the spontaneous one, thereby reducing the risk for spoilage, and providing new functionalities that improve the organoleptic properties of the end-products. Therefore, the application of starter cultures to the otherwise spontaneous cocoa bean fermentation process is desirable to increase the yield of well-fermented cocoa beans in a shorter time frame, thereby releasing the full flavor potential of the cocoa variety (De Vuyst and Weckx, 2016a,b). Starter cultures are composed of endogenous strains or exogenous strains, selected for a particular trait (Schwan, 1998; Lefeber *et al*., 2012; Kresnowati *et al*., 2013; Meersman *et al*., 2016; da Veiga Moreira *et al*., 2017). Most starter cultures applied consist of a *Saccharomyces* species, a *Lactobacillus* species, and an *Acetobacter* species (Schwan, 1998; Lefeber *et al*., 2012; Crafack *et al*., 2013; Sandhya *et al*., 2016; da Cruz Pedrozo Miguel *et al*., 2017; da Veiga Moreira *et al*., 2017; Ho *et al*., 2018). With the advent of molecular techniques, the specific functionalities that are inherent to a successful starter culture have been identified and screened for in other candidate functional starter cultures (Leroy and De Vuyst, 2004; Leroy *et al*., 2006; Johansen, 2018). In-depth knowledge of a fermentation process allows to reason about specific functionalities that may perturb it in a desirably way. For example, in cocoa bean fermentation processes, strains of *S. cerevisiae* are preferred that combine high glycolytic flux with temperature and ethanol tolerance (Meersman *et al*., 2015). However, the sensitivity of *S. cerevisiae* to lower pH values prompt the usage of additional yeast species as starter cultures to cover the whole cocoa bean fermentation process (Ho *et al*., 2018). Also, the pectinolytic activity of yeasts and their potential to produce a range of aroma compounds has engaged the selection and application of strains with these desirable traits (Crafack et al., 2013; Meersman *et al*., 2016). In-depth characterization combining whole-genome sequencing and physiological characterization is imperative for proper application of selected starter cultures (Johansen, 2018).

Chapter 2

Aims and objectives

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

The cocoa bean fermentation process is an essential step to obtain cured cocoa beans, the raw material for the production of cocoa-derived products, among which chocolate. Therefore, the cocoa bean fermentation process is economically important for the cocoa industry in general and for cocoa farmers in particular, as knowledge on increasing the quality of the cured cocoa beans may lead to an increase of their income. However, cocoa bean fermentation is one of the outsiders in the fermented food industry, as it still relies on a spontaneous fermentation process, limiting the yield of high-quality cured cocoa beans. To increase the control over the fermentation process, the application of functional starter cultures has been proposed that could prevail during the fermentation process and therefore enhance the formation of an optimally cured end-product. Yet, current candidate functional starter cultures are insufficiently characterized at the metabolic level to allow their optimal use in a cocoa bean fermentation process that is influenced by many factors, among which the fermentation technique, cocoa pulp composition, and environmental parameters. Therefore, it is important to increase the knowledge on the metabolism of the candidate functional starter cultures to obtain a better control over the fermentation process.

The aim of this PhD thesis was to obtain more insight into the metabolic activities of candidate functional starter cultures in cocoa fermentation conditions, in particular dedicated strains of acetic acid bacterial (AAB) and lactic acid bacterial (LAB) species, by applying an *in silico* research approach. Complementary to *in vivo* and *in vitro* experiments, *in silico* experiments allow to obtain a systemic view of the metabolism of particular starter culture strains by combining genome sequencing data with systems biology approaches. Therefore, genome-scale metabolic modeling of candidate functional starter cultures was performed to test the current knowledge regarding these strains, re-interpreting existing experimental data to obtain new metabolic insights, and predict new metabolic characteristics leading to testable hypotheses.

First, Chapter 3 aims to obtain a systemic view of the metabolism of *Acetobacter pasteurianus* 386B, an AAB strain that has been selected as a candidate functional starter culture for cocoa bean fermentation processes. This can be realized by constructing a genome-scale metabolic network that combines genome re-annotation, comparative genomics, and manual curation. Two other AAB strains, namely *Acetobacter ghanensis* LMG 23848^{T} and *Acetobacter senegalensis* 108B, have also been proposed as candidate functional starter cultures. To enable analysis of their specific metabolic properties, Chapter 4 deals with their genome-scale metabolic network reconstruction. In particular, this chapter assesses the possibility of a semi-automated genome-scale metabolic network reconstruction combining the genome-scale metabolic model of *A. pasteurianus* 386B obtained in Chapter 3 with large-scale comparative

genomics of the *Acetobacter* genus. In Chapter 5, genome-scale metabolic modeling of *A. pasteurianus* 386B allows to predict its metabolic activity under cocoa fermentation conditions and to interpret previously reported experimental data. Going a step further toward understanding the dynamics of the cocoa bean fermentation process, Chapter 6 aims to model the diauxic growth dynamics under cocoa fermentation conditions for *A. pasteurianus* 386B and *A. ghanensis* 23848^{T}. Finally, based on the work performed in the previous chapters, Chapter 7 deals with the *in silico* process of genome-scale metabolic network reconstruction and genome-scale metabolic modeling for two LAB candidate functional starter culture strains, namely *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80, to gain systemic understanding of their metabolism in the context of the cocoa bean fermentation process.

Ultimately, the research performed in this PhD thesis will enable the use of genome-scale metabolic models of the candidate functional starter cultures as *in silico* tools to interpret experimental data obtained from multi-omics studies concerning their metabolic activities during cocoa fermentation processes.

Chapter 3

Genome-scale metabolic reconstruction of *Acetobacter pasteurianus* 386B, a candidate functional starter culture for cocoa bean fermentation

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

Part of this chapter has been published in:

Pelicaen, R., Gonze, D., Teusink, B., De Vuyst, L. & Weckx, S. (2019). Genome-scale metabolic reconstruction of *Acetobacter pasteurianus* 386B, a candidate functional starter culture for cocoa bean fermentation. *Frontiers in Microbiology* **10**, 2801.

Supplementary material for this Chapter is available at:
https://www.frontiersin.org/articles/10.3389/fmicb.2019.02801/full#supplementary-material

Summary

Acetobacter pasteurianus 386B is a candidate functional starter culture for the cocoa bean fermentation process. To allow *in silico* simulations of its related metabolism in response to different environmental conditions, a genome-scale metabolic model for *A. pasteurianus* 386B was reconstructed. This is the first genome-scale metabolic model reconstruction for a member of the genus *Acetobacter*. The metabolic network reconstruction process was based on extensive genome re-annotation and comparative genomics analyses. The information content related to the functional annotation of metabolic enzymes and transporters was placed in a metabolic context by exploring and curating a Pathway/Genome Database of *A. pasteurianus* 386B using the Pathway Tools software. Metabolic reactions and curated gene-protein-reaction associations were bundled into a genome-scale metabolic model of *A. pasteurianus* 386B, named iAp386B454, containing 454 genes, 322 reactions, and 296 metabolites embedded in two cellular compartments. The reconstructed model was validated by performing growth experiments in a defined medium, which revealed that lactic acid as the sole carbon source could sustain growth of this strain. Further, the reconstruction of the *A. pasteurianus* 386B genome-scale metabolic model revealed knowledge gaps concerning the metabolism of this strain, especially related to the biosynthesis of its cell envelope and the presence or absence of metabolite transporters.

1 Introduction

Acetic acid bacteria (AAB) are obligately aerobic bacteria that are native to many different fermented food ecosystems (De Roos and De Vuyst, 2018). This is due to the typical carbohydrate richness of these ecosystems, as these bacteria are known to rapidly, and mostly incompletely, oxidize a wide range of carbohydrates. While *Gluconobacter* species are known to oxidize a wide range of sugars and polyols, *Acetobacter* species are specialized in the conversion of ethanol into acetic acid (Mamlouk and Gullo, 2013). As a result, species of the latter genus are present in high abundance in different ethanol-rich ecosystems, which can be desirable (*e.g.,* vinegar-making) or undesirable (*e.g.,* spoilage of wine or beer) (Mamlouk and Gullo, 2013; De Roos and De Vuyst, 2018). In all such ecosystems, a sufficient amount of oxygen is necessary since these bacteria rely completely on an aerobic respiratory chain for their energetic needs (Matsushita and Matsutani, 2016).

The cocoa bean fermentation process is a necessary step in the formation of cured cocoa beans, the raw material for cocoa-derived products. In this process, the conversion of ethanol to acetic acid mediated by *Acetobacter* species is crucial, as acetic acid is known to diffuse into the cocoa beans, contributing to the death of the cocoa bean embryo and the formation of cocoa flavor precursors inside the bean (De Vuyst and Weckx, 2016a). One particular species, *Acetobacter pasteurianus*, has been consistently isolated from different spontaneous cocoa bean fermentation processes around the world (Camu *et al*., 2007; Lefeber *et al*., 2011a; Meersman *et al*., 2013; Miescher Schwenninger *et al*., 2016; Ozturk and Young, 2017; Papalexandratou *et al*., 2013; Visintin *et al*., 2016). Its apparent adaptation to this fermentation process has been explained by its high ethanol, acid, and heat tolerance (Camu *et al*., 2007; Illeghems *et al*., 2013). The strain *A. pasteurianus* 386B has been isolated from a spontaneous cocoa bean fermentation process performed in Ghana (Camu *et al*., 2007), and has subsequently been selected as a candidate functional starter culture because of its rapid co-consumption of ethanol and lactate, the production of acetate and acetoin, and the achievement of high cell densities upon fermentation (Lefeber *et al*., 2010; Moens *et al*., 2014). To better understand its metabolic properties, the *A. pasteurianus* 386B genome was sequenced and functionally annotated. Metabolic pathway reconstruction highlighted the presence of an incomplete Embden-Meyerhof-Parnas (EMP) pathway, a modified tricarboxylic acid (TCA) cycle, and a truncated aerobic respiratory chain (Illeghems *et al*., 2013).

The metabolic pathways obtained only included the central carbon metabolism of *A. pasteurianus* 386B. Also, the macromolecule biosynthesis pathways of this strain in particular and of AAB in general are still largely unknown. Defined growth media have been used successfully to uncover the specific growth requirements of a range of microorganisms (Richards *et al*., 2014; van Niel and Hahn-Hägerdal, 1999; Verduyn *et al*., 1992). However, information regarding defined growth media for *Acetobacter* species is limited (Belly and Claus, 1972; Cleenwerck and De Vos, 2008; Gosselé *et al*., 1981; Heo and Son, 2002; Rainbow and Mitson, 1953; Son *et al*., 2003). Moreover, due to taxonomic reclassifications, it is often unclear to which extent the transfer of knowledge from one species to another may be performed.

Furthermore, the reconstructed metabolic pathways of *A. pasteurianus* 386B "on paper" are not amenable to obtain a system-wide view of the metabolism of this strain. A genome-scale metabolic model (GEM) is a modeling tool that allows to perform *in silico* metabolic flux simulations at steady-state, to explore the metabolic capabilities of the reconstructed metabolic network of a particular microorganism (Francke *et al*., 2005; Pitkänen *et al*., 2010; Thiele and Palsson, 2010). Such models have been used successfully to explain and even predict the metabolic properties of microorganisms (Teusink *et al*., 2005, 2006). For AAB, the only GEMs that are currently available are those for *Gluconobacter oxydans* 621H, an industrially important bacterium due to its property of oxidising a wide range of carbohydrates, and for *Komagataeibacter nataicola* RZS01, a bacterial cellulose producer (Wu *et al*., 2014b; Zhang *et al*., 2017).

To predict the specific growth rate of a bacterial cell population and its corresponding metabolic flux distribution, the flux balance analysis (FBA) algorithm demands the addition of a (species-specific) biomass reaction to the model, which is used as an approximation of the specific growth rate (Feist and Palsson, 2010). Therefore, genome-scale metabolic reconstruction of *A. pasteurianus* 386B was performed to retrieve, next to its central carbon metabolism, all macromolecule biosynthesis pathways leading to biomass production, and this in combination with *in vitro* growth experiments to validate the GEM properties. The *A. pasteurianus* 386B GEM obtained is the first GEM for a species of the genus *Acetobacter* and is expected to be a useful *in silico* tool to gain in-depth knowledge on the metabolic properties of this strain. Its improved application as a functional starter culture may then in turn improve the cocoa bean fermentation process as a whole.

2 Materials and methods

2.1 *Acetobacter pasteurianus* 386B genome re-annotation and *in silico* genome-scale metabolic reconstruction

The complete genome of *A. pasteurianus* 386B was sequenced and annotated previously, using a local installation of the bacterial genome annotation system GenDB v2.2 (Illeghems *et al*., 2013; Meyer, 2003). Since then, different genome annotations of this strain became publicly available in different databases. To perform a thorough re-annotation of the *A. pasteurianus* 386B genome and to be able to assess differences between the annotation sources, the annotation data were collected in a relational database (MySQL) that was built in-house (Figure 1).

Publicly available genome annotation sources included the Carbohydrate-Active enZYmes database (CAZy; Lombard *et al.*, 2014), the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa *et al.*, 2017), the Integrated Microbial Genomes platform of the Joint Genome Institute (JGI IMG; Markowitz *et al.*, 2012), the RefSeq database of the National Center for Biotechnology Information (NCBI; O'Leary *et al.*, 2016), the Pathosystems Resource Integration Center database (PATRIC; Wattam *et al.*, 2017), the proGenomes database (Mende *et al*., 2017), and TransportDB (Elbourne *et al*., 2017).

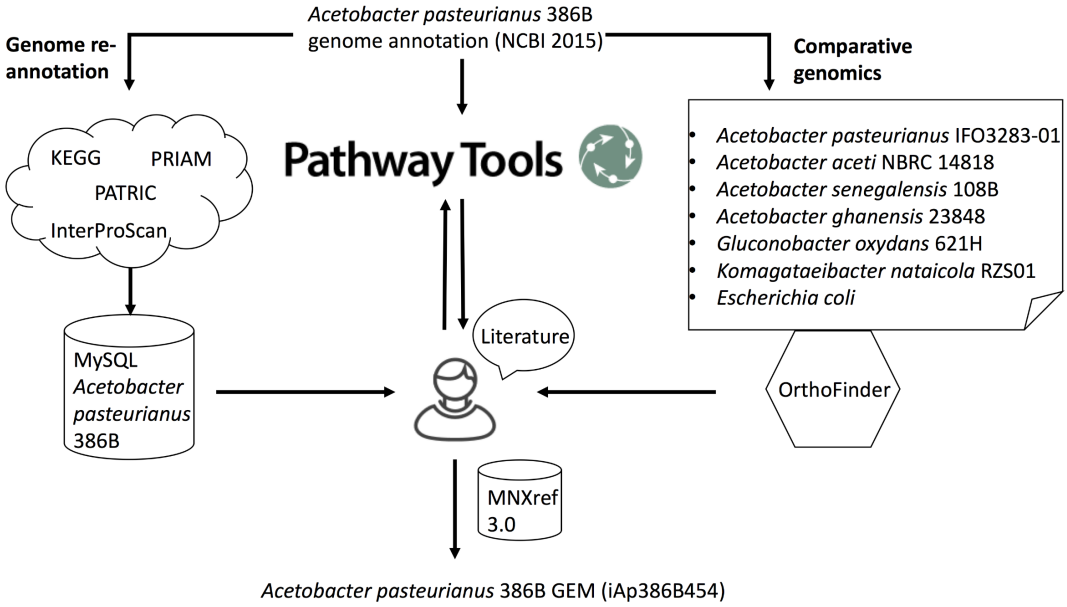


Figure 1. Schematic overview of the metabolic network reconstruction process. Genome re-annotation of *Acetobacter pasteurianus* 386B was performed using a combination of several databases (e.g. KEGG, PATRIC) and tools (e.g. PRIAM, InterProScan). Functional annotations were stored in a relational database. Comparative genomics using OrthoFinder allowed to identify orthogroups of protein sequences between *A. pasteurianus* 386B and other phylogenetically related bacterial species. Pathway Tools was used to predict metabolic pathways as well as the reactions and gene-protein-reaction associations that constitute these pathways. This information was stored in a Pathway/Genome database (PGDB). Edges entering the human curator node represent the information sources that were used to guide manual curation. Using additional information from the literature, the *A. pasteurianus* 386B PGDB was curated and its information transferred to a new genome-scale metabolic model of *A. pasteurianus* 386B, iAp386B454, using the MNXref 3.0 namespace of reactions and metabolites.

The latter database relies on the TC system for annotation, providing a defined ontology to describe transporter functions in analogy to the EC system for enzyme annotation (Saier *et al*., 2016). Two annotation versions were used in the case of the NCBI RefSeq annotation source, one published in April 2015 (further referred to as NCBI 2015) and one published in April 2017 (further referred as NCBI 2017). Differences in these genome annotations reflect improvements made in the NCBI prokaryotic genome annotation pipeline (Tatusova *et al*., 2016). Furthermore, the *A. pasteurianus* 386B genome was re-annotated in-house, using the subcellular localisation predictor CELLO (Yu *et al*., 2004), eggNOG-mapper (Huerta-Cepas *et al*., 2017), the enzyme annotation tool PRIAM (Claudel-Renard, 2003), and the tools embedded in InterProScan 5.22-61.0 (Jones et al., 2014) (Figure 1).

As a reference set of predicted protein-encoding genes, the NCBI 2015 genome annotation was used, containing 2,854 protein-encoding genes, as it is the same that was used by SRI International to reconstruct the *A. pasteurianus* 386B Pathway/Genome Database (PGDB) available on BioCyc (Caspi *et al*., 2016). Subsequently, for each protein-encoding gene, functional annotations from the different annotation sources were added to the relational

database. In addition, the *A. pasteurianus* 386B PGDB was manually curated in the Pathway Tools software, assisted by different automatic refinements using the Pathologic tool inside this software (Karp *et al*., 2016) (Figure 1). Features associated to a predicted MetaCyc pathway (Caspi *et al*., 2016), such as the taxonomic range, pathway score, pathway variants and pathway description, were manually assessed and compared to information available in the literature to decide whether the pathway in question should be omitted.

Also, comparative genomics was performed using OrthoFinder, allowing to predict orthogroups from protein-encoding genes in whole genomes (Emms and Kelly, 2015) (Figure 1). Protein sequences of a selection of bacterial species obtained from NCBI RefSeq were compared, namely *A. pasteurianus* 386B, *A. pasteurianus* IFO3283-01, *A. aceti* NBRC 14818, *A. senegalensis* 108B, *A. ghanensis* LMG 23848^{T}, *Gluconobacter oxydans* 621H, *Komagataeibacter nataicola* RZS01, and *Escherichia coli* str. K-12 substr. MG1655 (further referred to as *E. coli*). For the latter three bacterial strains, genome-scale metabolic models have been reconstructed before (Orth *et al*., 2011; Wu *et al*., 2014b; Zhang *et al*., 2017).
Finally, based on the information contained in the in-house built relational database, the *A. pasteurianus* 386B PGDB, and the outcome of the comparative genomics analysis, the *A. pasteurianus* 386B GEM was reconstructed and manually curated using the MNXref 3.0 namespace as a biochemical reaction repository (Bernard *et al*., 2014; Moretti *et al*., 2016) (Figure 1). Characterised enzymes and associated protein sequences available in the literature were used to perform sequence alignment searches using blastp (Altschul *et al*., 1990). The best blast hits in the *A. pasteurianus* 386B genome are indicated by their percentage similarity, percentage identity, and percentage coverage (Supplementary Table S1).

MetaCyc reaction components and GPR associations of biosynthesis pathways in the curated *A. pasteurianus* 386B PGDB were transferred to the *A. pasteurianus* 386B GEM by mapping the MetaCyc reaction components to MNXref 3.0 reaction identifiers. Reactions described in the literature but not present in the MNXref 3.0 reaction repository were manually added to the model. Enzyme names and enzyme commission (EC) numbers were assigned to each reaction, based on the ExplorEnz database (McDonald *et al*., 2009). The mass and charge balances were checked for all reactions in the resulting *A. pasteurianus* 386B GEM. The co-factor specificity of these reactions was manually curated, based on functional annotations and literature sources. Reaction reversibility constraints were manually curated, taking into account the directionality defined in the MetaCyc database, and to prevent reaction flux cycles from occurring in the GEM. The list of GPR associations was saved in a spreadsheet file (Supplementary File 2). The *A. pasteurianus* 386B GEM, named iAp386B454, was reconstructed and analysed using the COBRApy package version 0.11.3 (Ebrahim *et al*., 2013). Next, it was exported in the Systems Biology Markup Language (SBML) level 3 format (Supplementary File 3) and validated using the online SBML Validator (Bornstein *et al*., 2008). Reconstructed metabolic pathways were visualised using Escher (King et al., 2015).

2.2 Evaluation of the manual curation process and comparison to other reconstructions

The presumed increase in functional annotation quality of the genes included in iAp386B454 was evaluated based on the EC numbers of enzyme-encoding genes, as the EC system provides a defined enzymatic reaction classification (McDonald *et al*., 2009). EC numbers of reactions associated to 304 selected genes in the *A. pasteurianus* 386B GEM were compared to the EC numbers of the genes in the original annotation sources stored in the relational database. A number of classes were defined based on the hierarchical EC system that expressed the agreement between the EC number of a gene in the GEM and that in the annotation source. Then, for each annotation source, genes were assigned to the predefined classes. Preference was given to assign a gene to a more precise class if at least one of the EC numbers of the gene in the annotation source did fulfil the requirement set upon that class. Enzyme systems, for which the reaction is catalysed by a complex containing more than one enzyme (*e.g.*, pyruvate dehydrogenase), were excluded from the analysis, as these reactions inherited different EC numbers from their gene constituents.

The manually reconstructed *A. pasteurianus* 386B GEM was compared to GEMs obtained from a number of automatic GEM reconstruction tools. These included the RAST and ModelSEED tools in the KBase software (Arkin *et al*., 2018; Henry *et al*., 2010; Overbeek *et al*., 2014), CarveMe (Machado *et al*., 2018), MetaNetX (Ganter *et al*., 2013), and Pathway Tools (Karp *et al*., 2016). The NCBI 2015 annotation version of *A. pasteurianus* 386B was used for the reconstructions, which were performed without reaction gap-filling. COBRApy was used to parse the SBML files of the reconstructed GEMs, except for SBML files obtained from Pathway Tools, which were parsed using CBMPy version 0.7.19 (Olivier *et al*., 2005). Model properties were inferred from the reconstructions and compared to iAp386B454.

2.3 Biomass reaction

A biomass reaction was defined for *A. pasteurianus* 386B, using a combination of genomic and literature data (Supplementary File 4 and Supplementary Figure S1). Protein, DNA, RNA, lipid, fatty acid, peptidoglycan, and carbohydrate mass fractions were taken from a GEM of *K. nataicola* RZS01 (Zhang *et al*., 2017). Stoichiometric coefficients of the biomass reaction were obtained by converting the mass fraction of the different macromolecules into a molar fraction, using their estimated molar masses. For each macromolecule, a separate biosynthesis reaction was defined. Stoichiometric coefficients of these reactions resulted from estimations of the molar fractions of the macromolecule building blocks. Molar fractions of amino acids for proteins and of nucleotides for DNA and RNA were estimated using the genome sequence of *A. pasteurianus* 386B, as proposed before (Thiele and Palsson, 2010). For DNA, molar fractions of building blocks were estimated using the G+C percentage of the genome (*in casu* 52.86 %). For RNA, molar fractions of building blocks were estimated using their respective frequencies in rRNA-, tRNA-, and mRNA-encoding genes. RNA mass fractions were taken from *E. coli* (Milo and Phillips, 2016). The fatty acid and phospholipid compositions were taken from studies on related *Acetobacter* species (Hanada *et al*., 2001; Yamada *et al*., 1981). The average molar mass of a generic fatty acid was used to estimate the average molar mass of a generic phospholipid. Molar masses of macromolecule building blocks were queried in PubChem and ChEBI (Hastings *et al*., 2016; Kim *et al*., 2019). Energy requirements for

protein, DNA, and RNA biosyntheses were taken from *E. coli* (Neidhardt *et al*., 1990). Genes were manually assigned to the different macromolecule biosynthesis reactions for cellular processes that were not explicitly included in the model, among which tRNA loading, protein elongation, replication, transcription, and translation.

2.4 Growth experiments

Growth experiments were performed with *A. pasteurianus* 386B in a modified defined medium (Verduyn *et al*., 1992). This medium contained (per litre): $(NH_4)_2SO_4$, 5 g; KH_2PO_4, 1.375 g; $MgSO_4.7H_2O$, 0.5 g; EDTA, 15 mg; $ZnSO_4.7H_2O$, 4.5 mg; $CoSO_4.7H_2O$, 0.35 mg; $MnCl_2.4H_2O$, 1.0 mg; $CuSO_4.5H_2O$, 0.3 mg; $CaCl_2.2H_2O$, 4.5 mg; $FeSO_4.7H_2O$, 3.0 mg; MoO_3, 0.24 mg; H_3BO_3, 1.0 mg; KI, 0.1 mg; and 30 mM phosphate buffer (pH 6.0). The pH of the medium was set to 5.0. A filter-sterilised vitamin mixture was added after heat sterilisation (121°C, 2.1 bar, 20 min) of the medium. The final vitamin concentrations were (per litre): biotin, 0.0005 mg; calcium pantothenate, 0.01 mg; nicotinic acid, 0.01 mg; *myo*-inositol, 0.25 mg; thiamine-HCl, 0.01 mg; pyridoxine-HCl, 0.01 mg; and para-aminobenzoic acid, 0.002 mg. Eight different carbon sources (glucose, fructose, mannitol, citric acid, glycerol, lactic acid, ethanol, and sodium acetate) were used at a final concentration of 60 mM to assess if they could sustain growth. The pH of the citric acid, lactic acid (both by using NaOH to increase the pH), and sodium acetate (using HCl to decrease the pH) stock solutions was set to 5.0. To prepare the inocula, *A. pasteurianus* 386B was grown overnight at 30°C in 10 ml of an undefined medium (pH 5.5), which contained (per litre): lactic acid, 5 g; sodium acetate, 10 g; granulated yeast extract, 5 g; $MgSO_4.7H_2O$, 1 g; $NH_4H_2PO_4$, 20 g; and K_2HPO_4, 10 g. The shaking rate was set to 160 rpm. The overnight culture was centrifuged (4000 x *g*, 20 min, 4°C) and washed with a filter-sterilised saline solution (0.85 %, m/v, NaCl). The cells were resuspended in the sterile saline solution and inoculated in 2 mL of the defined medium mentioned above at an optical density at 600 nm (OD_{600}) of 0.01 in triplicate in test tubes with a total volume of 20 ml. *Acetobacter pasteurianus* 386B was allowed to grow at 30°C without shaking for 48 h as the test tubes contained only 2 ml culture medium, resulting in a relatively large surface to medium ratio. A threshold value for the OD_{600} of 0.1 was used to identify whether or not the strain had grown.

2.5 Model validation

In silico growth experiments were performed using flux balance analysis (FBA). With FBA, an optimization of the flux distribution of a GEM can be performed to maximize an objective function, typically the biomass reaction, thus predicting the specific growth rate of a bacterial cell population (Gottstein *et al*., 2016). FBA was performed with the *A. pasteurianus* 386B GEM, setting the biomass reaction as the objective of the simulation. Parameter values of this GEM included the consumption flux values of the different nutrients. Here, the consumption of ammonium as nitrogen source, sulfate as sulfur source, and phosphate as phosphate source were allowed without constraints. For aerobic respiration, an oxygen influx was allowed without constraint. D-glucose, D-mannitol, glycerol, D-lactate, ethanol, and acetate were tested as carbon sources separately to be able to compare the *in silico* results with the *in vitro* growth

experimental data. The maximum exchange flux of the carbon source was set to 60 C-mmol/g_{CDW}/h for each simulation. This value corresponded with a consumption of 10 mmol/g_{CDW}/h of glucose for *E. coli* (Varma *et al*., 1993).

3 Results and discussion

3.1 Genome re-annotation

An extensive re-annotation of the genome of *A. pasteurianus* 386B was performed, based on a combination of several databases and tools, comparative genomics analyses, and manual curations using different software packages (Figure 1). Of a total of 2,875 protein-encoding genes in the original genome annotation of *A. pasteurianus* 386B (Illeghems *et al*., 2013), 2,854 genes (NCBI 2015) were re-annotated in the current study, of which 454 genes were included in the *A. pasteurianus* 386B GEM based on their involvement in the metabolic pathways discussed below. In what follows, metabolic reactions in different parts of the *A. pasteurianus* 386B metabolism are discussed, based on the evidence that the enzyme-encoding genes were present in the genome.

3.2 Central carbon metabolism

The incomplete EMP pathway (Supplementary Figure S2) and the pentose phosphate pathway (Supplementary Figure S3) of *A. pasteurianus* 386B were reconstructed, as described previously (Illeghems *et al*., 2013). Periplasmic oxidation of D-glucose could be assigned to a pyrroloquinoline quinone (PQQ)-dependent glucose 1-dehydrogenase (EC 1.1.5.2) (Figure 2A). Based on the re-annotation effort of the current study, different gene candidates could be linked to this reaction (*e.g.*, APA386B_2133, Supplementary Table S1), among which genes with a previously unknown function (Illeghems *et al*., 2013). No transporter could be identified for D-gluconate uptake. However, since the gene encoding cytoplasmic gluconokinase was present in the genome (locus tag APA386B_1158), the D-gluconate transport reaction was nonetheless added to the *A. pasteurianus* 386B GEM. Cytoplasmic oxidation of D-mannitol could be assigned to a cytoplasmic mannitol 2-dehydrogenase. However, the possibility of periplasmic oxidation of D-mannitol to D-fructose was included in the GEM as well, since extracellular fructose formation from mannitol has been reported previously (Moens *et al*., 2014).

A gene encoding a lactate permease (TC 2.A.14), more precisely a lactate:H^+ symporter, was identified in the genome, which was homologous to the LctP transporter in *E. coli* (Dong *et al*., 1993; Núñez *et al*., 2002). Most probably *A. pasteurianus* 386B oxidizes D-lactate using a D-lactate dehydrogenase that is dependent on quinone (EC 1.1.5.12) or cytochrome *c* (EC 1.1.2.4), thus linking this reaction directly to the respiratory chain (Figure 2B). Overflow of lactate to acetoin has been found and metabolic flux analysis has revealed that acetoin is solely produced by decarboxylation of (2S)-2-acetolactate (Adler *et al*., 2014; Moens *et al*., 2014). Therefore, this metabolic route, for which the genes were present, and containing the enzymatic reactions of acetolactate synthase (EC 2.2.1.6) and acetolactate decarboxylase (EC 4.1.1.5),

was added to the *A. pasteurianus* 386B GEM (Supplementary Figure S4). Finally, an acetoin dehydrogenase complex (EC 2.3.1.190) was identified that could be involved in the oxidation of acetoin. A glycerol uptake transporter (TC 1.A.8) was identified in the genome that was homologous to the GlpF transporter of *E. coli* (Fu *et al*., 2000). Similarly as for lactate, oxidation of glycerol could feed electrons to the respiratory chain (Figure 2B); alternatively, glycerol may be used for glycerolipid biosynthesis.

Acetobacter pasteurianus 386B could oxidise ethanol to acetic acid in the periplasm by means of a membrane-bound PQQ-dependent alcohol dehydrogenase or in the cytoplasm by means of $NAD(P)^+$-dependent dehydrogenases (Figure 2C). Periplasmic acetate could be imported by an acetate/succinate:H^+ symporter, for which the encoding gene was identified (APA386B_1116) and which was homologous to the *E. coli* SatP transporter (Sá-Pessoa *et al*., 2013). Alternatively, excess cytoplasmic acetate may be exported by a primary active ABC transporter (TC 3.A; APA386B_103) (Nakano *et al*., 2006). A reaction cycle that regulates acetate overflow metabolism, involving acetic acid, acetyl phosphate, and acetyl-CoA, has been described in *E. coli* (De Mey *et al*., 2007; Valgepea *et al*., 2010). Putative orthologs for these enzymes were retrieved in *A. pasteurianus* 386B, namely genes encoding acetate kinase (EC 2.7.2.1), phosphate acetyltransferase (EC 2.3.1.8), and acetate CoA ligase (EC 6.2.1.1).

A major metabolic adaptation of acetic acid-resistant *Acetobacter* species is their modified tricarboxylic acid (TCA) cycle (Supplementary Figure S4), containing succinyl-CoA:acetate CoA-transferase (EC 2.8.3.18) and malate:quinone oxidoreductase (EC 1.1.5.4) (Mullins *et al*., 2008). Two enzymes involved in anaplerotic reactions were encoded in the genome, namely a reversible NAD^+-dependent malic enzyme (EC 1.1.1.38), interconverting malate and pyruvate, and phosphoenolpyruvate carboxylase (EC 4.1.1.31) that could carboxylate phosphoenolpyruvate to oxaloacetate. However, in the NCBI 2017 genome annotation version of *A. pasteurianus* 386B, the latter enzyme-encoding gene was annotated as a pseudogene, containing a frameshift mutation.

3.3 Aerobic respiration

Acetobacter pasteurianus is an obligate aerobe that has evolved a truncated aerobic respiratory chain (Figure 2D). In *A. pasteurianus* 386B, a single gene cluster *cyaBACD* (APA386B_1578 - APA386B_1581) was found that encodes cytochrome ba_3 ubiquinol oxidase (UOX ba_3; EC 7.1.1.3). This heme-copper terminal oxidase probably contains a heme A moiety, since the genes encoding heme O (*ctaB*; APA386B_608) and heme A (*ctaA*; APA386B_1565) synthase, which are remnants of a cytochrome *c* oxidase gene cluster, have been shown to be still functional in *A. pasteurianus* (Matsutani *et al*., 2014).

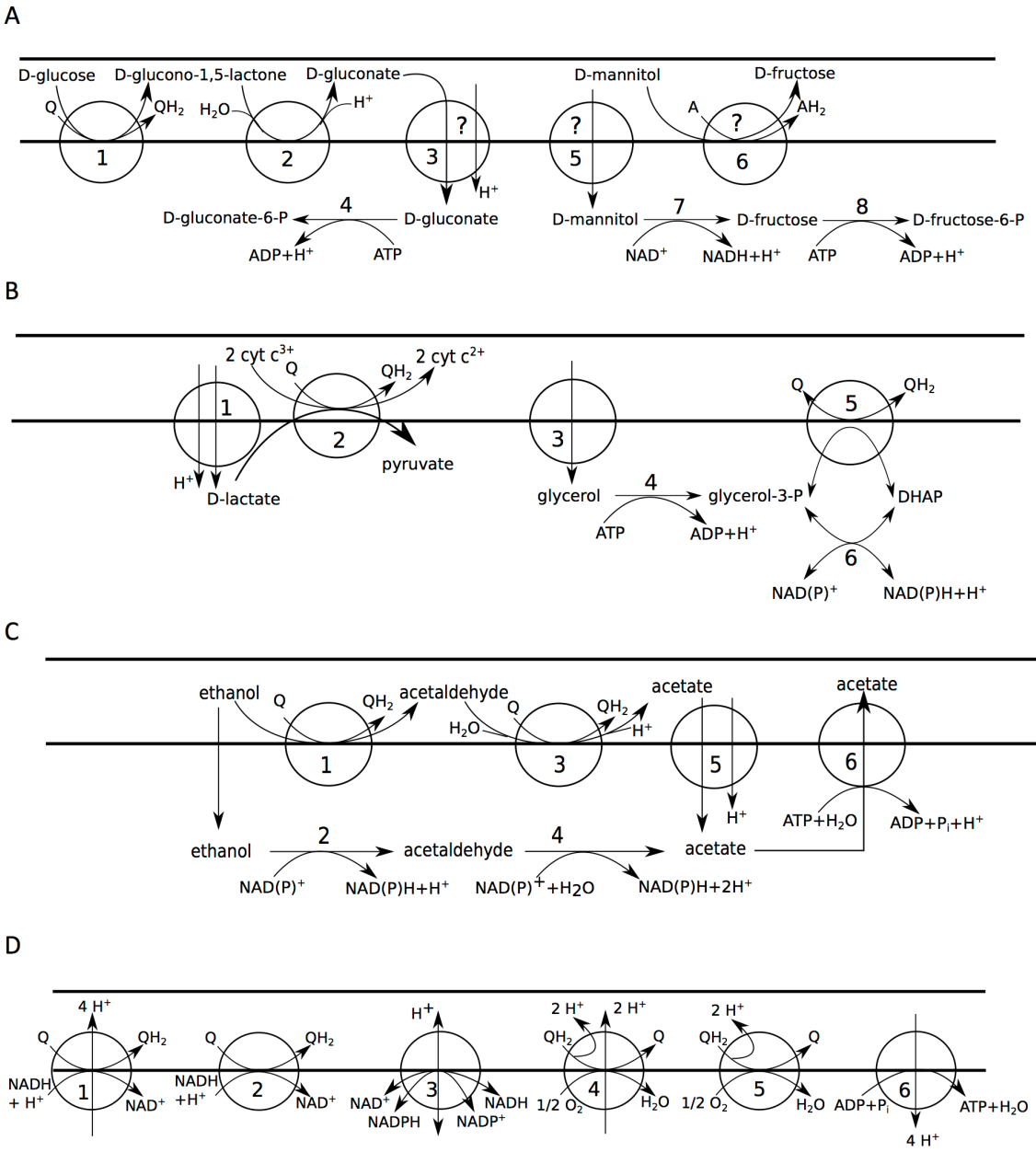


Figure 2. Carbon metabolism of *Acetobacter pasteurianus* 386B. Horizontal bars (from top to bottom) delineate the extracellular environment, the periplasm, and the cytosol, respectively. **(A)** Carbohydrate oxidation. (1) D-glucose:ubiquinone oxidoreductase (EC 1.1.5.2); (2) gluconolactonase (EC 3.1.1.17); (3) gluconate:H^+ symport; (4) gluconokinase (EC 2.7.1.12); (5) mannitol permease; (6) D-sorbitol dehydrogenase (mann_ox_rxn); (7) mannitol 2-dehydrogenase (EC 1.1.1.67) ; (8) fructokinase (EC 2.7.1.4). (**B**) Lactate and glycerol oxidation. (1) lactate:H^+ symporter; (2) D-lactate dehydrogenase (quinone) (EC 1.1.5.12), D-lactate dehydrogenase (cytochrome) (EC 1.1.2.4); (3) glycerol facilitator; (4) glycerol kinase (EC 2.7.1.30); (5) glycerol-3-phosphate dehydrogenase (EC 1.1.5.3); (6) glycerol-3-phosphate dehydrogenase [NAD(P)$^+$] (EC 1.1.1.94). (**C**) Ethanol oxidation. (1) alcohol dehydrogenase (quinone) (EC 1.1.5.5); (2) alcohol dehydrogenase [NAD(P)$^+$] (EC 1.1.1.1, 1.1.1.2); (3) aldehyde dehydrogenase (quinone) (EC 1.2.5.2); (4) aldehyde dehydrogenase [NAD(P)$^+$] (EC 1.2.1.3, EC 1.2.1.4); (5) succinate-acetate:H^+ symporter; (6) acetate ABC transporter. (**D**) Aerobic respiration. (1) NADH:ubiquinone reductase (H^+-translocating) (EC 7.1.1.2); (2) NADH:ubiquinone reductase (non-electrogenic) (EC 1.6.5.9); (3) proton-translocating NAD(P)$^+$ transhydrogenase (EC 7.1.1.1); (4) ubiquinol oxidase (H^+-transporting) (EC 7.1.1.3); (5) ubiquinol oxidase (electrogenic, non-H^+-transporting) (EC 7.1.1.7); (6) H^+-transporting two-sector ATPase (EC 7.1.2.2).

Additionally, *A. pasteurianus* 386B contained genes encoding two cytochrome *bd*-type oxidases, being cytochrome *bd* oxidase (APA386B_472 - APA386B_473) and its homolog, cyanide insensitive oxidase (CIO; APA386B_1010 - APA386B_1111). Since the reaction catalysed by both of these terminal oxidases have the same stoichiometry, an identical reaction (EC 7.1.1.7) was added to the GEM. However, it has been shown that these enzymes have different kinetic parameters, indicating a physiological distinct role (Miura *et al*., 2013). Furthermore, two sets of genes for a cytochrome bc_1 complex (E.C. 7.1.1.8) were found in the genome of *A. pasteurianus* 386B, as is also the case for *A. aceti* (Sakurai *et al*., 2011).

Next to ubiquinone, $NAD(P)^+$ is involved in respiratory chain reactions. Two types of NADH:ubiquinone reductases were encoded in the *A. pasteurianus* 386B genome, namely one that is proton-translocating (EC 7.1.1.2) and one that is not (EC 1.6.5.9). Furthermore, a membrane-bound proton-translocating $NAD(P)^+$ transhydrogenase (EC 7.1.1.1) was present in the genome. Finally, the genes encoding ATP synthase (EC 7.1.2.2) were present in two gene clusters (APA386B_1266 - APA386B_1270 and APA386B_1604 - APA386B_1608), which is also the case for *Rhodospirillum rubrum* (Falk and Walker, 1988).

3.4 *Acetobacter pasteurianus* 386B grows in a defined medium

From all carbon sources tested during the *in vitro* growth experiments, *A. pasteurianus* 386B was only able to grow on a defined medium containing lactic acid as the sole carbon source added. Thus, except for the possible need for some micronutrients, *A. pasteurianus* 386B had no specific auxotrophies and was able to form all its biomass compounds from ammonium as the sole nitrogen source, sulfate as the sole sulfur source, phosphate as the sole phosphate source, and lactic acid as the sole carbon source.

3.5 Biosynthesis pathways

In what follows, biosynthesis pathways are described that were included in the *A. pasteurianus* 386B GEM and for which there was genetic evidence based on the genome re-annotation. These pathways allowed to simulate the growth of *A. pasteurianus* 386B *in silico* using the defined medium conditions mentioned above. Focus is on those steps of the pathways for which there was evidence that the gene-protein-reaction associations in *A. pasteurianus* 386B were different from the reference pathways described in the literature.

3.5.1 Fatty acid and phospholipid biosynthesis

The major fatty acids in bacterial cells of the genus *Acetobacter* are *cis*-vaccenic acid (C18:1), palmitic acid (C16:0), and stearic acid (C18:0) (Yamada *et al*., 1981). Further, the presence of a detectable amount of myristic acid (C14:0) distinguishes the genus *Acetobacter* from the genus *Gluconobacter* (Yamada *et al*., 1981). The fatty acid biosynthesis pathway for saturated fatty acids in *A. pasteurianus* 386B was similar to the one described in *E. coli* (Janßen and Steinbüchel, 2014). For unsaturated fatty acid biosynthesis, the FabA (EC 4.2.1.59/5.3.3.14) and FabB (EC 2.3.1.41) enzymes are critical for their formation in *E. coli* (Feng and Cronan,

2009). Although *cis*-vaccenic acid (C18:1) was the major fatty acid in the fatty acid profile of *A. pasteurianus*, the genes encoding FabA and FabB could not be found in the *A. pasteurianus* 386B genome. Even though unsaturated fatty acid biosynthesis in *A. pasteurianus* 386B seemed to be unclear, reactions of the FabA/FabB pathway were nonetheless added to the model (Supplementary Figure S5). As each fatty acid elongation cycle can be described by a stoichiometric reaction (Janßen and Steinbüchel, 2014), lumped reactions were added to the GEM to describe the formation of myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0) and *cis*-vaccenic acid (C18:1).

Bacterial cell membranes are composed of amphiphilic lipids, most commonly glycerophospholipids (Sohlenkamp and Geiger, 2016). Here, the phospholipid biosynthesis pathway of *Sinorhizobium meliloti* was taken as a reference (Geiger *et al*., 2013). In *A. pasteurianus* 386B, glycerol 3-phosphate could be produced by a quinone-dependent (EC 1.1.5.3) or $NAD(P)^+$-dependent (EC 1.1.1.94) glycerol-3-phosphate dehydrogenase. The *A. pasteurianus* 386B genome encoded the PlsX/PlsY/PlsC system for phosphatidic acid biosynthesis, as described before (Parsons and Rock, 2013). Although long-chain acyl-acyl carrier protein (acyl-ACP) are the end-products of fatty acid biosynthesis and are initially transferred to a phosphate moiety by phosphate:acyl-ACP acyltransferase (PlsX), no generic reaction was available in MNXRef 3.0. Therefore, two reactions (EC 3.6.1.7 and EC 3.1.2.14) were added to the GEM to simulate acyl phosphate and acyl-ACP formation to represent the link between fatty acid biosynthesis and glycerophospholipid biosynthesis. Subsequent acylation of glycerol 3-phosphate is performed by the membrane-bound glycerol-3-phosphate acyltransferase (PlsY; EC 2.3.1.n3). Then, 1-acyl-glycerol-3-phosphate acyltransferase (PlsC; EC 2.3.1.51) forms phosphatidic acid (Hara *et al*., 2008; Paoletti *et al*., 2007; Yoshimura et al., 2007). *Acetobacter aceti* contains an unusually high amount of phosphatidylcholine in its membrane, attributed to its acetic acid resistance (Hanada *et al*., 2001). In *A. pasteurianus* 386B, the three methylation steps of the S-adenosyl-L-methionine (SAM)-dependent methylation pathway forming phosphatidylcholine from phosphatidylethanolamine were most probably catalyzed by the same enzyme, namely phosphatidylethanolamine N-methyltransferase (Pmt, EC 2.1.1.17), as the enzyme from *A. pasteurianus* 386B (encoded by the gene with locus tag APA386B_612) belonged to the same family as the one in *Rhodobacter sphaeroides* and shared high sequence identity to the Pmt enzyme in *A. aceti* (Geiger *et al*., 2013; Hanada *et al*., 2001).

3.5.2 Amino acid biosynthesis

The presence of biosynthesis pathways for proteinogenic amino acids in the *A. pasteurianus* 386B PGDB was confirmed by the growth of *A. pasteurianus* 386B in a defined medium containing ammonium as the sole nitrogen source (Section 3.4). Two genes were retrieved that encoded potential ammonium ion channels (TC 1.A.11; APA386B_239 or APA386B_740). In addition, an ammonium assimilation pathway for *de novo* biosynthesis of proteinogenic amino acids was assumed to be present. In *E. coli*, two pathways are known for ammonium assimilation (van Heeswijk *et al*., 2013), one occurs via a $NADP^+$-dependent glutamate dehydrogenase (EC 1.4.1.4) that allows to form L-glutamate directly from 2-oxoglutarate and

ammonium, the other via the formation of L-glutamate and L-glutamine through a cycle of reactions involving glutamine synthetase (EC 6.3.1.2) and glutamate synthase (EC 1.4.1.13), which produces a net amount of L-glutamate. Only the latter pathway was found in the *A. pasteurianus* 386B genome (Figure 3A and Supplementary Figure S6).

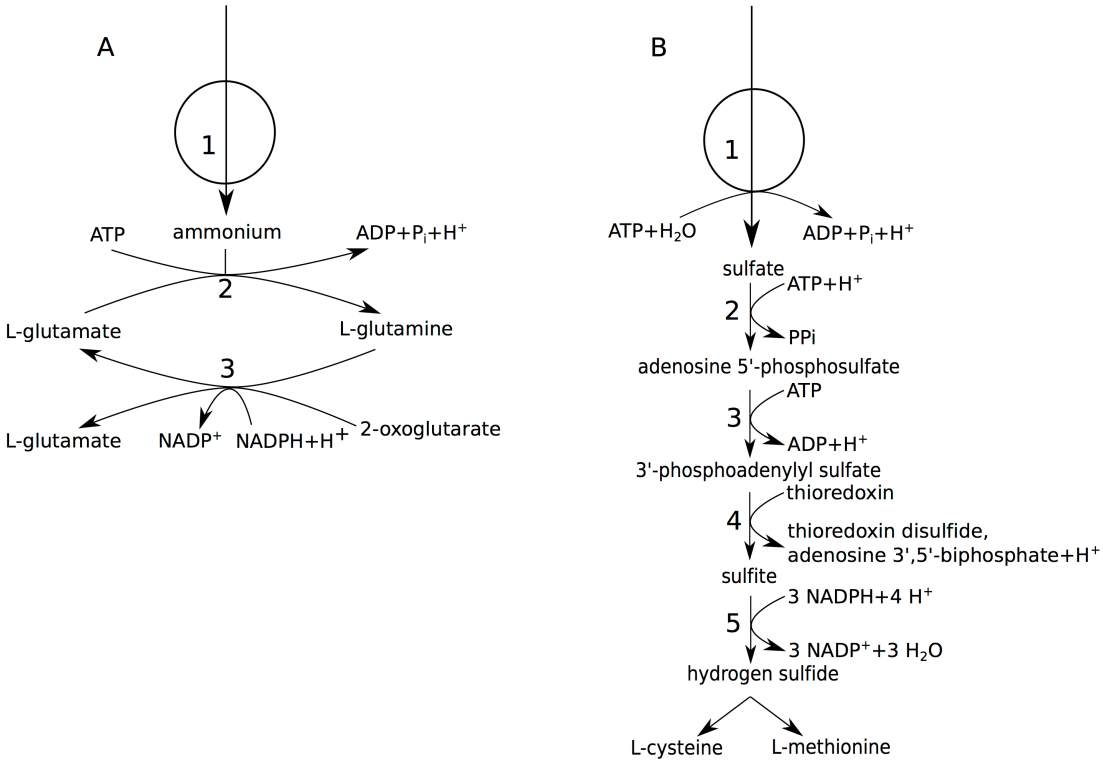


Figure 3. Nitrogen and sulfur metabolism of *Acetobacter pasteurianus* 386B. **(A)** Ammonium assimilation. (1) ammonium permease; (2) glutamine synthetase (EC 6.3.1.2); (3) glutamate synthase (NADPH-dependent) (EC 1.4.1.13). **(B)** Sulfate assimilation. (1) sulfate ABC transporter; (2) sulfate adenylyltransferase (EC 2.7.7.4); (3) adenylyl-sulfate kinase (EC 2.7.1.25); (4) phosphoadenylyl-sulfate reductase (thioredoxin-dependent) (EC 1.8.4.8); (5) assimilatory sulfite reductase (NADPH-dependent) (EC 1.8.1.2).

A number of pyridoxal phosphate-dependent aminotransferases were identified based on the genome annotation. These were involved in the formation of L-aspartate and L-alanine and were assigned to specific reactions based on their presence in different orthogroups. Aspartate aminotransferase (EC 2.6.1.1) was linked to three genes, likely encoding this enzyme (APA386B_861, APA386B_862, and APA386B_942). These were not predicted to be homologous to the *E. coli* aspartate aminotransferase gene (b0928, *aspC*) but shared considerable sequence identity with curated SwissProt sequences from other bacteria with identical function (> 50 % identity). The gene encoding glutamate-pyruvate aminotransferase (EC 2.6.1.2; APA386B_991) was a predicted ortholog of the gene in *E. coli* (b2379, *alaC*). Finally, the genes encoding aspartate 4-decarboxylase (EC 4.1.1.12; APA386B_482, APA386B_1928) had no *E. coli* homolog but were related to a bifunctional aspartate aminotransferase and aspartate 4-decarboxylase of *Comamonas testosteroni*. However, the

enzyme characterised had only a minor activity as an aminotransferase (Wang and Lee, 2007). Thus, the *A. pasteurianus* 386B genes were only associated to the reaction catalyzed by aspartate 4-decarboxylase.

In the biosynthesis pathways of L-valine, L-leucine, and L-isoleucine, a bifunctional enzyme (EC 2.2.1.6) catalyses the formation of (2S)-2-acetolactate and (S)-2-hydroxy-2-ethyl-3-oxobutanoate, the latter metabolite only involved in L-isoleucine biosynthesis (Barak and Chipman, 2012). The *A. pasteurianus* 386B genome contained three genes, which likely encode the enzymes for these reactions. The genes with locus tags APA386B_835 and APA386B_836 were most likely involved in amino acid biosynthesis, since these genes each had an *E. coli* ortholog encoding subunits of a bifunctional acetolactate synthase/acetohydroxybutanoate synthase enzyme complex (*ilvI, ilvH*). However, the gene with locus tag APA386B_1863 was closely related to two 'catabolic' acetolactate synthases, identified in *Klebsiella pneumoniae* and *Lactococcus lactis* (Peng *et al*., 1992; Snoep *et al*., 1992). This evidence, combined with the fact that the neighbouring gene APA386B_1862 was annotated as encoding an acetolactate decarboxylase, forming acetoin from (2S)-2-acetolactate, could be an indication that the physiological role of the APA386B_1863-encoded enzyme would only be related to acetoin formation. Finally, the branched-chain amino acid aminotransferase (EC 2.6.1.42/2.6.1.6; APA386B_1001) could be involved in the last step of the formation of these amino acids.

The aromatic amino acids L-phenylalanine, L-tyrosine, and L-tryptophan share a common initial pathway, which produces chorismate from erythrose 4-phosphate (Yang and Pittard, 2008). Comparative genomics revealed that the gene with locus tag APA386B_1330, encoding a dehydroquinate dehydratase (EC 4.2.1.10), had an ortholog in *G. oxydans* 621H (locus tag GOX0437) that has been identified as encoding a periplasmic enzyme involved in the oxidation of quinate (Adachi *et al*., 2008). The cytoplasmic dehydroquinate dehydratase of *G. oxydans* 621H (locus tag GOX1351) could not be found in the *A. pasteurianus* 386B genome. In the last reaction step of the biosynthesis of L-tyrosine and L-phenylalanine, an amino group transfer occurs with glutamate as amino group donor. In *E. coli*, three enzymes have been identified that could catalyse this reaction, namely the aromatic aminotransferase (TyrB), the aspartate aminotransferase (AspC), and the branched-chain amino acid aminotransferase (IlvE), whereby the latter is only involved in the biosynthesis of L-phenylalanine (Yang and Pittard, 2008). In *A. pasteurianus* 386B, only an ortholog of the gene encoding IlvE was found (APA386B_1001). Thus, the L-tyrosine-forming reaction was tentatively associated to the aspartate aminotransferases encoded in the genome (APA386B_861, APA386B_862, and APA386B_942).

Since *A. pasteurianus* 386B was able to grow in a defined medium containing sulfate as the sole sulfur source, sulfate assimilation was assumed (Figure 3B). Indeed, import of sulfate ions could occur via a sulfate permease (TC 2.A.53) as well as a probable ortholog of the *E. coli cysPUWA* ABC sulfate transporter. However, since the stoichiometry of the anion:H^+ symporter is not known in *Acetobacter*, only the ABC transporter was added to the *A. pasteurianus* 386B GEM. The enzymes of the assimilatory sulfate reduction pathway, which

reduce sulfate to hydrogen sulfide, were encoded in the genome. One exception was the flavoprotein subunit of sulfite reductase (EC 1.8.1.2), which was not found in the genome re-annotation. Whereas L-methionine is produced by the transsulfuration pathway in *E. coli*, only cystathionine *β*-lyase (EC 4.4.1.8), but not homoserine *O*-succinyltransferase (EC 2.3.1.46), could be identified in *A. pasteurianus* 386B. In *Corynebacterium glutamicum*, next to the transsulfuration pathway, a direct sulfhydrylation pathway is present that produces L-methionine, of which the key enzyme is *O*-acetylhomoserine sulfhydrylase (EC 2.5.1.49) (Hwang et al., 2002). The gene with locus tag APA386B_753 was annotated as encoding a putative bifunctional *O*-acetylhomoserine sulfhydrylase (EC 2.5.1.49)/*O*-succinylhomoserine sulfhydrylase (EC 2.5.1.48). However, the presence of a homoserine *O*-acetyltransferase (E.C. 2.3.1.31; APA386B_2138) upstream in the direct sulfhydrylation pathway gave additional evidence to annotate the gene APA386B_753 as encoding an *O*-acetylhomoserine sulfhydrylase (E.C. 2.5.1.49). Since this pathway was complete, the direct sulfhydrylation pathway was added to the *A. pasteurianus* 386B GEM instead of the transsulfuration pathway. Finally, L-methionine could be synthesised from L-homocysteine by methionine synthase (EC 2.1.1.13). SAM could be formed from L-methionine by methionine adenosyltransferase (EC 2.5.1.6). This reaction is part of the SAM cycle, which links L-methionine biosynthesis to glycerophospholipid biosynthesis, and involves the key reaction catalysed by adenosylhomocysteinase (EC 3.3.1.1) (Reddy *et al*., 2008).

3.5.3 Nucleotide biosynthesis

Based on the *A. pasteurianus* 386B PGDB, the nucleotide biosynthesis pathways were curated and added to the *A. pasteurianus* 386B GEM (Supplementary Figure S7). A broad-substrate-range nucleoside diphosphate kinase (EC 2.7.4.6) was present in the genome of *A. pasteurianus* 386B that could phosphorylate different nucleoside diphosphate acceptors, using ATP as phosphate donor, to form the respective nucleoside triphosphate products (Armenta-Medina *et al*., 2014). Similarly, a broad-substrate-range thioredoxin-dependent ribonucleoside-diphosphate reductase (EC 1.17.4.1) could reduce ribonucleoside diphosphate acceptors to their respective deoxyribonucleoside diphosphate forms. The thioredoxin molecule may be recycled by an $NADP^+$-dependent thioredoxin reductase (EC 1.8.1.9). The precursor for purine biosynthesis is 5-phospho-α-D-ribose 1-phosphate (PRPP), which is also the starting point of histidine biosynthesis and involved in tryptophan biosynthesis (Kilstrup *et al*., 2005). From PRPP, a linear pathway was obtained to produce inosine monophosphate, from which adenosine monophosphate (AMP) or guanosine monophosphate (GMP) could be produced. Two additional adenosine salvage reactions were identified, which were related to side-products formed in other pathways. First, sulfate reduction to hydrogen sulfide could yield adenosine 3',5'-bisphosphate as a by-product of the initial adenylation of sulfate by ATP. This compound could be recycled to AMP by 3',5'-bisphosphate nucleotidase (EC 3.1.3.7), which was previously not annotated in the *A. pasteurianus* 386B genome. Second, the SAM cycle could produce adenosine, which could be recycled to AMP by adenosine kinase (EC 2.7.1.20).

Uridine diphosphate (UDP) production is the result of a linear pathway, in which aspartate is the substrate for pyrimidine biosynthesis and which could be converted into cytidine

triphosphate (CTP) via CTP synthase (EC 6.3.4.2) or deoxythymidine triphosphate (dTTP) by a dedicated pathway (Kilstrup *et al*., 2005). Cytidine monophosphate (CMP) could be salvaged by cytidylate kinase (EC 2.7.4.25), as CMP was a side-product from phospholipid and lipopolysaccharide (LPS) biosynthesis. Consecutive cytidine diphosphate (CDP) reduction by ribonucleoside-diphosphate reductase and phosphorylation by nucleoside diphosphate kinase could produce dCTP.

3.5.4 Peptidoglycan and lipopolysaccharide biosynthesis

Biosynthesis pathways of membrane components in *E. coli* were used as a reference for their reconstruction in *A. pasteurianus* 386B (Supplementary Figures S8 and S9). In general, the biosynthesis of peptidoglycan starts with the formation of glucosamine 6-phosphate by glutamine-fructose-6-phosphate transaminase (EC 2.6.1.16), which is subsequently biochemically activated to form UDP-*N*-acetylglucosamine (UDP-GlcNAc). UDP-GlcNAc is subsequently converted into UDP-*N*-acetylmuramic acid (UDP-MurNAc) and used as a basis to attach the first alanine residue by UDP-*N*-acetylmuramate L-alanine ligase (EC 6.3.2.8), which was previously not annotated in the *A. pasteurianus* 386B genome. In the next reactions, D-glutamate, meso-diaminopimelate, and D-alanyl-D-alanine are successively attached to UDP-MurNAc (Barreteau *et al*., 2008). Finally, another molecule of UDP-GlcNAc is linked to UDP-MurNAc to form the peptidoglycan monomer (Vollmer *et al*., 2008). This last step was included in the macromolecule reaction forming peptidoglycan (Supplementary Figure S1). Whereas it involves *in vivo* the linkage of the peptidoglycan precursor to undecaprenyl-phosphate (Typas et al., 2012), a growing peptidoglycan chain requires the recycling of undecaprenyl-phosphate, so its biosynthesis was not included in the model.

For lipopolysaccharide biosynthesis, UDP-GlcNAc is an important precursor to form $(Kdo)_2$-lipid A via the Raetz pathway (Whitfield and Trent, 2014). In *A. pasteurianus* 386B, UDP-2,3-diacylglucosamine pyrophosphatase (EC 3.6.1.54) could form the intermediary lipid X. In contrast to *E. coli* that has the *lpxH* gene, in *A. pasteurianus* 386B and other α-proteobacteria this enzyme is encoded by the *lpxI* gene, which has a different reaction mechanism (Metzger and Raetz, 2010). Lipid X could be converted into lipid IV and two molecules of CMP-ketodeoxyoctonate (CMP-Kdo) could be transferred to lipid IV by a bifunctional 3-deoxy-D-manno-octulosonic-acid transferase (EC 2.4.99.12 and EC 2.4.99.13). Finally, two fatty acyl chains could be added to form $(Kdo)_2$-lipid A. Whereas *E. coli* has two distinct enzymes encoded by *lpxL* and *lpxM*, one for each fatty acyl chain transfer, of which *lpxM* is not required for growth (Raetz *et al*., 2007), *A. pasteurianus* 386B possessed only one putative ortholog, APA386B_2689, which was most similar to *lpxL*. Subsequently, a core oligosaccharide unit, of which the composition may differ between microorganisms, is synthesized and attached to lipid A (Raetz and Whitfield, 2002). Here, the pathway present in *A. pasteurianus* 386B diverged from the one known in *E. coli*, since only four out of ten described enzymatic reactions to synthesize the *E. coli* core oligosaccharide could be linked to putative *A. pasteurianus* 386B orthologs. Furthermore, two enzymes were missing in the pathway to produce ADP-L-glycero-D-manno-heptose. These reactions were nonetheless added to the model to allow for biomass formation (Supplementary Figure S9). Genes encoding enzymes

to form the LPS precursors UDP-glucose and UDP-galactose were present, including the genes encoding phosphoglucomutase (EC 5.4.2.2), UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9), and UDP-glucose 4-epimerase (EC 5.1.3.2). Biosynthesis reactions for the formation of lipoprotein were not added to the *A. pasteurianus* 386B GEM, since there was no information about its abundance in the biomass composition data used.

3.6 Model validation

In vitro and *in silico* growth experiments were compared to validate the *A. pasteurianus* 386B GEM (Table 1). This strain has been routinely cultivated in a cocoa pulp simulation medium containing lactic acid, ethanol, and mannitol as the main carbon sources (Lefeber *et al*., 2010; Moens *et al*., 2014). Here, the carbon sources were tested separately to assess their influence on the predicted flux distribution in isolation. For all tested carbon sources, periplasmic proton exchange was necessary to obtain growth *in silico*.

Table 1. *In vitro* growth experiments and *in silico* predicted specific growth rate using different carbon sources.

Carbon source	*In vitro* growth	*In silico* specific growth rate (h^{-1})
D-glucose	No	0.66
D-mannitol	No	0.66
Glycerol	No	1.12
Lactic acid	Yes	0.85
Ethanol	No	0.00
Acetic acid	No	0.00

Growth of *A. pasteurianus* 386B on lactic acid as the sole carbon source was found *in silico* as well as *in vitro*. CO_2 and H_2O were the sole metabolites secreted by the model, which was in accordance to the experimental results obtained with another *A. pasteurianus* strain (Adler *et al*., 2014). *Acetobacter pasteurianus* 386B was not able to grow on ethanol as the sole carbon source, which confirms the growth characteristics of *A. pasteurianus* and is in contrast to other species of the genus *Acetobacter* (Cleenwerck *et al*., 2008). This result was also obtained when acetate was the sole carbon source. These results are probably related to the absence of genes encoding enzymes of the glyoxylate cycle in the *A. pasteurianus* 386B genome, since this cycle is known to be crucial for growth on C2 sources such as ethanol and acetate (Illeghems *et al*., 2013).

In contrast to the results of the *in vitro* growth experiments, *in silico* growth was possible on D-glucose and D-mannitol as the sole carbon sources. In both cases, FBA predicted their catabolism by the pentose phosphate pathway, leading to the formation of fructose-6-phosphate

and glyceraldehyde-3-phosphate. The latter was further catabolized in the lower part of the Embden–Meyerhof–Parnas (EMP) pathway. Due to the absence of a phosphofructokinase enzyme, fructose-6-phosphate was converted into glucose-6-phosphate, leading to a reaction flux cycle involving the pentose phosphate pathway and the upper part of the EMP pathway. A high $NADPH+H^+$ production flux had to be balanced to allow growth, which was mainly performed by a proton-translocating $NAD(P)^+$ transhydrogenase. The contrasting results obtained for D-glucose and D-mannitol might be explained by the fact that no specific transporter could be identified for these metabolites or their oxidation products. In general, functional annotation of transporters is difficult because a limited number of transporters have been functionally characterised (Gelfand and Rodionov, 2008). It is therefore possible that D-glucose and D-mannitol are almost exclusively oxidized in the periplasm, forming D-gluconate and D-fructose, respectively (Moens *et al*., 2014). In addition, the occurrence of a reaction flux cycle in the FBA prediction could be an indication of the incapability of sugar catabolism to sustain growth of *A. pasteurianus* 386B.

A similar discrepancy for *in vitro* and *in silico* growth was found for glycerol. The predicted specific growth rate on glycerol was the highest of all carbon sources tested, but *in silico* growth was only possible if the reaction catalyzed by the aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3) was defined as being reversible. Here, a glycerol facilitator could be identified in the genome of *A. pasteurianus* 386B. No apparent reason for the absence of *in vitro* growth could be identified in the flux distribution. However, the assumption of the presence of a catabolic glycerol-3-phosphate dehydrogenase may be faulty or glycerol catabolism was not efficient enough for the bacterial cells to grow on glycerol as the sole carbon source.

3.7 Evaluation of the manual curation process and *A. pasteurianus* 386B GEM properties

Manual curation of GPR associations in the *A. pasteurianus* 386B GEM was performed based on different information sources. Hereto, EC numbers were used, as these provide a defined classification of enzymatic reactions. However, GPR associations may be complex and it is thus expected that curation will reveal inconsistencies in the functional annotations provided by the different annotation sources. For genes linked to reactions in the GEM, the differences between the manually curated EC numbers and the EC numbers provided by each annotation source were compared (Figure 4). Some annotation sources, such as eggNOG-mapper and NCBI 2017, did not provide much EC information. EggNOG-mapper annotations are characterised by long annotation notes, mostly descriptive and evading EC classification. For the NCBI annotation, the outcome is related to the RefSeq annotation version used that apparently only contained very few EC numbers. Although overall the EC numbers in the *A. pasteurianus* 386B GEM corresponded to the ones in the different annotation sources, differences were apparent at the main class or sub-subclass EC classification. The former indicated, for example, the occurrence of multi-functional enzymes, linked to more than one reaction, with EC numbers belonging to different classes. Also, proton-translocating enzymes were recently re-classified in a new EC class of translocases (EC 7). The latter highlighted the subtleties of the manual curation process in defining the substrate and co-factor specificities.

The properties of the resulting *A. pasteurianus* 386B GEM, the curated *A. pasteurianus* 386B PGDB, and the models of the automatic reconstruction tools are summarised in Table 2. Compartmentalisation differed between the different reconstructions, probably due to a difference in ways to represent a GEM from Gram-negative bacteria. However, since compartmentalisation represents important physiological barriers, it most likely influences the GEM simulation results. As could be expected, the number of genes in the automatically reconstructed models was higher than for the *A. pasteurianus* 386B GEM presented in the current study, as this reconstruction was performed manually and because some pathways were deliberately not included, such as the co-factor biosynthesis pathways. This latter decision was taken based on the assumption that the biosynthesis and degradation fluxes of the reaction co-factors were small compared to their involvement in the metabolic redox reactions. In contrast, the number of orphan reactions was much lower for the *A. pasteurianus* 386B GEM compared to the other models, providing causal links between reactions and enzyme-encoding genes in the genome. Also, the number of dead-end metabolites was low compared to the automatic reconstructions. Although dead-end metabolites could be an indication of redundancy in the reconstruction, their presence may also reflect the uncertainty in the reactants and products of the different enzymatic reactions in the models, thus providing a compendium of possible reactants and products, as for example in the curated *A. pasteurianus* 386B PGDB.

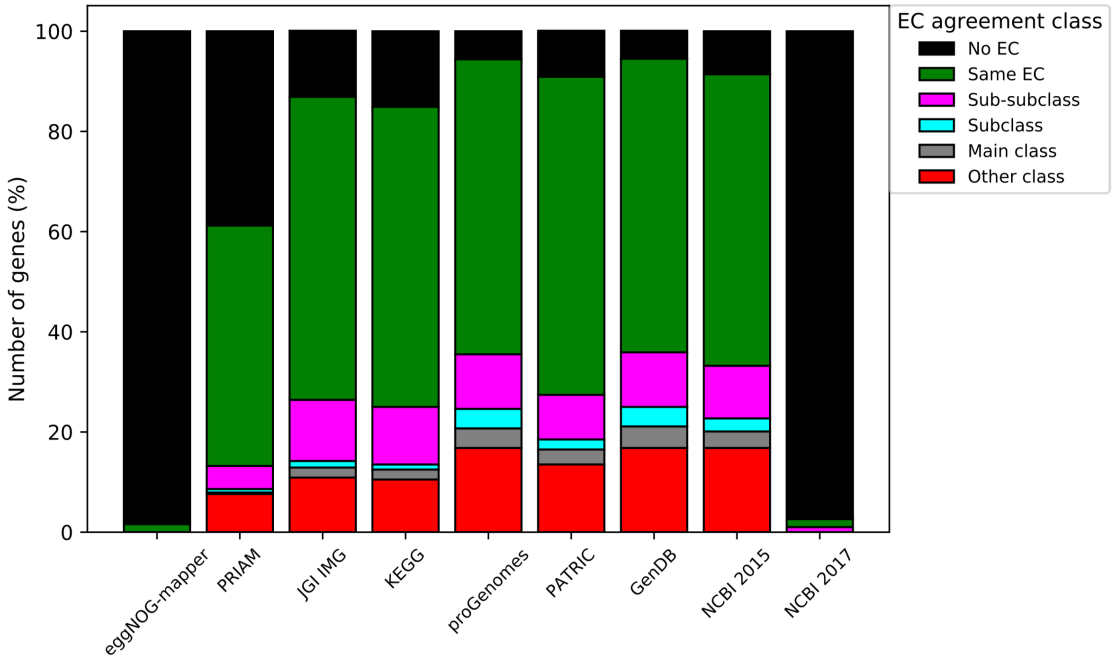


Figure 4. Comparison of EC number annotations of genes in the *Acetobacter pasteurianus* 386B GEM with EC numbers present in other annotation sources. The levels of agreement between annotated EC numbers were classified in a sub-subclass (three first numbers identical), subclass (two first numbers identical), main class (first number identical), and other class (different first number), according to the EC classification system. The number of genes classified in each class is shown for each of the annotation sources considered.

Table 2. Properties of currently available reconstructed genome-scale metabolic models of *Acetobacter pasteurianus* 386B. Curated reconstructions include iAp386B454 and PGDB_curated. Draft reconstructions include CarveMe, KBase, MNX, and PGDB (Pathway Tools). Exchange reactions represent a possible exchange of metabolites with the environment. Irreversible reactions are based on the flux bounds in the SBML file. Orphan reactions have no GPR association. Dead-end metabolites have only a single reaction partner. NA, not available.

Property	iAp386B454	CarveMe	KBase	MNX	PGDB	PGDB_curated
Compartments	2	3	2	3	0	0
Pathways	NA	NA	NA	NA	294	213
Genes	454	611	697	820	723	829
Reactions	322	1424	1061	2712	1514	1818
Exchange reactions (% of reactions)	17 (5 %)	135 (9 %)	98 (9 %)	245 (9 %)	0 (0 %)	0 (0 %)
Irreversible reactions (% of reactions)	135 (42 %)	946 (66 %)	477 (45 %)	1451 (54 %)	1052 (69 %)	1493 (82 %)
Orphan reactions (% of reactions)	32 (10 %)	473 (33 %)	107 (10 %)	778 (29 %)	468 (30 %)	308 (16 %)
Metabolites	296	1078	1026	1535	1608	1759
Dead-end metabolites (% of metabolites)	2 (1 %)	19 (2 %)	305 (30 %)	90 (6 %)	701 (43 %)	719 (40 %)

4 Conclusion

During the genome re-annotation of *A. pasteurianus* 386B, the functional annotation of predicted enzymes and transporters was targeted, as these are critical for an accurate genome-scale metabolic network reconstruction. To improve the quality of the re-annotation, information from multiple, different annotation sources was combined, which proved to be a good strategy to guide manual curation of GPR associations. This methodology was combined with the prediction of orthogroups, using genomes of related species, which was further fine-tuned with information from the literature. Finally, using the Pathway Tools software allowed to set the annotation content in a metabolic pathway context. In this way, possible links between the different biosynthesis pathways, necessary for the biomass formation, were identified.

The re-annotated *A. pasteurianus* 386B genome was used to compile a curated GEM, named iAp386B454, containing 454 genes, 322 reactions, and 296 metabolites embedded in two cellular compartments. The GEM is available in SBML level 3 format and as a curated Pathway Tools PGDB and represents the first *in silico* genome-scale metabolic network reconstruction of a species of the genus *Acetobacter*. The reconstructed model was validated by performing growth experiments in a defined medium, which revealed that lactic acid as the sole carbon source could sustain growth of this strain.

Nevertheless, it became clear that some knowledge gaps remained, for example for the reconstruction of the biosynthesis pathways of cell constituents, especially for the cell envelope. Also, the combination of genome re-annotation and growth experiments could not resolve the presence of all metabolite transporters. The results obtained in this study will help to guide future research to close these knowledge gaps in *A. pasteurianus*.

Chapter 4

Application of comparative genomics of *Acetobacter* species facilitates genome-scale metabolic reconstruction of *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

Summary

Genome sequence data of a selection of strains of *Acetobacter* species were used to perform a comparative genomics analysis. Combining the predicted orthologous groups of protein-encoding genes from the *Acetobacter* genomes with gene-reaction rules of genome-scale metabolic models (GEMs) from two reference microorganisms, namely a previously manually curated model of *Acetobacter pasteurianus* 386B (iAp386B454) and two manually curated models of *Escherichia coli* (EcoCyc and iJO1366), allowed to predict the set of reactions present in *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B, two candidate starter culture strains for cocoa bean fermentation processes. The predicted presence of reactions was manually curated using genome re-annotation data, followed by the reconstruction of species-specific GEMs. This approach additionally revealed possible differences concerning the carbon core metabolism and redox metabolism among *Acetobacter* species, pointing to a hitherto unexplored metabolic diversity. Specifically, the presence or absence of reactions related to citrate catabolism and to the glyoxylate cycle for assimilation of C2 compounds provided not only new insights into the cocoa bean fermentation process but also interesting guidelines for future research. The *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B GEMs, reconstructed in a semi-automated way, provided a proof-of-concept toward accelerated formation of GEMs of candidate functional starter cultures for food fermentation processes.

1 Introduction

Acetic acid bacteria (AAB) are Gram-negative, obligately aerobic bacteria that incompletely oxidize a range of alcohols and reducing sugars. The region- and stereo-selectivity of these oxidation reactions has led to numerous biotechnological applications, most notably the conversion of D-sorbitol into L-sorbose (Shinjoh and Toyama, 2016). The genus *Acetobacter* comprises currently more than 20 reported species (Yamada, 2016), and is common to food fermentation processes, among which traditional vinegar production, kombucha fermentation, and the cocoa bean fermentation process (De Vuyst and Weckx, 2016; Coton *et al.*, 2017; Lu *et al.*, 2018). For the latter fermentation process, several *Acetobacter* species have been isolated and selected as potential candidate functional starter cultures, including *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B (Camu *et al.*, 2007; Moens *et al*., 2014). To gain more insight into their contribution to the cocoa bean fermentation process, their genomes have been sequenced and functionally annotated, revealing a possible role in citrate assimilation during the fermentation process (Illeghems *et al.*, 2016). However, *A. ghanensis* and *A. senegalensis* are outcompeted during cocoa bean fermentation, whereas *A. pasteurianus* prevails (Lefeber *et al*., 2011a).

So far, the gold standard in the taxonomic analysis of AAB are polyphasic approaches, combining phenotypic, chemotaxonomic, and genotypic data (Cleenwerck and De Vos, 2008). One specific metabolic feature that is typically tested for differentiating *Acetobacter* species is growth on ethanol as the sole carbon source. Therefore, a defined medium containing ethanol as the sole carbon source and ammonium as the sole nitrogen source is often used (Cleenwerck, 2002). However, carbon source usage may also be deduced through genomic analysis, since carbon assimilation pathways consist of a defined sequence of biochemical reactions for which the respective enzyme-encoding genes must be present in the genome (Serres and Riley, 2006).

Growth of *Acetobacter* species on C2 compounds, such as ethanol and acetate, has so far been related to the presence of the glyoxylate cycle enzymes (Saeki *et al.*, 1997; Sakurai *et al.*, 2013). The glyoxylate cycle is a shunt of the tricarboxylic acid (TCA) cycle and consists of only two enzymes, namely isocitrate lyase (EC 4.1.3.1), which converts isocitrate into glyoxylate and succinate, and malate synthase (EC 2.3.3.9), which fixates acetyl-CoA to glyoxylate forming malate. The glyoxylate cycle circumvents the CO_2-producing reactions of the TCA cycle, thus leading to a net succinate production. The presence of genes encoding the glyoxylate cycle enzymes in *Acetobacter* species genomes and their relationship to the growth of these species in a medium containing ethanol as the sole carbon source has not been thoroughly investigated.

A genome-scale metabolic model (GEM) provides a well-defined data structure of the metabolic network inferred from a bacterial genome (Thiele and Palsson, 2010). Next to the use of a GEM as knowledge base of the biochemical reaction potential of a microbial strain, a GEM is also useful to perform simulations of the metabolism of a microbial strain, for instance, using constraint-based modeling methods. Recently, the use of GEMs to compare the reaction potential in different microorganisms has been introduced, for which the term pan-reactome

was coined (Prigent *et al.*, 2018; Seif *et al.*, 2018; Gu *et al.*, 2019). In this case, the GEM reconstruction process is performed for different strains in parallel using different metabolic resources, for instance, GEMs of reference species and metabolic databases.

In the present study, the concept mentioned above is applied to reconstruct GEMs for *A. senegalensis* 108B and *A. ghanensis* LMG 23848^T, taking advantage of a comparative genomics analysis that encompasses 36 strains belonging to 27 species of the *Acetobacter* genus and a previously reconstructed and manually curated GEM of *A. pasteurianus* 386B (Pelicaen *et al.*, 2019). These GEMs will be useful to gain more insight into the metabolic adaptations of *A. senegalensis* 108B and *A. ghanensis* LMG 23848^T to cocoa fermentation conditions.

2 Materials and methods

2.1 Strains

Three *Acetobacter* strains originating from spontaneous cocoa bean fermentation processes, namely *Acetobacter pasteurianus* 386B, *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B, were used throughout this study (Camu *et al*., 2007; Illeghems *et al*., 2013, 2016; Moens *et al*., 2014).

2.2 Growth experiments

Growth experiments were performed in a modified defined medium (Verduyn *et al.*, 1992). This medium contained (per litre): $(NH_4)_2SO_4$, 5 g; KH_2PO_4, 1.375 g; $MgSO_4.7H_2O$, 0.5 g; EDTA, 15 mg; $ZnSO_4.7H_2O$, 4.5 mg; $CoSO_4.7H_2O$, 0.35 mg; $MnCl_2.4H_2O$, 1 mg; $CuSO_4.5H_2O$, 0.3 mg; $CaCl_2.2H_2O$, 4.5 mg; $FeSO_4.7H_2O$, 3 mg; MoO_3, 0.24 mg; H_3BO_3, 1 mg; and KI, 0.1 mg, dissolved in 30 mM citrate buffer (pH 6.0). The final pH of the medium was set to 5.0. A filter-sterilised vitamin mixture was added after heat sterilisation of the medium (120°C, 20 min). The final vitamin concentrations were (per litre): biotin, 0.0005 mg; calcium pantothenate, 0.01 mg; nicotinic acid, 0.01 mg; *myo*-inositol, 0.25 mg; thiamine-HCl, 0.01 mg; pyridoxine-HCl, 0.01mg; and para-aminobenzoic acid, 0.002 mg. Eight different carbon sources, namely glucose, fructose, mannitol, citric acid, glycerol, lactic acid, ethanol, and sodium acetate, were used at a final concentration of 30 mM to assess if they could sustain growth of the three *Acetobacter* strains mentioned above. The pH of the citric acid, lactic acid, and sodium acetate stock solutions was set to 5.0.

For inoculum build-up, the three *Acetobacter* strains were grown overnight at 30 °C in a standard incubator in an undefined medium (Lefeber *et al.*, 2012), which contained (per litre): D-glucose, 20 g; D-mannitol, 20 g; lactic acid, 10 g; granulated yeast extract, 10 g; soy peptone, 5 g; $MgSO_4.7H_2O$, 1 g; $(NH_4)_2HPO_4$, 1 g; KH_2PO_4, 1 g; and $Na_3C_6H_5O_7$, 1 g. The pH of the medium was set to 5.5. The shaking rate was set to 160 rpm. The overnight culture was centrifuged (4,000 x *g*, 20 min, 4°C), and washed with a filter-sterilized saline solution (0.85 %, m/v, NaCl). Then, the cells were resuspended in sterile saline solution and inoculated in 2

ml of the defined medium mentioned above at an optical density at 600 nm (OD_{600}) of 0.01, in duplicate, in test tubes with a total volume of 20 ml. The three *Acetobacter* strains were allowed to grow at 30°C for 48 h and 120 h without shaking. A threshold value for the OD_{600} of 0.1 was used to identify whether or not the strains had grown.

A similar growth experiment was performed in a modified mineral medium with 30 mM phosphate buffer (pH 6.0), in triplicate, for the three *Acetobacter* strains. In this case, for inoculum build-up, the three *Acetobacter* strains were grown overnight at 30 °C in an undefined medium (pH 5.5), which contained (per litre): lactic acid, 5 g; sodium acetate, 10 g; granulated yeast extract, 5 g; $MgSO_4.7H_2O$, 1 g; $NH_4H_2PO_4$, 20 g; and K_2HPO_4, 10 g (Pelicaen *et al*., 2019). A threshold value for the OD_{600} of 0.1 was used to identify whether or not the strains had grown.

2.3 Re-annotation of the *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B genomes

The protein-encoding genes of the *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B genomes were originally annotated using the bacterial genome annotation system GenDB v2.2 (Meyer, 2003; Illeghems *et al.*, 2016). In the present study, these genomes were re-annotated using a previously in-house developed bioinformatics workflow (Pelicaen *et al*., 2019). The NCBI RefSeq genome annotation version was taken as a basis for re-annotation of *A. ghanensis* LMG 23848^T (accessed in April 2017, 2454 protein-encoding genes) and *A. senegalensis* 108B (accessed in October 2017, 3444 protein-encoding genes). The amino acid sequences of each set of protein-encoding genes were annotated using a combination of tools, namely BlastKOALA from the KEGG database (Kanehisa *et al.*, 2016, 2017), the TransAAP tool to predict transport proteins (Elbourne *et al.*, 2017), the subcellular localization predictor CELLO (Yu *et al.*, 2004), eggNOG-mapper (Huerta-Cepas *et al.*, 2017), the enzyme annotation tool PRIAM (Claudel-Renard, 2003), the RAST annotation pipeline from the KBase software (Overbeek *et al.*, 2014; Arkin *et al.*, 2018), and the tools embedded in InterProScan 5.22-61.0 (Jones *et al.*, 2014). Furthermore, since the publication of the genome sequences of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B (Illeghems *et al.*, 2016), these genomes have been re-annotated by several annotation pipelines and the functional annotation data were stored in dedicated databases related to those pipelines. As such, these additional functional annotations were retrieved from the Carbohydrate-Active enZYmes (CAZy) database (Lombard *et al.*, 2014), the Pathosystems Resource Integration Center (PATRIC) database (Wattam *et al.*, 2017), and the Universal Protein Resource (UniProt) database (The UniProt Consortium, 2019).

In addition, the genome annotation pipelines used by GenDB, NCBI RefSeq (NCBI prokaryotic genome annotation pipeline; Tatusova *et al.*, 2016) and PATRIC (RASTtk; Brettin *et al.*, 2015) represent independent methods combining gene prediction and gene annotation. Therefore, functional annotation data of these sources were combined in an in-house relational database (MySQL) specific for *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B.

2.4 Comparative genomics of *Acetobacter* species based on orthogroups

From all *Acetobacter* species mentioned in the NCBI Taxonomy database (accessed May 2019), the genome sequence of 96 strains of *Acetobacter* species was available in the NCBI RefSeq database. All those genome sequences were retrieved, as well as the genome sequences of *Gluconobacter oxydans* 621H, *Komagataeibacter nataicola* RZS01, and *Escherichia coli* str. K-12 substr. MG1655 (further referred to as *E. coli*). The latter three strains were used as outgroup. Next, the *Acetobacter* genomes were ranked according to their quality, based on the NCBI RefSeq database category and assembly level information of the genomes. For each *Acetobacter* species, its highest ranked genome was selected, and if this genome was not from the type strain, the type strain genome was also selected. Next, OrthoFinder (Emms and Kelly, 2015) was used to predict orthogroups from the protein-encoding genes of the selected *Acetobacter* genomes and from the three outgroup genomes. Protein identifiers from the NCBI RefSeq database were linked to gene identifiers based on the Genbank files of each genome. Subsequently, OrthoFinder was used to infer a rooted species tree based on the predicted orthogroups (Emms and Kelly, 2017, 2018).

2.5 Reconstruction of genome-scale metabolic models for *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B

The relationship between genes, proteins, and reactions can be described using gene-protein-reaction (GPR) associations, linking the different entities, possibly even depicting their relative stoichiometry (Thiele and Palsson, 2010). However, in practice, GPR associations are typically implemented as Boolean rules that define reaction presence based on gene presence in an annotated genome (Machado *et al.*, 2016). Albeit potentially oversimplifying the actual GPR associations, these Boolean rules (gene-reaction rules) allow to quickly evaluate the presence of a reaction in the GEM as a function of *in silico* gene deletions (Ebrahim *et al.*, 2013). Here, this concept was exploited to predict the reaction presence based on gene presence in the annotated genomes. To convert the text Boolean expression into a symbolic rule that allows its computational assessment, the Python library SymPy was used.

As a first step in the reconstruction of the GEMs for *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B, the manually curated GEM of *A. pasteurianus* 386B (iAp386B454; Pelicaen *et al*., 2019) was used as a reference to evaluate which of the reactions of this GEM were present in the *Acetobacter* genomes considered. To minimize the number of false negative predictions of reaction presence, a separate analysis was performed, whereby only high-quality *Acetobacter* genomes with an NCBI assembly level corresponding to a complete genome or chromosome level were considered.

Next, the predictions obtained were used for the reconstruction of GEMs of *A. ghanensis* LMG 23848^T, further referred to as iAg23848, and of *A. senegalensis* 108B, further referred to as iAs108B. New gene-reaction rules were associated to the transferred iAp386B454 reactions. For multi-copy orthogroups, *i.e.*, orthogroups containing multiple genes for each genome, the logic OR assumption was made, expressing that these genes are co-orthologs of the *A.*

pasteurianus 386B genes. Cases for which *A. pasteurianus* 386B genes occurred in the same orthogroup as the ones of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B and had a logic AND relation in the iAp386B454 gene-reaction rules, were systematically checked and manually curated in the iAg23848 and iAs108B GEMs. Reactions without GPR, including spontaneous reactions as well as reactions for macromolecular biosynthesis and the iAp386B454 biomass reaction, were transferred as such to the iAg23848 and iAs108B GEMs.

2.6 Gap filling of genome-scale metabolic models and flux balance analysis

Further manual curation was performed for the iAg23848 and iAs108B GEMs in a semi-automated way. First, GEM gap filling was performed using iAp386B454 as a template. Hereto, the GEM gap filling method of COBRApy was used (Ebrahim *et al.*, 2013), which computes the minimal amount of reactions that need to be added from the iAp386B454 template to obtain *in silico* growth. Then, gap filling results were manually curated by combining functional annotation information of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B, stored in their respective relational databases and in the GEMs. Only the reactions for which genome annotation evidence was found were added to the GEMs. This also included reactions necessary to reconcile *in silico* and *in vitro* growth experiments. Finally, flux balance analysis (FBA) and parsimonious FBA with biomass formation as the objective function were performed to examine the flux distributions of the newly constructed GEMs under simulated *in vitro* growth conditions. The total allowable carbon consumption flux was set to 60 C-mmol/g_{CDW}/h.

2.7 Presence of the glyoxylate cycle and aerobic respiratory chain reactions

To predict the presence/absence of reactions of the glyoxylate cycle and the aerobic respiratory chain, a similar analysis as described in Section 3.4 was performed but now with an *E. coli* GEM as reference. For *E. coli*, two curated reconstruction sources were leveraged, namely the EcoCyc Pathway/Genome Database embedded in Pathway Tools version 21.5, and the iJO1366 GEM of the BIGG database (Orth *et al.*, 2011; Keseler *et al.*, 2017).

3 Results

3.1 Growth experiments

Growth experiments (Table 1) under defined medium conditions, using a citrate buffer, indicated that *A. pasteurianus* 386B showed the same growth phenotype as in defined medium with a phosphate buffer (Pelicaen *et al.*, 2019). Only the addition of lactic acid as a sole carbon source could sustain the growth of *A. pasteurianus* 386B under these conditions. For *A. ghanensis* LMG 23848^T, increasing the incubation time from 48 h to 120 h allowed for growth of the bacterial cell population only on glucose as the sole carbon source, using a citrate buffer. In contrast, the growth phenotype of *A. senegalensis* 108B markedly differed when using a phosphate buffer or citrate buffer. Under growth conditions with a phosphate buffer, this strain only showed growth on lactate and citrate as the sole carbon sources, but exchanging the

phosphate buffer for a citrate buffer led to growth on all single carbon sources tested, at least after 120 h of incubation. The results indicated the influence of the medium buffer used, and the possibility of citrate co-consumption in the case that a citrate buffer was applied.

Table 1. Indication of bacterial cell population growth of strains of three different *Acetobacter* species in a defined medium supplemented with different carbon sources, denoted as a binary code: 0, no growth; 1, growth. A, modified mineral medium with phosphate buffer, after 48 h of incubation; B, modified mineral medium with citrate buffer, after 48 h of incubation; C, modified mineral medium with citrate buffer, after 120 h of incubation.

Carbon source	*Acetobacter pasteurianus* 386B			*Acetobacter ghanensis* LMG 23848^T			*Acetobacter senegalensis* 108B		
	A	B	C	A	B	C	A	B	C
Glucose	0	0	0	0	0	1	0	1	1
Fructose	0	0	0	0	0	0	0	1	1
Mannitol	0	0	0	0	0	0	0	1	1
Citric acid	0	0	0	0	0	0	1	0	1
Glycerol	0	0	0	0	0	0	0	1	1
Lactic acid	1	1	1	0	0	0	1	1	1
Ethanol	0	0	0	0	0	0	0	1	1
Acetic acid	0	0	0	0	0	0	0	1	1

3.2 Re-annotation of the *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B genomes

Functional annotation information of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B was stored in their respective relational databases. For *A. ghanensis* LMG 23848^T, it comprised the functional annotations of 2698 protein-encoding genes originally annotated with GenDB, extended with 160 unique protein-encoding genes from NCBI RefSeq, and 306 unique protein-encoding genes from PATRIC. For *A. senegalensis* 108B, it comprised the functional annotations of 3605 protein-encoding genes originally annotated with GenDB, extended with 340 unique protein-encoding genes from NCBI RefSeq, and 625 unique protein-encoding genes from PATRIC.

3.3 Comparative genomics of *Acetobacter* species based on orthogroups

The genome selection procedure resulted in 36 genomes (Table 2).

Importantly, 16 of these genomes had only a contig level quality, of which eight were type strain genomes. Based on the protein-encoding genes in these genomes, OrthoFinder predicted the presence of 6149 orthogroups, of which 155 were orthogroups that contained at least one

protein-encoding gene for each genome considered, and 98 of these orthogroups contained exactly one protein-encoding gene for each genome (so-called single-copy orthogroups).

Table 2. Selection of *Acetobacter* genomes used in the OrthoFinder analysis.

Strain of *Acetobacter* species	NCBI assembly level	NCBI RefSeq identifier
Acetobacter aceti TMW2.1153	Complete Genome	GCF_002005445.1_ASM200544v1
Acetobacter aceti NBRC 14818^{T}	Contig	GCF_004341595.1_ASM434159v1
Acetobacter ascendens LMG 1590^{T}	Complete Genome	GCF_001766235.1_ASM176623v1
Acetobacter cerevisiae LMG 1625^{T}	Contig	GCF_001580535.1_ASM158053v1
Acetobacter cerevisiae LMG 1545	Contig	GCF_001581105.1_ASM158110v1
Acetobacter cibinongensis 4H-1^{T}	Contig	GCF_000963925.1_ASM96392v1
Acetobacter fabarum KR	Contig	GCF_002276555.1_ASM227655v1
Acetobacter ghanensis LMG 23848^{T}	Chromosome	GCF_001499675.1_Acetobacter_ghanensis
Acetobacter indonesiensis 5H-1^{T}	Scaffold	GCF_000963945.1_ASM96394v1
Acetobacter malorum CECT 7742	Scaffold	GCF_001642635.1_ASM164263v1
Acetobacter malorum LMG 1746^{T}	Contig	GCF_001580615.1_ASM158061v1
Acetobacter nitrogenifigens DSM 23921^{T}	Scaffold	GCF_000429165.1_ASM42916v1
Acetobacter okinawensis JCM 25146^{T}	Contig	GCF_000613865.1_ASM61386v1
Acetobacter orientalis 21F-2^{T}	Scaffold	GCF_000963965.1_ASM96396v1
Acetobacter orleanensis CCM 3610	Scaffold	GCF_002358055.1_ASM235805v1
Acetobacter orleanensis JCM 7639^{T}	Scaffold	GCF_000964205.1_ASM96420v1
Acetobacter oryzifermentans SLV-7^{T}	Complete Genome	GCF_001628715.1_ASM162871v1
Acetobacter papayae JCM 25143^{T}	Contig	GCF_000613285.1_ASM61328v1
Acetobacter pasteurianus Ab3	Complete Genome	GCF_001183745.1_ASM118374v1
Acetobacter pasteurianus 386B	Complete Genome	GCF_000723785.2_AP1
Acetobacter pasteurianus subsp. *pasteurianus* LMG 1262^{T}	Scaffold	GCF_000285275.1_ASM28527v1
Acetobacter persici TMW2.1084	Complete Genome	GCF_002006565.1_ASM200656v1
Acetobacter persici JCM 25330^{T}	Contig	GCF_000613905.1_ASM61390v1
Acetobacter pomorum BDGP5	Complete Genome	GCF_002456135.1_ASM245613v1
Acetobacter senegalensis 108B	Complete Genome	GCF_001499615.1_Acetobacter_senegalensis_108B

Table 2. Selection of *Acetobacter* genomes used in the OrthoFinder analysis (continued).

Strain of *Acetobacter* species	NCBI assembly level	NCBI RefSeq identifier
Acetobacter senegalensis LMG 23690^T	Contig	GCF_001580995.1_ASM158099v1
Acetobacter sp. B6	Contig	GCF_004014775.1_ASM401477v1
Acetobacter sp. BCRC 14118	Contig	GCF_003332175.1_ASM333217v1
Acetobacter sp. DmW_043	Contig	GCF_002153485.1_ASM215348v1
Acetobacter sp. DsW_059	Contig	GCF_002153695.1_ASM215369v1
Acetobacter sp. DsW_063	Contig	GCF_002153745.1_ASM215374v1
Acetobacter sp. DsW_54	Contig	GCF_002153575.1_ASM215357v1
Acetobacter sp. JWB	Complete Genome	GCF_003323795.1_ASM332379v1
Acetobacter syzygii 9H-2^T	Scaffold	GCF_000964225.1_ASM96422v1
Acetobacter tropicalis NBRC 16470^T	Scaffold	GCF_000787635.2_ASM78763v2
Acetobacter tropicalis BDGP1	Complete Genome	GCF_002549835.1_ASM254983v1

Based on all predicted orthogroups, OrthoFinder inferred a rooted species tree (Figure 1). The outgroup genomes chosen, namely *E. coli*, *G. oxydans* 621H, and *K. nataicola* RZS01, were successfully placed outside the *Acetobacter* species subtree by OrthoFinder. Moreover, the tree locations of genomes of hereto undescribed *Acetobacter* species provided a first indication to which described *Acetobacter* species they are most related to.

3.4 Reconstruction of genome-scale metabolic models for *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B

The predicted orthogroups were used in combination with the iAp386B454 GEM reference to evaluate the iAp386B454 GEM reaction presence or absence in eleven high-quality *Acetobacter* genomes. The most frequent missing reactions in the central metabolism were the ones catalyzed by proton-translocating NAD(P)$^+$ transhydrogenase (EC 7.1.1.1; missing in *A. aceti* TMW2.1153, *A. tropicalis* BDGP1, *A. pasteurianus* Ab3, and *A. senegalensis* 108B), phosphoglucomutase (EC 5.4.2.2; missing in *A. aceti* TMW2.1153, *A. tropicalis* BDGP1, *A. pasteurianus* Ab3, and *A. senegalensis* 108B), ornithine cyclodeaminase (EC 4.3.1.12; missing in *A. persici* TMW2.1084, *A. aceti* TMW2.1153, and *A. ghanensis* LMG 23848^T), a putative periplasmic mannitol oxidation reaction (missing in *A. ascendens* LMG 1590^T, *A. aceti* TMW2.1153, and *A. pasteurianus* Ab3), ribose-5-phosphate isomerase (EC 5.3.1.6; missing in *A. persici* TMW2.1084, *A. tropicalis* BDGP1, and *A. senegalensis* 108B), and fructokinase (EC 2.7.1.4; missing in *A. ascendens* LMG 1590^T, and *A. ghanensis* LMG 23848^T). Interestingly, both phosphate acetyltransferase (EC 2.3.1.8) and acetate kinase (EC 2.7.2.1) were missing in *A. aceti* TMW2.1153.

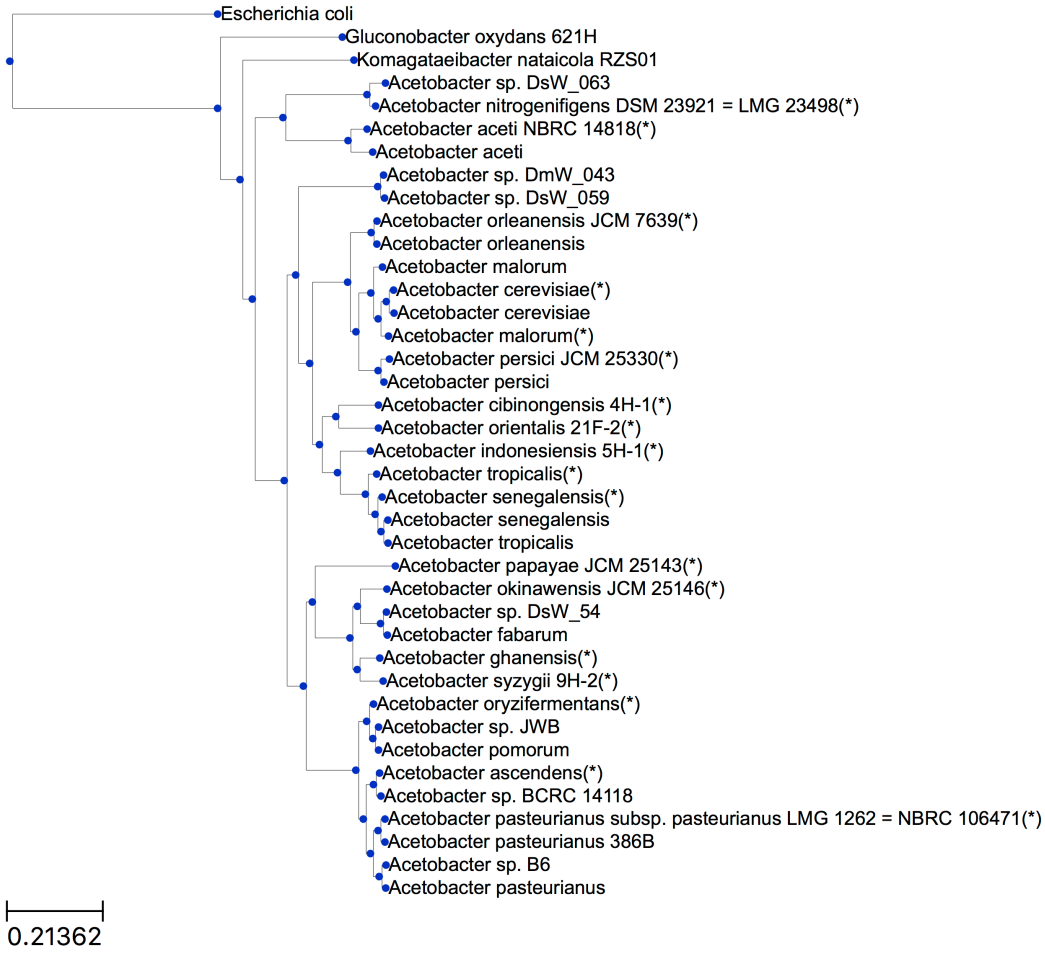


Figure 1. Rooted species tree inference by OrthoFinder. Branch lengths represent the average number of substitutions per site. Type strain genomes are indicated with an asterisk (*). For genomes for which the strain name is not indicated, the reader is referred to Table 2.

Although the lactate:H^+ symporter was missing in *A. aceti* TMW2.1153, the genome contained genes encoding a quinone-dependent (EC 1.1.5.12) and cytochrome *c*-dependent (EC 1.1.2.4) D-lactate dehydrogenase. This strain was originally isolated from a water kefir ecosystem and other *A. aceti* strains have been typically found in an ethanol-rich habitat, examples including the vinegar ecosystem. The absence of the lactate:H^+ symporter in the *A. aceti* TMW2.1153 genome might indicate genome evolution toward excluding lactate as a substrate, other means of lactate transport in the cell, or, possibly, the false absence of a gene encoding this function. In contrast, characteristic reactions for the *Acetobacter* genus, including the membrane-bound pyrroloquinoline quinone (PQQ)-dependent ethanol dehydrogenase (EC 1.1.5.5) and succinyl-CoA:acetate CoA-transferase (EC 2.8.3.18) were found in all high-quality *Acetobacter* genomes considered.

Combining the predicted orthogroups and the iAp386B454 GEM allowed to reconstruct GEMs of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B. Model properties before and after curation-based gap filling are shown in Table 3. The concomitant increase of the number of reactions and decrease of the number of dead-end metabolites indicate a successful gap filling

process. A complication arose in this analysis due to the occurrence of twelve *A. pasteurianus* 386B gene identifiers in the iAp386B454 GEM, which are not present in the *A. pasteurianus* 386B NCBI RefSeq genome version used for OrthoFinder analysis. The ten reactions associated to these genes were computationally filtered and manually checked to see whether the presumed orthologs were present in the *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B genomes, and if so, the reactions and associated gene-reaction rules were manually added to their respective GEMs. This was the case for D-glucose:ubiquinone oxidoreductase (EC 1.1.5.2), phosphoenolpyruvate carboxylase (EC 4.1.1.31), and 3-isopropylmalate dehydratase (EC 4.2.1.33), both for *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B.

Table 3. Properties of genome-scale metabolic models (GEMs) of *Acetobacter ghanensis* LMG 23848^T and *Acetobacter senegalensis* 108B after automated reconstruction and curation-based gap-filling.

GEM property	*A. ghanensis* LMG 23848^T		*A. senegalensis* 108B	
	After reconstruction	After gap-filling	After reconstruction	After gap-filling
Compartments	2	2	2	2
Genes	306	326	339	361
Reactions	294	312	308	326
Exchange reactions	17 (6 %)	17 (5 %)	17 (6 %)	18 (6 %)
Irreversible reactions	121 (41 %)	127 (41 %)	132 (43 %)	139 (43 %)
Orphan reactions	37 (13 %)	37 (12 %)	37 (12 %)	38 (12 %)
Metabolites	294	297	295	298
Dead-end metabolites	26 (9 %)	9 (3 %)	13 (4 %)	4 (1 %)

For *A. ghanensis* LMG 23848^T, a quinone-dependent D-lactate dehydrogenase (EC 1.1.5.12) was absent in the genome, but a cytochrome *c*-dependent enzyme (EC 1.1.2.4) was identified. No acetolactate decarboxylase (EC 4.1.1.5) was found in the *A. ghanensis* LMG 23848^T genome. Instead, a gene encoding diacetyl reductase (EC 1.1.1.303) was present. Finally, no quinone-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.5.3) was found. For *A. senegalensis* 108B, no proton-translocating NAD(P)$^+$ transhydrogenase (EC 7.1.1.1) was found in the genome. However, a soluble NAD(P)$^+$ transhydrogenase (EC 1.6.1.1) was present. Additionally, no gene could be found encoding gluconokinase (EC 2.7.1.12), possibly explaining why *A. senegalensis* 108B could not grow under defined medium conditions with glucose as the sole carbon source. The fact that citrate could sustain growth of this strain paved the way to find a genomic cause. Indeed, a putative citrate:H$^+$ symporter and a gene cluster encoding citrate lyase was found in the genome, as previously reported (Illeghems *et al.*, 2016). The gene cluster encoding citrate lyase was also present in two *Acetobacter* species isolated from fruit or fruit-derived fermented foods, namely *A. persici* JCM 25330^T (peach fruit) and *A. malorum* CECT 7742 (strawberry vinegar). Finally, the *A. senegalensis* 108B genome

encoded a phosphoenolpyruvate carboxykinase (EC 4.1.1.49), which was not found in *A. pasteurianus* 386B, thus providing another anaplerotic link between the TCA cycle and the gluconeogenesis pathway in *A. senegalensis* 108B.

3.5 Gap filling of genome-scale metabolic models and flux balance analysis

The results of the growth experiments for *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B were used to fill the reaction gaps in these models using the iAp386B454 GEM as a reaction repository. The FBA gap filling results indicated that 17 iAp386B454 GEM reactions were needed to obtain *in silico* growth of *A. ghanensis* LMG 23848^T with glucose as the sole carbon source. For *A. senegalensis* 108B, ten iAp386B454 GEM reactions were necessary when growth was required with lactate as the sole carbon source. Manual curation allowed the addition of all but two reactions for *A. ghanensis* LMG 23848^T, for which no genomic evidence could be found. These reactions were the ones catalyzed by asparagine synthase (EC 6.3.5.4) and ornithine cyclodeaminase (EC 4.3.1.12). Both are involved in amino acid biosynthesis, the first being necessary for L-asparagine biosynthesis and the latter responsible for L-proline biosynthesis. Interestingly, the gene encoding the alternative L-proline biosynthesis enzyme NAD(P)-dependent pyrroline-5-carboxylate reductase (EC 1.5.1.2) was annotated with an internal stop codon in the NCBI RefSeq annotation. Other biosynthesis pathways for these amino acids may exist for *A. ghanensis* LMG 23848^T, since *in vitro* growth was obtained under defined medium conditions without the addition of amino acids. Conversely, for *A. senegalensis* 108B, genomic evidence was found for all gap-filling results. However, this evidence had to be retrieved from genes in the NCBI RefSeq annotation for which no amino acid sequence was available (*e.g.*, due to a predicted frameshift) or genes only annotated in the GenDB or PATRIC genome annotation versions.

Flux balance analysis with the iAg23848 GEM on glucose as the sole carbon source resulted in no growth, unless L-proline and L-asparagine were additionally available, in which case the predicted specific growth rate was 0.70 h^{-1} and both amino acids' consumption was balanced by the protein biosynthesis reaction. Interestingly, the flux distribution revealed obligatory NADPH oxidation by the proton-translocating NAD(P)$^+$ transhydrogenase. However, genome annotation analysis revealed that evidence for the presence of this reaction was limited, since the *A. ghanensis* LMG 23848^T genome missed the β-subunit of this enzyme. Growth of *A. ghanensis* LMG 23848^T under defined medium conditions with glucose was only found when a citrate buffer was used. Allowing additional *in silico* consumption of citrate revealed that constraining the proton-translocating NAD(P)$^+$ transhydrogenase did not prevent growth, resulting in the prediction of a specific growth rate of 0.58 h^{-1} with the glutamate synthase reaction (EC 1.4.1.13) as the largest NADPH consumer.

The iAs108B GEM was solved with FBA under two growth conditions, namely with lactate or with citrate as the sole carbon sources. No additional nutrients had to be added to obtain *in silico* growth, indicating that this model contained all reactions necessary for biomass formation. FBA with lactate resulted in a specific growth rate of 0.86 h^{-1}. The soluble NAD(P)$^+$ transhydrogenase produced NADPH, whereas phosphoenolpyruvate carboxykinase supplied

all necessary phosphoenolpyruvate. FBA with citrate resulted in a specific growth rate of 0.75 h^{-1}. Most of the citrate consumed was directly used in the TCA cycle, whereas only 12 % was consumed by citrate lyase. No acetate was produced by the model; all intracellular acetate produced was consumed by succinyl-CoA:acetate CoA-transferase.

3.6 Presence of the glyoxylate cycle and aerobic respiratory chain reactions

The presence of genes encoding the glyoxylate cycle enzymes was assessed in all 36 *Acetobacter* genomes. Presence/absence was compared to experimental data related to the growth of the strains in a defined medium containing ethanol as the sole carbon source and obtained for the type strains of a number of *Acetobacter* species (Table 4; Cleenwerck et al., 2008). For eight out of 19 *Acetobacter* species that were able to grow, three species (*A. estunensis*, *A. peroxydans*, and *A. lovaniensis*) have currently no genome sequenced, two species for which a genome sequence was available (*A. aceti* and *A. nitrogenifigens*) encoded the glyoxylate cycle, and finally, three species with a genome sequence available (*A. senegalensis*, *A. cibinongensis*, and *A. fabarum*) did not encode the glyoxylate cycle. The latter three species could thus correspond to false positive experimental results based on growth with ethanol as the sole carbon source, albeit that the genome sequence available for *A. fabarum* is not from the type strain.

The enzymes involved in the aerobic respiratory chain as known for *E. coli* were present in all *Acetobacter* genomes considered, comprising NADH dehydrogenase type I (EC 7.1.1.2), NADH dehydrogenase type II (EC 1.6.5.9), cytochrome bo_3 oxidase (EC 7.1.1.3), and cytochrome *bd* oxidase (EC 7.1.1.7). Concerning NADPH metabolism, *E. coli* contains both a proton-translocating $NAD(P)^+$ transhydrogenase and a soluble $NAD(P)^+$ transhydrogenase. From the comparative genomics analysis, it was apparent that, at least for the high-quality *Acetobacter* genomes, the presence of these enzymes was mutually exclusive. An interesting exception was *A. aceti*, for which both reactions were predicted to be present.

4 Discussion

The reconstruction of a GEM for *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B, two candidate starter culture strains for cocoa bean fermentation processes, provided *in silico* tools to gain insight into the metabolic properties of these strains. In the current study, the GEM reconstruction process was accelerated using a semi-automated approach with experimental validation. Growth experiments under defined medium conditions identified individual carbon sources that allowed the growth of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B. These growth conditions could then be used in simulations using FBA with the newly reconstructed iAg23848 and iAs108B GEMs, providing insights into the metabolic flux distributions of the models.

Table 4. Overview of the presence of the glyoxylate cycle enzymes in *Acetobacter* genomes. +: growth; -: no growth; w: weak growth according to Cleenwerck *et al.,* 2008. If the results of only one strain are reported, this is always the type strain.

Acetobacter species (number of strains tested in Cleenwerck *et al*., 2008)	Growth on ammonium + ethanol according to Cleenwerck *et al*., 2008	Number of genomes included in the OrthoFinder analysis	Number of genomes containing a gene encoding isocitrate lyase (EC 4.1.3.1)	Number of genomes containing a gene encoding malate synthase (EC 2.3.3.9)
A. fabarum (4)	+	1	0	0
A. lovaniensis (1)	+	0	0	0
A. ghanensis (3)	-	1	0	0
A. syzigii (1)	-	0	0	0
A. pasteurianus (7)	-	3	0	0
A. pomorum (1)	-	1	0	0
A. peroxydans (2)	+	0	0	0
A. indonesiensis (2)	-	1	0	0
A. orientalis (1)	-	1	0	0
A. cibinongensis (1)	w	1	0	0
A. tropicalis (2)	-	2	0	0
A. senegalensis (3)	+	2	0	0
A. orleanensis (4)	-	2	0	0
A. malorum (1)	-	2	0	0
A. cerevisiae (4)	-	2	0	0
A. nitrogenifigens (1)	+	1	1	1
A. oeni (1)	-	0	0	0
A. aceti (4)	+	2	2	2
A. estunensis (3)	+	0	0	0

For *A. senegalensis* 108B, the prediction of citrate assimilation based on a previous genome analysis (Illeghems et al., 2016) was confirmed. In general, conversion of citrate present in the cocoa pulp-bean mass is ascribed to lactic acid bacteria that proliferate at the start of the cocoa bean fermentation process (De Vuyst and Weckx, 2016a). Citrate consumption by AAB species may contribute to their survival during this stage of the cocoa bean fermentation process (Illeghems *et al*., 2016). For *A. ghanensis* LMG 23848^T, the absence of a quinone-dependent D-lactate dehydrogenase (EC 1.1.5.12) is probably compensated by the presence of a cytochrome *c*-dependent enzyme (EC 1.1.2.4), which probably allows lactate consumption. However, this may be an explanation for the lower lactate consumption rate of *A. ghanensis* LMG 23848^T, compared to *A. pasteurianus* 386B, under cocoa pulp-simulating conditions (Moens *et al.*, 2014). In general, both lactate and acetate are overoxidized during a late stage of the cocoa bean fermentation process (Moens *et al*., 2014; De Vuyst and Weckx, 2016a).

Additionally, no acetolactate decarboxylase (EC 4.1.1.5) was found in the *A. ghanensis* LMG 23848^T genome, although acetoin was formed (Moens *et al.*, 2014). In *A. pasteurianus*, it has been shown that acetoin production was solely depending on acetolactate decarboxylase, in contrast to the theoretically proposed acetaldehyde carboligation pathway, as found in *Saccharomyces cerevisiae* (De Ley, 1959; Gunawan *et al*., 2007). However, in *A. ghanensis* LMG 23848^T as well as in *A. senegalensis* 108B, yet another pathway may be responsible for acetoin formation, since in both genomes a gene encoding diacetyl reductase (EC 1.1.1.303) was present, as is the case in lactic acid bacteria (Keenan and Lindsay, 1968). Finally, no quinone-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.5.3) could be found, possibly explaining why *A. ghanensis* LMG 23848^T could not grow under defined medium conditions with glycerol as the sole carbon source.

Re-annotation of the genomes of *A. ghanensis* LMG 23848^T and *A. senegalensis* 108B proved to be critical to reconstruct their GEMs. Reaction presence predictions using the iAp386B454 GEM as a reference were successful, but led to metabolic network gaps in essential biosynthesis reactions. Without the current re-annotation effort, it would be impossible to speculate about their presence. Keeping this in mind, the mutual occurrence of genome sequence errors in query and reference genomes blurs the outcome of this functional comparative genomics analysis, since it is ultimately based on the prediction of the presence or absence of genes in the genomes considered. However, considering that genomic data are the only well-structured and publicly available data source at hand to compare the metabolic properties of *Acetobacter* species, this analysis may provide interesting insights for future research. The predicted differences in the carbon and redox metabolic potential of the *Acetobacter* species revealed the possible metabolic diversity in this genus. Specifically, the presence or absence of the lactate:H^+ symporter, fructokinase, gluconokinase, ribose-5-phosphate isomerase, and citrate lyase may represent adaptations to different habitats and niches. Also, the presence or absence of a proton-translocating and soluble NAD(P)$^+$ transhydrogenase may represent an important and hitherto not sufficiently explored adaptation to the usage of different carbon sources and their influence on the redox metabolism of *Acetobacter* species. In *E. coli*, both enzymes are present, each having a specific functional role, whereby the first enzyme produces NADPH and the latter re-oxidizes excess NADPH (Sauer *et al.*, 2004). The complete coverage of the key reactions of the aerobic respiratory chain among the *Acetobacter* genomes is a strong indication that this metabolic feature is a defining physiological factor of this genus.

The predicted presence or absence of the glyoxylate cycle enzymes represent an interesting case where genomic and experimental evidence probably go hand in hand. Alternative pathways have been described for the assimilation of C2 compounds in microorganisms, namely the ethylmalonyl-CoA pathway and the methylaspartate cycle (Alber *et al.*, 2006; Khomyakova *et al.*, 2011). However, these pathways have not been found in AAB, and no genomic evidence for their presence could be found in the *Acetobacter* genomes considered. If the glyoxylate cycle is the only C2 assimilation pathway present in species of the genus *Acetobacter*, then the phenotypic test concerning the growth of an *Acetobacter* species in a medium with ethanol as the sole carbon source, commonly used to infer the taxonomy of an

unknown *Acetobacter* strain, should be backed by genomic evidence. The results presented here indicate that this was not the case for some *Acetobacter* species, thus revealing potential false positive experimental data (Cleenwerck *et al*., 2008). *Acetobacter senegalensis* 108B was not able to grow under defined medium conditions with ethanol as the sole carbon source. In addition, no genomic evidence was found for the glyoxylate cycle in the re-annotated *A. senegalensis* 108B genome. The experimental results presented here are in contrast to previous results obtained for *A. senegalensis* 108B and for the type strain of *A. senegalensis* (Camu *et al.*, 2007; Ndoye *et al.*, 2007). Thus, this is an argument that genomic evidence should be taken into account for future species descriptions of members of the *Acetobacter* genus.

From a practical viewpoint, in fermentation processes such as vinegar production, the use of strains that can solely oxidize ethanol to acetic acid, but not assimilate it into biomass components, may be favorable (Mullins and Kappock, 2013; Sakurai *et al.*, 2013). Similarly, for the cocoa bean fermentation process, the absence of the glyoxylate cycle in the genomes of *A. ghanensis* LMG 23848^{T}, *A. senegalensis* 108B, and *A. pasteurianus* 386B (Illeghems *et al*., 2013; Pelicaen *et al*., 2019), may represent a competitive advantage of these strains to quickly oxidize the available ethanol, produced by yeast species, to acetic acid. Genomic screening of strains may provide a more straight-forward strategy to select those strains for which no evidence of the glyoxylate cycle can be found, to use them as starter cultures in such fermentation processes.

5 Conclusion

In this study, the possibility of integrating two different data structures based on the same genomic data source was explored. A first data structure comprises GEMs, allowing to catalogue the entire biochemical reaction potential of a microbial strain. The mathematical form of these models has the added value to allow *in silico* metabolic flux simulations. Also, gene-reaction rules, whether or not manually curated, allow to explore the causal links between the genome and the reactome. A second data structure, the orthogroups, predicts the phylogenetic relatedness between protein-encoding genes from different species' genomes, from which estimates of their shared function can be made. Combining the manually curated gene-reaction rules of a GEM of *A. pasteurianus* 386B and *E. coli* and a set of predicted orthogroups of a selection of *Acetobacter* genomes, allowed to predict the presence of reactions that characterized the metabolism of *A. pasteurianus* 386B and *E. coli* in these *Acetobacter* species. Evaluation of biochemical reaction presence in bacterial genomes represents a new frontier in genome annotation, since this evaluation is a step closer to the actual biological functionalities that the genome encodes for. The results obtained stress the need for decent data curation and species description.

Chapter 5

Genome-scale metabolic modeling of *Acetobacter pasteurianus* 386B reveals its metabolic adaptation to cocoa fermentation conditions

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

Supplementary material for this Chapter is available in the Annexes.

Summary

Acetobacter pasteurianus 386B has been selected as a candidate functional starter culture to better control the cocoa fermentation process. Previously, its genome has been sequenced and a genome-scale metabolic model (GEM) has been reconstructed. To understand its metabolic adaptation to cocoa fermentation conditions, different flux balance analysis (FBA) simulations were performed and compared with experimental data. In particular, metabolic flux distributions were simulated for two phases that characterize the growth of *A. pasteurianus* 386B under cocoa fermentation conditions, predicting a switch in respiratory chain usage in between these phases. The possible influence on the resulting energy production was shown using a reduced version of the GEM. FBA simulations revealed the importance of the compartmentalization of the ethanol oxidation reactions, namely in the periplasm or in the cytoplasm, and highlighted the potential role of ethanol as a source of carbon, energy, and NADPH. Regarding the latter, the physiological function of a proton-translocating $NAD(P)^+$ transhydrogenase was further investigated *in silico*. This study revealed the potential of using a GEM to simulate the metabolism of *A. pasteurianus* 386B, and may provide a general framework toward a better physiological understanding of functional starter cultures in food fermentation processes.

1 Introduction

The cocoa fermentation process is of high economic importance, as it is fundamental to transform raw cocoa beans into cured ones through fermentation and drying, from which products such as chocolate can be made (Santander Muñoz *et al*., 2019; De Vuyst and Weckx, 2016a). Although the demand for chocolate is increasing worldwide, the cocoa fermentation process is currently still a spontaneous process, yielding fermented dry cocoa beans with variable quality. This fermentation process takes about four to six days and is characterized by successive metabolic activities of yeasts, lactic acid bacteria (LAB), and acetic acid bacteria (AAB). The latter are primarily responsible for the formation of acetic acid (and acetoin) and for the increase in temperature during the fermentation process, which both contribute to the biochemical formation of cocoa flavor precursors inside the cocoa beans (De Vuyst and Weckx, 2016a; Aprotosoaie *et al*., 2016).

Field experiments using cocoa pulp-bean mass as well as laboratory fermentation experiments using a cocoa pulp simulation medium (CPSM) have indicated that yeasts and LAB grow on the glucose, fructose, and citric acid present in the cocoa pulp, whereby yeasts secrete mainly ethanol and LAB mainly lactic acid at the initial stage of the fermentation process (Lagunes Gálvez *et al*., 2007; Camu *et al*., 2008; Adler *et al*., 2014; Ho *et al*., 2014; Moens *et al*., 2014; Aprotosoaie *et al*., 2016; De Vuyst and Weckx, 2016a). The ethanol and lactic acid produced are subsequently consumed by AAB to sustain their growth. Therefore, AAB depend on the amount of oxygen available, since these bacteria are obligate aerobes, relying on aerobic respiration for their energetic needs (Matsushita and Matsutani, 2016). Two main factors contribute to air ingress into the cocoa pulp-bean mass during the fermentation process, namely the liquefaction of the mucilaginous cocoa pulp and the practice of turning the cocoa pulp-bean mass (Camu *et al*., 2008; De Vuyst and Weckx, 2016a).

From a collection of *Acetobacter* isolates obtained from spontaneous cocoa fermentation processes, *A. pasteurianus* 386B has been characterized physiologically and subsequently selected as a candidate functional starter culture, combining a relatively high experimental biomass yield and rapid co-consumption of ethanol and lactic acid to produce acetic acid and acetoin (Lefeber *et al*., 2010; Moens *et al*., 2014). In addition, it has been suggested that *A. pasteurianus* has a second phase of growth, whereby the acetic acid and acetoin produced are overoxidized once lactic acid and ethanol are depleted (Adler *et al*., 2014; Moens *et al*., 2014). A similar diauxic growth curve has been shown for *Acetobacter aceti* growing on ethanol, and for *Gluconobacter oxydans* growing on glucose (Sakurai *et al*., 2012; Hanke *et al*., 2013; Mullins and Kappock, 2013).

Nowadays, genome-scale metabolic models (GEMs) are used to gain system-level insights into the metabolic capacity of microorganisms (Palsson, 2006). A GEM is an *in silico* representation of the metabolism of a bacterial cell population, based on genomic information (Thiele and Palsson, 2010; Gottstein *et al*., 2016; de Jong *et al*., 2017). To simulate the intracellular metabolic reaction flux distribution, flux balance analysis (FBA) may be used

(Orth *et al*., 2010). FBA performs an optimization of the metabolic reaction flux values of the GEM to obtain the optimal value of an objective function of choice, typically the maximization of the biomass reaction flux, in which case the model returns the predicted specific growth rate. Simulations are performed at steady-state, meaning that the intracellular metabolite concentrations are assumed to remain invariable over time. Additional constraints that can be applied to the model include reaction directionality and reaction capacity constraints. Their differential application allows to simulate different growth conditions the microorganism under study may encounter. In this way, the influence of these growth conditions on the metabolism of the characterized strain can be investigated *in silico*.

Previously, the *A. pasteurianus* 386B genome has been sequenced and annotated (Illeghems *et al*., 2013). Subsequent re-annotation and reconstruction of a GEM has been performed with the aim to obtain a system-level view on its metabolism (Pelicaen *et al*., 2019). The present study aims at the understanding of the metabolic adaptation of *A. pasteurianus* 386B to cocoa fermentation conditions through an *in silico* investigation using the GEM reconstructed previously. Further, the influence of the biomass reaction used on the resulting metabolic flux predictions and the respiratory chain energetics were assessed in more detail.

2 Materials and methods

2.1 Genome-scale metabolic models and constraint-based modeling methods

Three different GEMs were used to perform FBA simulations (Table 1).

Table 1. Overview of genome-scale metabolic models used in this study. The models iAp386Bcore and iAp386Benergy were derived from iAp386B454 to only simulate central carbon metabolism and energy production, respectively.

Model	iAp386B454	iAp386Bcore	iAp386Benergy
Compartments	2	2	2
Genes	454	143	94
Reactions	322	91	44
Exchange reactions	17	14	3
Metabolites	296	70	41

The main *A. pasteurianus* 386B GEM, iAp386B454, consists of a genome-scale metabolic network reconstruction of *A. pasteurianus* 386B including a biomass reaction (Pelicaen *et al*., 2019). A first derived GEM, iAp386Bcore, was extracted from iAp386B454 containing all metabolic reactions of *A. pasteurianus* 386B involved in central carbon metabolite conversions. To simulate biomass formation in iAp386Bcore, a previously defined biomass equation was incorporated, containing a set of universal biomass precursors (Edirisinghe *et al*.,

2016). A second derived GEM, iAp386Benergy, focusing on energy production, was extracted from iAp386B454 containing all reactions needed for the production of ATP from a set of carbon substrates, more precisely acetoin, lactate, ethanol, and acetate. The GEMs were analyzed using the COBRApy package version 0.11.3 (Ebrahim *et al*., 2013). Three constraint-based modeling methods were used to explore the optimal solution space of the GEMs. FBA was used to find the optimal value of an objective function taking into account a set of constraints. Parsimonious FBA (pFBA) was applied to minimize the cumulative sum of reaction fluxes, using the optimal value of the objective function as an additional constraint. Flux variability analysis (FVA) was used to find the maximal and minimal flux values of each reaction, using a fraction of the optimal value of the objective function as an additional constraint. More precisely, the loopless variant of FVA was used, which omits reaction flux cycles from the analysis, and the fraction of the optimum was set to 99 %. Optimization of the GEMs was performed using the CPLEX solver from IBM (Armonk, New York, USA). The GEM's consumption of ammonium as nitrogen source, sulfate as sulfur source, and phosphate as phosphate source was allowed without any constraints. For aerobic respiration, oxygen influx was allowed without constraint. The mapping of simulated metabolic flux data onto reconstructed metabolic pathways was visualized using Escher (King *et al*., 2015). The carbon recovery was calculated as the ratio of the sum of C-mol product metabolites to the sum of C-mol substrate metabolites.

2.2 Simulation of the two growth phases of *A. pasteurianus* under CPSM conditions

Previous research has indicated that *A. pasteurianus* 386B and *A. pasteurianus* NCC 316 have two phases of growth when grown in CPSM supplemented with lactate and ethanol (Adler *et al*., 2014; Moens *et al*., 2014). Metabolic flux data of the latter strain were used in the present study as a reference to simulate each phase for *A. pasteurianus* 386B. To be able to compare the resulting flux distributions in both phases, the maximum exchange flux of the carbon sources was set to 60 C-mmol/g_{CDW}/h and exchange fluxes were rescaled according to this maximum (Supplementary Tables S1 and S2). This value was chosen as it corresponds to a consumption of 10 mmol/g_{CDW}/h of glucose for *Escherichia coli* (Varma *et al*., 1993). To simulate the first growth phase of *A. pasteurianus* 386B, the reported stoichiometric yields and standard deviations (SD) of carbon substrates and products of *A. pasteurianus* NCC 316 were used to define the iAp386B454 exchange reaction constraints. This resulted in two sets of exchange reaction constraints, namely one set of consumption flux constraints for lactate and ethanol and one set of production flux constraints for acetate and acetoin. To simulate the second growth phase of *A. pasteurianus* 386B, analysis of metabolite time-series data of *A. pasteurianus* NCC 316 have shown that acetoin, acetate as well as pyruvate are co-consumed (Adler *et al*., 2014). Here, consumption fluxes of each respective carbon source were set proportionally to their respective peak metabolite concentrations during fermentation, taking into account the maximum carbon source consumption of 60 C-mmol/g_{CDW}/h. For each simulated growth phase, the biomass reaction was set as the objective function.

2.3 Physiological role of lactate and ethanol

To investigate the contribution of lactate and ethanol to biomass formation in iAp386B454, different simulations were performed. In each case, the maximal consumption flux of lactate and ethanol was set to 12 mmol/g_{CDW}/h. First, iAp386B454 was left unconstrained, except for the lactate and ethanol consumption fluxes. Second, ethanol and acetate consumption in the cytoplasm was not allowed. Third, the fraction of ethanol that could be taken up in the cytoplasm was fixed, allowing the GEM to find the optimal ethanol to lactate consumption flux ratio leading to a maximal biomass formation. In the latter case, acetate consumption in the cytoplasm was not allowed. For each step-size increase of 2 mmol/g_{CDW}/h in lactate consumption flux, the ethanol consumption flux was varied with a step-size of 0.1 mmol/g_{CDW}/h. The resulting co-consumption flux values were used to simulate biomass formation with FBA.

2.4 Theoretical energy production

The *A. pasteurianus* 386B aerobic respiratory chain is composed of a set of respiratory enzymes with NADH and ubiquinol as main electron donors, including NADH:ubiquinone oxidoreductase (NADH dehydrogenase) type I (NDH type I) and type II (NDH type II) for NADH oxidation, as well as ubiquinol oxidase (UOX) ba_3, UOX *bd*, and a putative cyanide-insensitive oxidase for ubiquinol oxidation (Illeghems *et al*., 2013; Pelicaen *et al*., 2019). Respiratory enzymes were classified according to their respective electron donor (NADH or ubiquinol) and their theoretical proton-pumping efficiency ($H^+/2e^-$, H^+/O). Since UOX *bd* and cyanide-insensitive oxidase have the same theoretical H^+/O value, they were collectively defined as UOX *bd*-type. This allowed to define four distinct respiratory chains, each composed of an NADH dehydrogenase and a terminal oxidase. FBA was used with the iAp386Benergy model to calculate energy production with different substrates in different metabolic modes and compare the results to a previous theoretical analysis (Adler *et al*., 2014). For each metabolic mode, the relevant substrate consumption and product formation reactions were defined and specific reactions were blocked to assure identical oxidation pathways when varying the respiratory chain used (Supplementary Table S3). Different energetic parameters of the aerobic respiratory chain were computed using the iAp386Benergy simulation results. These included:

i) the H^+/O ratio, defining the number of translocated protons by the electrogenic respiratory enzymes for reduction of one oxygen atom:

$$Y_{\frac{H^+}{O}} = \frac{1}{2} * \frac{v_{H^+}}{v_{O_2}}$$

(ii) the P/O ratio, representing the amount of high-energy phosphate bounds produced by ATP synthase for each reduced oxygen atom:

$$Y_{\frac{ATP}{O}} = \frac{1}{2} * \frac{v_{ATP\ synthase}}{v_{O_2}}$$

(iii) the ATP production yield on the specified substrate:

$$Y_{\frac{ATP}{S}} = \frac{v_{ATP}}{v_S}$$

with v_{H^+} the translocation flux (mmol/g_{CDW}/h) of periplasmic protons by the electrogenic respiratory enzymes, v_{O_2} the oxygen uptake flux (mmol/g_{CDW}/h), $v_{ATP\ synthase}$ the flux of ATP synthase (mmol/g_{CDW}/h), v_{ATP} the total ATP production flux (mmol/g_{CDW}/h), and v_S the substrate consumption flux (mmol/g_{CDW}/h). The stoichiometry of the ATP synthase encoded in iAp386Benergy corresponded with a $Y_{ATP/H}^{+}$ = 1/4, which is the current consensus value for *Escherichia coli* ATP synthase (Steigmiller *et al*., 2008). A similar formula as for (iii) was used to compute the NADH and ubiquinol-9 yield on the substrate specified. The proton-translocating NAD(P)$^+$ transhydrogenase reaction was blocked to be able to compare the simulation results with a previous theoretical analysis (Adler *et al*., 2014). Exchange reactions for carbon dioxide and water were added to the model, as well as a proton exchange flux to simulate the exact protonation state of the carbon source. FBA was performed with maximization of the flux of an ATP hydrolysis reaction as the objective function of the simulation.

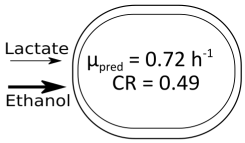
3 Results

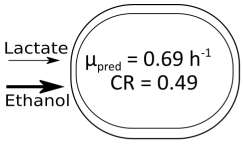
3.1 Simulation of two growth phases of *A. pasteurianus* 386B under CPSM conditions

The growth of *A. pasteurianus* 386B under cocoa fermentation conditions is characterized by two phases of growth, namely a first phase where ethanol and lactic acid are co-consumed, followed by a second phase where the acetic acid and acetoin produced are overoxidized.

Using the consumption flux constraints of lactate and ethanol only, *in silico* simulation of the first growth phase of *A. pasteurianus* 386B with iAp386B454 and iAp386Bcore resulted in a complete consumption of lactate and ethanol and led to the formation of carbon dioxide, water, and acetate (Figure 1A). For both models, no acetoin was produced. However, the localization of ethanol oxidation and the origin of the secreted acetate was predicted to be different for both models, depending on the constraint-based modeling method used. FBA using iAp386B454 predicted intracellular NAD(P)$^+$-dependent ethanol oxidation, leading to the formation of intracellular acetate that was assimilated in the tricarboxylic acid (TCA) cycle by succinyl-CoA:acetate CoA-transferase or converted into acetyl-CoA by means of acetate kinase and phosphate acetyltransferase. As not all acetate could be assimilated, acetate was secreted by the model as an overflow metabolite using the acetate ABC transporter. In contrast, pFBA predicted mainly periplasmic ethanol oxidation, leading to secreted periplasmic acetate. FBA and pFBA simulations with iAp386Bcore predicted periplasmic oxidation of ethanol and uptake of periplasmic acetate in the cytoplasm. For all simulations, NADH and ubiquinol were oxidized by means of a respiratory chain consisting of NDH type I and UOX ba_3. However, FVA showed flux variability for the different respiratory chain reactions, indicating that the resulting FBA flux distribution was not unique. The predicted specific growth rate was 0.72 h^{-1} with a carbon recovery of 0.49 for iAp386B454 and 0.69 h^{-1} with an identical carbon recovery for iAp386Bcore. The difference in predicted specific growth rates reflected the different biomass equations used in both models.

A





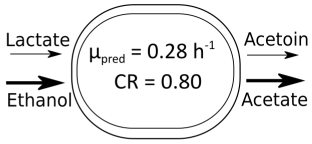
iAp386B454

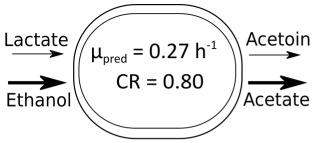
IN	pFBA	FVA max	FVA min
Lactate*	7.64	7.64	7.56
Ethanol*	18.55	19.33	18.55
Oxygen	21.88	47.86	19.32
Phosphate	0.51	0.51	0.50
Ammonium	6.70	6.70	6.63
Sulfate	0.14	0.14	0.14
OUT			
Acetoin	0.00	0.00	-0.04
Acetate	-11.64	0.00	-14.16
CO_2	5.93	31.09	2.75
H_2O	41.21	66.99	38.45
H^+	-10.17	1.54	-12.63

iAp386Bcore

IN	pFBA	FVA max	FVA min
Lactate*	7.64	7.64	7.56
Ethanol*	18.55	19.33	18.55
Oxygen	23.40	55.23	23.31
Phosphate	2.90	2.90	2.88
Ammonium	NA	NA	NA
Sulfate	NA	NA	NA
OUT			
Acetoin	0.00	0.00	-0.04
Acetate	-14.61	0.00	-15.55
CO_2	0.00	31.09	0.00
H_2O	31.48	63.23	31.35
H^+	-16.77	-2.06	-17.61

B





iAp386B454

IN	pFBA	FVA max	FVA min
Lactate*	7.64	7.64	7.61
Ethanol*	18.55	19.33	18.55
Oxygen	20.06	23.84	20.06
Phosphate	0.20	0.20	0.20
Ammonium	2.63	2.63	2.60
Sulfate	0.06	0.06	0.06
OUT			
Acetoin*	-2.32	-2.32	-2.34
Acetate*	-16.90	-16.25	-17.33
CO_2	4.85	7.83	4.85
H_2O	29.96	33.66	29.89
H^+	-11.68	-11.01	-12.11

iAp386Bcore

IN	pFBA	FVA max	FVA min
Lactate*	7.64	7.64	7.61
Ethanol*	18.55	19.33	18.55
Oxygen	22.29	26.73	22.28
Phosphate	1.14	1.14	1.13
Ammonium	NA	NA	NA
Sulfate	NA	NA	NA
OUT			
Acetoin*	-2.32	-2.32	-2.34
Acetate*	-17.24	-16.25	-17.33
CO_2	4.16	7.83	4.16
H_2O	27.78	32.19	27.73
H^+	-13.45	-12.42	-13.53

Figure 1. Parsimonious FBA solution of the iAp386B454 and iAp386Bcore models to simulate the first growth phase of *Acetobacter pasteurianus* 386B under cocoa fermentation conditions. Flux values of exchange reactions are listed as well as their respective maximal and minimal flux values, computed by loopless flux variability analysis. All fluxes are expressed in mmol/g_{CDW}/h. Constrained exchange fluxes are indicated with an asterisk and shown in the model scheme. The predicted specific growth rate and carbon recovery (CR) are reported for each simulation. CO_2, carbon dioxide. NA, not applicable. A, only consumption fluxes were constrained; B, consumption as well as production fluxes were constrained.

The flux distributions obtained were optimal in the sense that available substrates were used most efficiently by the reconstructed metabolic network of *A. pasteurianus* 386B to maximize the flux of its biomass reaction. FBA thus simulated a metabolic flux distribution that maximized the biomass yield of *A. pasteurianus* 386B on lactate and ethanol but it did not predict the acetoin formation that has been found previously (Adler *et al*., 2014; Moens *et al*.,

2014). To be able to simulate the metabolic flux distribution of *A. pasteurianus* 386B more accurately, additional production flux constraints were added for acetate and acetoin (Figure 1B). In this case, iAp386B454 with pFBA predicted a specific growth rate of 0.28 h^{-1} with a carbon recovery of 0.80. The predicted carbon dioxide production flux was 4.85 mmol/g_{CDW}/h. Furthermore, the metabolic flux distribution was characterized by a cytoplasmic ethanol uptake fraction of 16 %. The predicted respiratory chain solely consisted of UOX ba_3 and UOX *bd*. Omitting the production flux constraint for acetate gave similar results, indicating that carbon loss due to acetoin secretion led to a relatively low flux in the TCA cycle thus limiting the amount of assimilated acetate.

In silico simulation of the second growth phase of *A. pasteurianus* 386B using FBA resulted in a complete consumption of acetoin, acetate, and pyruvate (Supplementary Figure S1). In this simulation, pyruvate and periplasmic proton consumption was necessary to achieve growth and the carbon sources were fully oxidized to carbon dioxide and water. This was accompanied by a lower predicted specific growth rate (0.14 h^{-1}) and a higher oxygen consumption flux (56.19 mmol/g_{CDW}/h) compared to the first growth phase. Production of acetaldehyde and acetyl-CoA by acetoin dehydrogenase led to consumption of acetate and acetyl-CoA in almost balanced succinyl-CoA:acetate CoA-transferase and citrate synthase reaction fluxes of the TCA cycle. To prevent the occurrence of futile cycles in the predicted flux distribution, an ATP hydrolysis reaction was activated in the model that consumed 86 % of the ATP produced. This indicated an excess energy formation.

Comparison of the flux distributions in both growth phases (Figure 2) revealed a relatively low TCA cycle flux for the first growth phase (Figure 2A) compared to the second growth phase (Figure 2B). Also, in contrast with the first growth phase, the respiratory chain of the second growth phase was solely composed of NDH type II and UOX *bd*, since only these reactions were present in the optimal solution space, as obtained by FVA.

3.2 Physiological role of lactate and ethanol

As *A. pasteurianus* 386B has two sets of ethanol-oxidizing enzymes, one cytoplasmic set and one periplasmic set, the influence of these sets of enzymes on the resulting specific growth rate was investigated. Co-consumption of lactate and ethanol was simulated, whereby the amount of ethanol that could be oxidized in the cytoplasm was constrained. Then, at different fixed flux values of lactate consumption, the ethanol consumption flux value was varied and the predicted specific growth rate was recorded (Figure 3). For the maximal flux value of substrate co-consumption tested in the simulations, iAp386B454 predicted a specific growth rate of 1.05 h^{-1} when cytoplasmic consumption of ethanol and acetate was allowed (Figure 3A), and a specific growth rate of 0.60 h^{-1} when this was not allowed (Figure 3B). In the first case, ethanol was preferentially oxidized in the cytoplasm with $NADP^+$ as electron acceptor and cytoplasmic acetate was consumed through the TCA cycle. In the second case, ethanol could only be oxidized in the periplasm with quinone as electron acceptor, whereas periplasmic acetate was secreted. The amount of biomass formed per unit of ethanol consumed was higher in the first

case than in the second case; in the latter case, ethanol could only be used as energy source and not as carbon source.

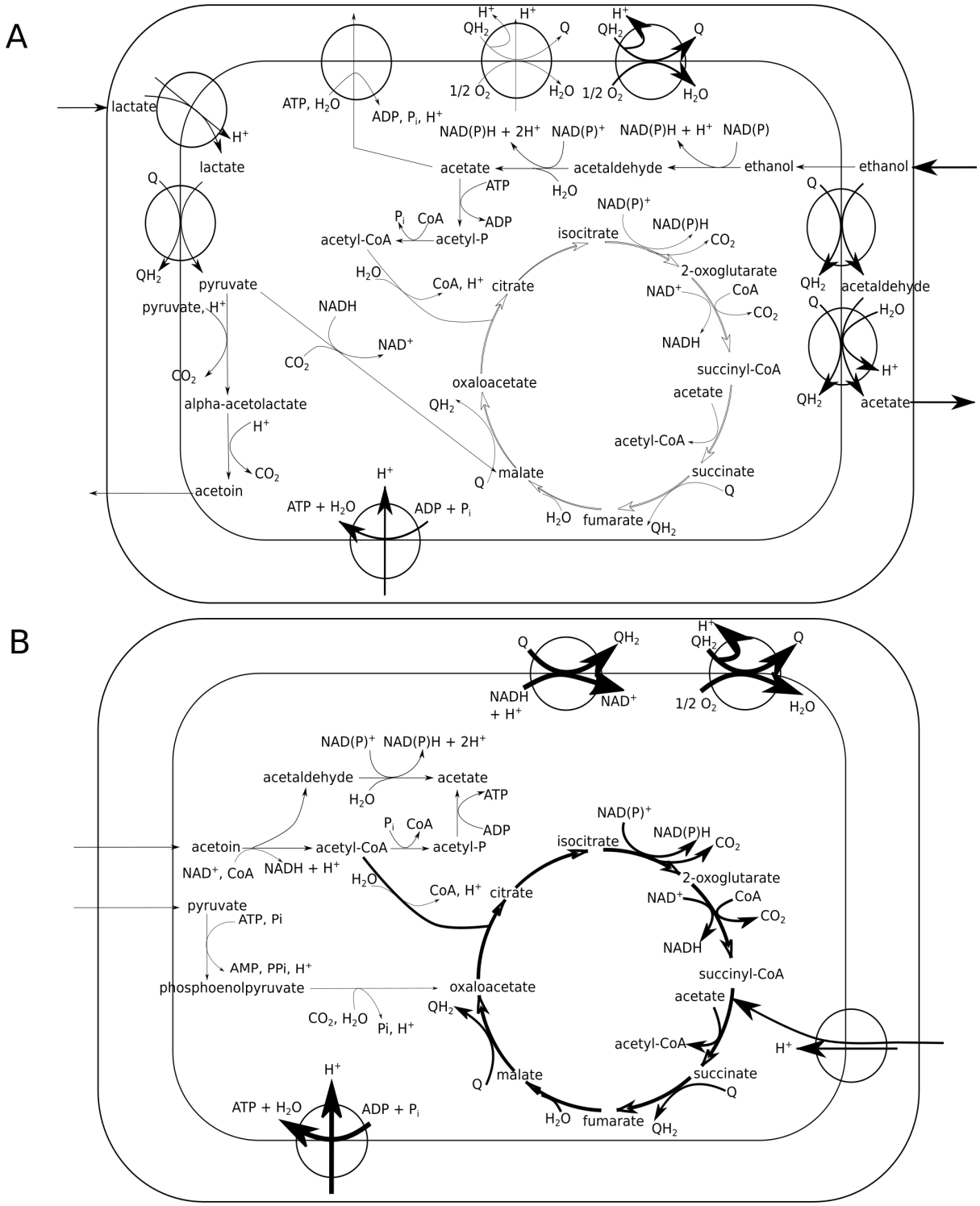


Figure 2. Parsimonious FBA solution of iAp386B454 in the two phases of growth of *Acetobacter pasteurianus* 386B under cocoa fermentation conditions. The resulting flux distributions were mapped onto the *A. pasteurianus* 386B metabolic network. The arrow thickness is proportional to the predicted reaction flux values. These were normalized with respect to the maximal flux value obtained for both growth phases. A, first growth phase of *A. pasteurianus* 386B; B, second growth phase of *A. pasteurianus* 386B.

In both cases, a saturation was obtained for the predicted specific growth rate value as a function of the lactate consumption flux. Also, proton exchange switched from periplasmic proton consumption at low ethanol consumption flux to periplasmic proton secretion at high ethanol consumption flux for a fixed lactate consumption flux. The NADPH balance at maximal co-consumption of lactate and ethanol revealed that cytoplasmic ethanol oxidation allowed NADPH production that was partly consumed by a proton-translocating $NAD(P)^+$ transhydrogenase (Figure 3A). Blocking cytoplasmic ethanol oxidation led to NADPH production by a proton-translocating $NAD(P)^+$ transhydrogenase and by a flux cycle involving the pentose phosphate pathway and upper Embden-Meyerhof-Parnas pathway (Figure 3B). The extreme case of solely periplasmic ethanol oxidation indicated another role of ethanol oxidation, namely as a potential NADPH source.

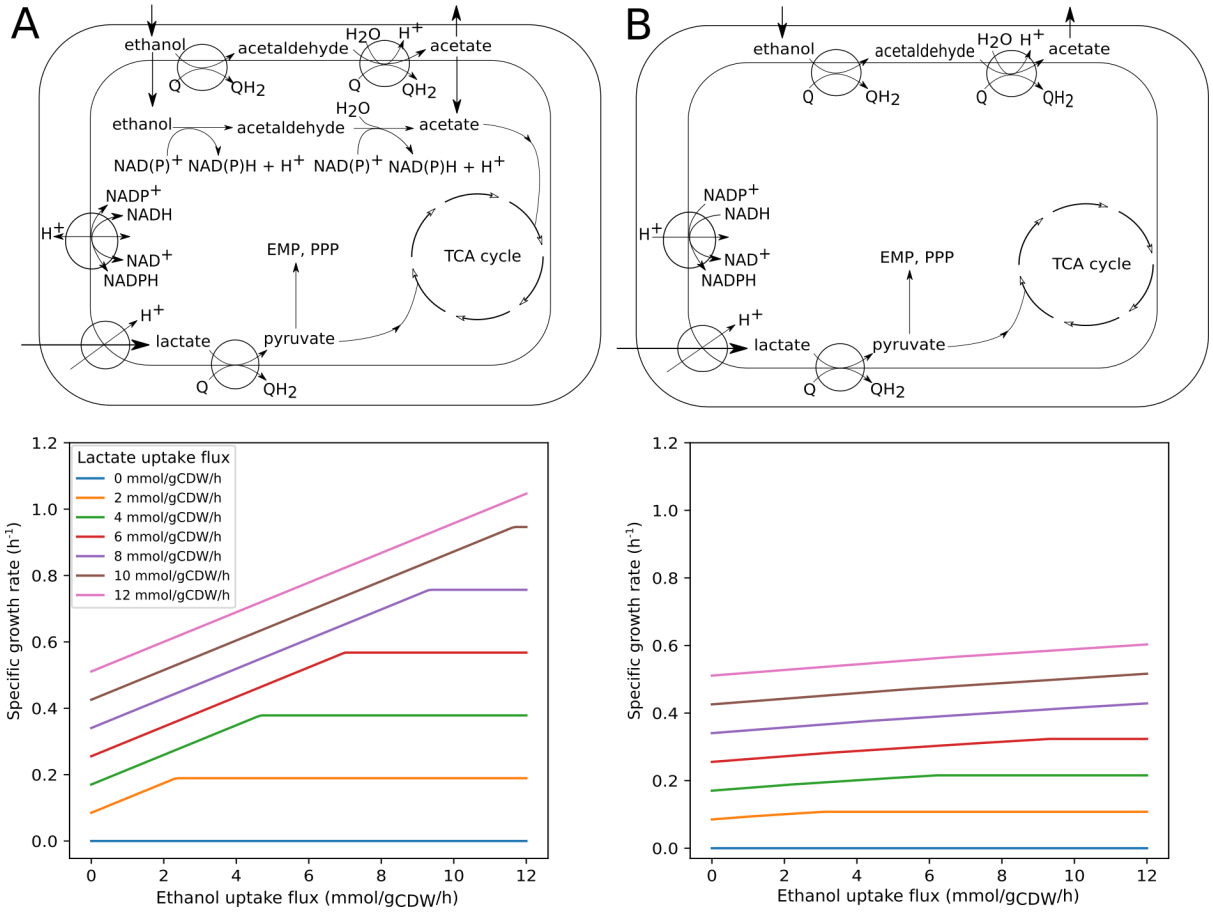


Figure 3. *In silico* analysis of the physiological role of lactate and ethanol in *Acetobacter pasteurianus* 386B. FBA simulations were performed for different ethanol consumption flux values at fixed lactate consumption flux values. A, unconstrained iAp386B454 model; B, constrained iAp386B454 model enforcing only periplasmic ethanol oxidation.

Varying the cytoplasmic ethanol uptake fraction for iAp386B454 and computing the optimal ethanol to lactate consumption flux ratio, leading to a maximal biomass production, revealed that the combination of low lactate consumption flux and low cytoplasmic ethanol oxidation

flux led to a peak in the optimal ethanol to lactate consumption flux ratio (Figure 4). At a high lactate consumption flux, the optimal ethanol to lactate consumption flux ratio was independent of the amount of ethanol that could be oxidized in the cytoplasm. This indicated the importance of reaction compartmentalization in the GEM and its influence on the predicted metabolic flux distribution. It also further pinpointed the role of ethanol to relieve lactate as a carbon, energy, and NADPH source.

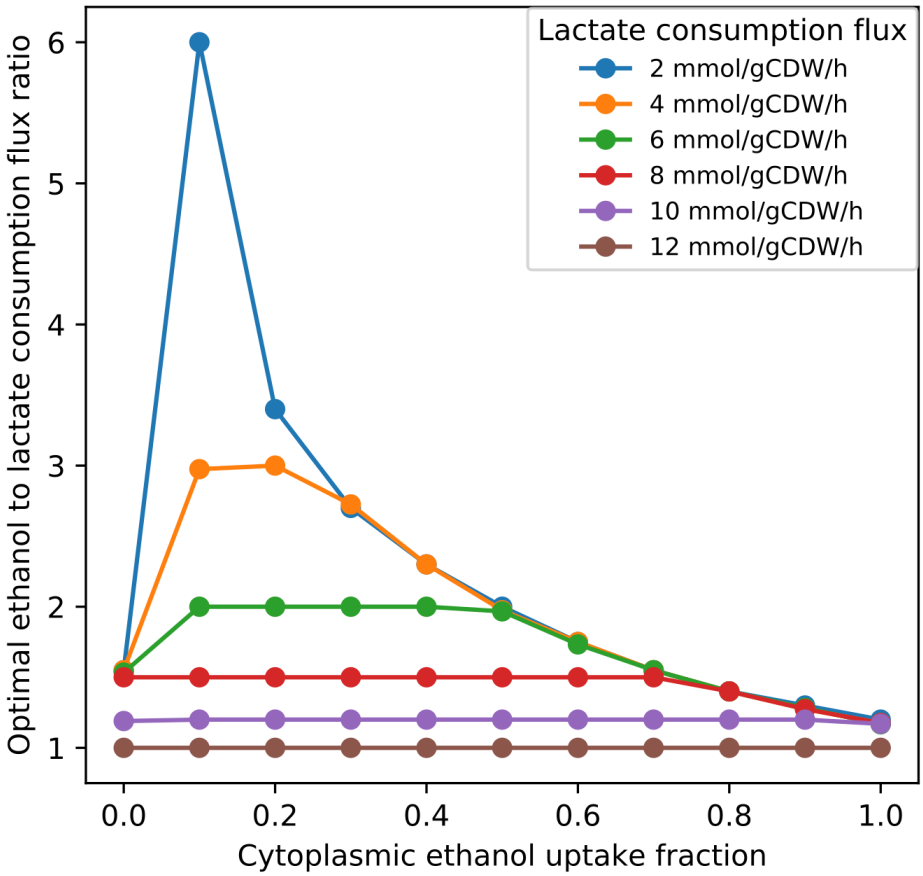


Figure 4. Influence of the cytoplasmic ethanol uptake fraction on the optimal ethanol to lactate consumption flux ratio for the growth of *Acetobacter pasteurianus* 386B, using iAp386B454. The optimal ethanol to lactate consumption flux ratio depicts the ratio of the ethanol to lactate consumption fluxes for which biomass formation was maximal, given a certain fraction of ethanol consumed in the cytoplasm.

3.3 Theoretical energy production of *A. pasteurianus* 386B

FBA was performed to calculate the maximal energy production in the form of ATP from different substrates and through different respiratory pathways, using iAp386Benergy (Table 2).

Table 2. *In silico* simulation of different metabolic modes and resulting energetic outcomes of *Acetobacter pasteurianus* 386B using different substrates. Different respiratory chains were defined using iAp386Benergy. The resulting NADH, ubiquinol-9, and ATP yields are listed. The data from a previous study on *A. pasteurianus* NCC 316 (Adler et al., 2014) are added to allow a comparison with the data obtained during the current study. TCA, tricarboxylic acid cycle; CO_2, carbon dioxide.

Mode of metabolism	Key enzymes			NDH type I + UOX ba_3	NDH type I + UOX *bd*-type	NDH type II + UOX ba_3	NDH type II + UOX *bd*-type	*A. pasteurianus* NCC 316 (Adler et al., 2014)
		NADH yield (mol/mol substrate)	Ubiquinol-9 yield (mol/mol substrate)	ATP yield (mol/mol substrate)				
Ethanol - acetic acid (intracellular)	NAD-dependent oxidation	2	2	4	3	2	1	None
Ethanol - acetic acid (periplasmatic)	Membrane-bound dehydrogenases	0	2	2	1	2	1	1
Ethanol – CO_2	TCA cycle	4	6	10	7	6	3	2
Acetic acid – CO_2	TCA cycle	2	4	6	4	4	2	1
Lactic acid - acetic acid	Pyruvate decarboxylase	1	2	3	2	2	1	1
Lactic acid – CO_2	Pyruvate dehydrogenase, acetate kinase	3	6	10	7	7	4	4
Lactic acid - acetoin	Acetolactate synthase	0	1	1	0.5	1	0.5	0.5
Acetoin – CO_2	Acetoin dehydrogenase, acetate kinase	6	10	17	12	11	6	5

Overall, the most optimal respiratory chain for maximal ATP formation in all metabolic modes considered was composed of NDH type I and UOX ba_3. The least optimal respiratory chain in all metabolic modes was composed of NDH type II and UOX *bd*-type. The two remaining respiratory chains (NDH type I + UOX *bd*-type and NDH type II + UOX ba_3) had intermediate ATP yields. The specific value depended on the type of substrate and the respiratory pathway used, which specified the production flux of NADH and ubiquinol in each metabolic mode. For ethanol, intracellular oxidation of ethanol was energetically more favorable than periplasmic oxidation and was maximal at complete oxidation to carbon dioxide. Oxidation of

acetic acid to carbon dioxide was less efficient than oxidation of ethanol to carbon dioxide, due to a lower amount of NADH produced. For the oxidation of lactic acid to carbon dioxide and of acetoin to carbon dioxide, an extra ATP was produced via substrate-level phosphorylation by acetate kinase. Oxidation of acetoin to carbon dioxide was the energetically most efficient metabolic mode for all simulations. Acetoin dehydrogenase, which oxidizes acetoin to form acetyl-CoA and acetaldehyde, did not lead to direct loss of carbon dioxide but increased the flux through the TCA cycle, thereby increasing the production flux of NADH and ubiquinol. The resulting stoichiometric proton pumping efficiency (H^+/O) and energetic efficiency (P/O) parameters were computed for each respiratory chain and metabolic mode considered (Table 3). This analysis enabled to show the influence of different metabolic modes and respiratory chains on the resulting energetic parameters.

Table 3. *In silico* simulation of different metabolic modes and the resulting energetic outcomes of *Acetobacter pasteurianus* 368B using different substrates, in view of the stoichiometric proton pumping efficiency (H^+/O) and energetic efficiency (P/O) of the different respiratory chains for each metabolic mode. CO_2, carbon dioxide.

Mode of metabolism	NDH type I + UOX *ba*$_3$		NDH type I + UOX *bd*-type		NDH type II + UOX *ba*$_3$		NDH type II + UOX *bd*-type	
	H^+/O	P/O	H^+/O	P/O	H^+/O	P/O	H^+/O	P/O
Ethanol - acetic acid (intracellular)	8.0	2.0	6.0	1.5	4.0	1.0	2.0	0.5
Ethanol - acetic acid (periplasmatic)	4.0	1.0	2.0	0.5	4.0	1.0	2.0	0.5
Ethanol – CO_2	6.7	1.7	4.7	1.2	4.0	1.0	2.0	0.5
Acetic acid – CO_2	6.0	1.5	4.0	1.0	4.0	1.0	2.0	0.5
Lactic acid - acetic acid	6.0	1.5	4.0	1.0	4.0	1.0	2.0	0.5
Lactic acid – CO_2	6.0	1.5	4.0	1.0	4.0	1.0	2.0	0.5
Lactic acid - acetoin	4.0	1.0	2.0	0.5	4.0	1.0	2.0	0.5
Acetoin – CO_2	6.4	1.6	4.4	1.1	4.0	1.0	2.0	0.5

4 Discussion

The GEM iAp386B454 and the derived iAp386Bcore and iAp386Benergy models allowed to investigate the metabolic properties of *A. pasteurianus* 386B under different *in silico* conditions, mimicking cocoa fermentation conditions and investigating the physiological role of the different carbon and energy sources.

In CPSM containing lactic acid, ethanol, and mannitol as carbon sources, *A. pasteurianus* 386B preferentially consumes lactic acid and ethanol at a relatively high rate, producing acetoin and acetate (Moens *et al.*, 2014). A similar result has been obtained for *A. pasteurianus* NCC 316 in CPSM containing lactic acid and ethanol, for which in addition it has been shown that

ethanol is primarily oxidized in the periplasm (Adler *et al*., 2014). In the case of acetate, the parsimonious solution of pFBA related to periplasmic ethanol oxidation reflected the adaptation of *A. pasteurianus* 386B to high ethanol concentrations, *i.e.,* oxidation with a restricted set of enzymes. In the case of acetoin, however, simulating this growth phase when using consumption flux constraints only did not lead to its secretion. Therefore, an additional set of production flux constraints had to be added to the GEM to be able to predict a metabolic flux distribution that was conform this prior knowledge. FBA predicted a relatively low flux of the TCA cycle, mostly originating from lactate, producing the necessary biomass precursors, and energy generation by periplasmic ethanol oxidation. The cytoplasmic ethanol uptake fraction of 16% predicted with pFBA was close to the value of 14% obtained using ^{13}C-metabolic flux analysis for *A. pasteurianus* NCC 316 (Adler *et al*., 2014). Alternatively, the addition of intracellular reaction constraints to limit the cytoplasmic ethanol uptake fraction or the TCA cycle flux would give similar results, but these constraints are more difficult to estimate experimentally. Thus, *a posteriori* maximization of the biomass reaction, including consumption flux constraints for lactate and ethanol and at least a production flux constraint for acetoin, allowed to predict the metabolic phenotype of *A. pasteurianus* NCC 316 under CPSM conditions using the *A. pasteurianus* 386B GEM.

In-depth investigation of the contribution of lactate and ethanol to biomass formation in the first growth phase revealed their role as potential source of carbon, energy, and NADPH. Constraining the amount of ethanol that could be oxidized in the cytoplasm led to a peak in the optimal ethanol to lactate ratio at low lactate consumption flux values. This was in line with previous results that indicate a high stoichiometric consumption of ethanol compared to lactate and the underlying functional separation of ethanol and lactate metabolism, whereby ethanol is mostly oxidized in the periplasm and lactate is oxidized in the cytoplasm (Adler *et al*., 2014; Moens *et al*., 2014). The low flux of the TCA cycle forced the model to produce NADPH via a proton-translocating NAD(P)$^+$ transhydrogenase. The downregulation of the TCA cycle in the presence of ethanol has already been suggested for *A. pasteurianus* and *A. aceti* (Sakurai *et al*., 2011; Adler *et al*., 2014). Thus, in the presence of ethanol, the contribution of lactate to biomass formation would be limited due to the overflow metabolism of acetoin. In FBA terms, this corresponded to a relatively low lactate consumption flux, leading to a high ethanol to lactate ratio to maximize growth. Downregulating the TCA cycle in the presence of ethanol may thus represent an adaptation to a rapid consumption of high quantities of ethanol, resulting in acetic acid accumulation in the growth medium. This strategy entails carbon loss, but it may be effective in an environment that is rich in carbohydrates, which could be useful in certain environments, such as spoiled wine, from which *Acetobacter* spp. are commonly isolated (Deppenmeier and Ehrenreich, 2009; Barata *et al*., 2012).

Since the functional separation of ethanol and lactate metabolism entailed *in silico* NADPH production by a proton-translocating NAD(P)$^+$ transhydrogenase, the preferred *in vivo* reaction directionality of the enzyme under this condition may be NADPH production with concurrent usage of the proton motive force. This is in contrast with previous hypotheses (Illeghems *et al*., 2013; Adler *et al*., 2014), but is consistent with the physiological function of this enzyme in *E.*

coli and other microorganisms (Sauer *et al*., 2004; Spaans *et al*., 2015). Furthermore, the cytoplasmic ethanol dehydrogenase of *A. pasteurianus* NBRC 101655 has been described as being strictly NAD-dependent (Chinnawirotpisan *et al*., 2003). As a hypothetical source of NADPH, *in vivo* preference of NAD-dependent ethanol oxidation may further support our hypothesis.

A different flux distribution was found when simulating the second growth phase of *A. pasteurianus* 386B. Here, consumption of acetoin, acetate, and pyruvate, led to an increase in the oxygen consumption flux and a decrease in the specific growth rate, compared to the first growth phase. A higher proportion of acetate and acetyl-CoA led to an increased flux of the TCA cycle, which resulted in a higher production flux of the reduced co-factors NAD(P)H and ubiquinol. The flux obtained through a lower efficient respiratory chain, consisting of NDH type II and UOX *bd*-type, as well as the activity of an ATP hydrolysis reaction indicated dissipation of the potential redox energy. The respiratory chain activity thus shifted from ubiquinol oxidation in the first phase of growth, to an increase in NADH oxidation for the second phase. Because of the nature of the carbon sources in the second growth phase of *A. pasteurianus* 386B, the model predicted a necessary higher oxygen consumption flux for growth to be optimal. The importance of oxygenation for acetate removal has been shown for *A. aceti*, in which the role of a modified TCA cycle in AAB as an acetic acid resistance mechanism has been exemplified (Mullins *et al*., 2008). Thus, in its second growth phase, *A. pasteurianus* 386B coupled acetate removal with energy production.

Using iAp386Benergy, ATP yields of different respiratory pathways as a function of different metabolic modes were calculated in an automated and transparent way. The simulations indicated that the most optimal respiratory chain was composed of NDH type I and UOX ba_3. Consistent with the optimality principle, this respiratory chain was also predicted by FBA using iAp386B454, applying lactate and ethanol consumption flux constraints solely. Different ATP yields as a function of respiratory pathways and substrates may reflect the possible adaptation of *A. pasteurianus* 386B in terms of energy generation to different environmental conditions. Comparing these results with the ones obtained previously (those mentioned in Table 2), it could be concluded that the previous results were actually a special case, for which a specific respiratory chain was assumed. Adler *et al.* (2014) have used a P/O value of 0.5 to calculate the ATP equivalents resulting from different modes of metabolism. In terms of respiratory chain enzymes, this value corresponds best to a respiratory chain consisting of NDH type II and UOX *bd*-type. However, even in this case the ATP yield obtained differed for the oxidation of ethanol and acetate to carbon dioxide. This is because Adler *et al.* (2014) have assumed production of acetyl-CoA by an ATP-consuming acetyl-CoA synthase, but FBA predicted a balanced modified TCA cycle. The same authors have proposed that the two-step oxidation of lactate and ethanol *via* acetoin and acetate, respectively, to carbon dioxide was energetically not so different from a one-step oxidation to carbon dioxide (Adler *et al*., 2014). This metabolic strategy seems adequate in a competing environment, where optimization of speed rather than yield is more favorable (Deppenmeier and Ehrenreich, 2009). However, the energetic difference that it entails may have been underestimated if the respiratory chain used by *A.*

pasteurianus in general would switch in between growth phases, as predicted with FBA for iAp386B454.

Of interest were the differences between the experimental measurements of the H^+/O parameter and the computed ones using iAp386Benergy. In *A. pasteurianus* LMG 1365 (now reclassified as *A. peroxydans*), the H^+/O measured is equal to 1.9 when acetate is used as electron donor (Luttik et al., 1997). *In silico* acetate oxidation to carbon dioxide using iAp386Benergy indicated a H^+/O range of 2 to 6, depending on the respiratory chain defined. For *G. oxydans*, a H^+/O measurement was performed with mannitol as electron donor, which effectively measured the H^+/O of the terminal cytochrome bo_3 (ba_3) oxidase. Whereas a theoretical value of 4 was expected, a much lower value of 1.26 was obtained (Richhardt *et al*., 2013). The occurrence of a low H^+/O ratio in combination with a high oxidase activity has been described as being typical for AAB terminal oxidases and is probably related to their evolutionary adaptation to substrate-rich habitats (Matsutani *et al*., 2014). Thus, the theoretical H^+/O values to predict *in silico* ATP yields with iAp386Benergy on different substrates were most probably an overestimation of their values *in vivo*, but nonetheless provided useful as guidelines for energy generation by *A. pasteurianus* in general, since experimental values are hard to obtain.

In view of the cocoa fermentation process, acetic acid could be considered as both a friend and foe. It is necessary for the formation of cocoa bean flavor precursors inside the beans but in case there is an excess, more effort is needed in the chocolate production process to remove acetic acid, if not it may lead to unpleasant sourness in the chocolates produced. The current *A. pasteurianus* 386B GEM predicted that a second growth phase, mainly centered around acetate overoxidation, would lead to a shift in the respiratory chain used and a simultaneous increase in the oxygen consumption flux. Thus, acetic acid depletion in the cocoa pulp-bean mass may require extensive oxygenation during the stage of the fermentation process in which AAB thrive.

5 Conclusion

In silico constraint-based modeling using a GEM of *A. pasteurianus* 386B allowed to gain more insights into the metabolic adaptations that allow this strain to thrive under cocoa fermentation conditions.

It was shown that *A. pasteurianus* 386B experienced two phases of growth under cocoa fermentation conditions. In the first phase, a high stoichiometric ratio of ethanol to lactate consumption occurred, leading to the formation of acetate and acetoin, with implications on the TCA cycle flux and the respiratory chain usage. Constraint-based analysis of lactate and ethanol consumption fluxes indicated that lactate, but also ethanol, played a role of potential source of carbon, energy, and NADPH. In the second phase, acetate and acetoin were consumed but another carbon source, presumably pyruvate, with a higher carbon content had to be co-consumed to allow growth. The increase of the TCA flux led to a switch in respiratory chain usage and increased oxygen consumption.

The presence of a proton-translocating $NAD(P)^+$ transhydrogenase may have important metabolic consequences, forming a bridge between energy metabolism and biosynthesis pathways. Since there was *a priori* no reason to add an irreversibility constraint to this reaction, the results obtained highlighted the potential of constraint-based modeling to decipher the physiological role of metabolic enzymes and pathways, thereby possibly enabling an improved functional annotation of these enzymes. The physiological function of the proton-translocating $NAD(P)^+$ transhydrogenase in *A. pasteurianus* 386B thus remains to be discovered.

Overall, the insights obtained during this study will lead to a better application of *A. pasteurianus* 386B to fulfill its role in controlling cocoa fermentation processes.

Chapter 6

Dynamic modeling of *Acetobacter pasteurianus* 386B and *Acetobacter ghanensis* LMG 23848^T, two candidate functional starter cultures for cocoa bean fermentation processes

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

Supplementary material for this Chapter is available in the Annexes.

Summary

Dynamic modeling of monoculture fermentations with ordinary differential equation and dynamic flux balance analysis models provides quantitative estimations of microbial growth parameter values and metabolic flux distributions. *Acetobacter pasteurianus* 386B and *Acetobacter ghanensis* LMG 23848^T were previously selected as candidate functional starter cultures for the cocoa bean fermentation process. Genome-scale metabolic models have previously been made for these two strains and flux balance analysis simulations performed. The current study provides the necessary dynamic modeling tools to obtain quantitative estimates of their microbial growth parameter values, allowing comparison to other candidate functional starter cultures and related microorganisms. Due to a limitation in the available experimental data for *A. pasteurianus* 386B, monoculture experimental data of another strain, *A. pasteurianus* NCC 316 was used, grown in similar cocoa fermentation conditions. The diauxic growth curve of these strains found in cocoa fermentation conditions could successfully be fitted using additional assumptions, most notably an increase in the growth-associated maintenance, which are in need of experimental verification. Dynamic FBA results revealed a different survival strategy of *A. pasteurianus* NCC 316 and *A. ghanensis* LMG 23848^T by modulating the lactate consumption rate. Also, a possible and hitherto unknown role of the cytochrome bc_1 complex in *A. ghanensis* LMG 23848^T could be the oxidation of reduced cytochrome *c* produced by a cytochrome *c*-dependent D-lactate dehydrogenase.

1 Introduction

Cocoa bean fermentation is a prerequisite for the production of cocoa-derived products, among which chocolate (De Vuyst and Weckx, 2016a). So far, this fermentation process is still a spontaneous one, whereby inoculation of the cocoa pulp-bean mass relies on the environmental microbiota at the location of fermentation. A successive growth of enterobacteria, yeasts, lactic acid bacteria, and acetic acid bacteria (AAB) occurs commonly. The different process-technological parameters at hand, such as the fermentation method used (*e.g.*, heaps, boxes, platforms, and baskets), the fermentation duration, and the mixing of the fermenting cocoa pulp-bean mass enabling air to ingress, are expected to influence the cocoa bean fermentation outcome (Camu *et al*., 2008; Aprotosoaie *et al*., 2016; De Vuyst and Weckx, 2016a).

To better control and potentially accelerate the cocoa bean fermentation process, the use of starter cultures has been proposed (Schwan, 1998; de Melo Pereira *et al*., 2016; De Vuyst and Weckx, 2016a). Therefore, *in situ* experimental studies have been performed, to assess the influence of the addition of a starter culture on the autochthonous community, with divergent results. In some cases, the starter culture does not prevail during fermentation, and the influence on the resulting metabolite dynamics is not always clear (Schwan, 1998; Lefeber *et al*., 2012; Crafack *et al*., 2013; Sandhya *et al*., 2016; da Veiga Moreira *et al*., 2017; Ho *et al*., 2018). Species of the AAB genus *Acetobacter*, mostly *Acetobacter pasteurianus*, are almost invariably used in starter culture cocktails for controlled cocoa bean fermentation processes (Schwan, 1998; Lefeber *et al*., 2012; Crafack *et al*., 2013; Sandhya *et al*., 2016; da Veiga Moreira *et al*., 2017; Ho *et al*., 2018). They combine interesting features, since they co-consume lactic acid and ethanol, produced by lactic acid bacteria and yeast species, respectively, leading to the production of mainly acetic acid. In addition, they are relatively tolerant to high temperatures and low pH values, both occurring during fermentation of the cocoa pulp-bean mass (De Vuyst and Weckx, 2016a).

Two strains of *Acetobacter* species isolated from cocoa bean fermentation processes performed in Ghana, namely *A. pasteurianus* 386B and *Acetobacter ghanensis* LMG 23848^T, have shown to possess superior ethanol consumption rates under cocoa pulp-simulating growth conditions (Camu *et al*., 2007; Lefeber *et al*., 2010; Moens *et al*., 2014). Whereas *A. pasteurianus* 386B combines this trait with a rapid lactate consumption rate, the consumption of lactate is relatively slower for *A. ghanensis* LMG 23848^T. Moreover, both strains show a second phase of growth after ethanol depletion, in which acetic acid is overoxidized to carbon dioxide and water (Moens *et al*., 2014). Similar growth dynamics have been found for *A. pasteurianus* NCC 316 when grown in the same cocoa pulp simulation medium as mentioned above (Adler *et al*., 2014). This type of growth dynamics, whereby accumulated product metabolites are consumed at a later phase of growth, is similar to a diauxic growth shift (Siegal, 2015). It has previously been found for the growth of *Escherichia coli* on glucose, whereby acetic acid is first secreted and later consumed. This co-occurring shift in the metabolic flux distribution of *E. coli* has been investigated using a combination of experimental and modeling techniques (Enjalbert *et al*., 2015; Succurro *et al*., 2019). However, the mechanistic principles behind a diauxic shift are much less known. For *Acetobacter* species, a hypothesis has been proposed, stating that

ethanol would act as a signal molecule and that its presence would repress the tricarboxylic acid (TCA) cycle, in turn inhibiting the oxidation of acetate (Saeki *et al*., 1999; Adler *et al*., 2014).

Predictive models that would combine microbial growth properties and physicochemical dynamics could be very useful to gain insight into processing factors that may improve the cocoa bean fermentation process. A first conceptual and technical hurdle consists of the construction of models that capture sufficient mechanistic aspects to describe the fermentation dynamics. Phenomenological growth models, using ordinary differential equations (ODEs), are used to describe the microbial population and metabolites dynamics of a fermentation process, whereby typically the microbial growth parameters of the models are microorganism- and growth condition-specific, providing a coarse-grained but overall view of the microorganism's growth characteristics (Peleg and Corradini, 2011). Quantitative modeling of the cocoa bean fermentation process using phenomenological growth models has been attempted to gain insight into the mechanisms that are at hand during the fermentation process. By fitting the microbial growth parameters of the model to experimental data of *in vivo* cocoa bean fermentation processes, information regarding the fermentation conditions can be extracted and interpreted (Moreno-Zambrano *et al*., 2018).

A second application hurdle is that, although crucial to release the cocoa flavor potential, the microbial fermentation process has only an indirect impact on the final flavor of cocoa-derived products (Santander Muñoz *et al*., 2019). This is because the fermentation process takes place early in the cocoa value chain, and thus its influence may be limited by the additional steps needed to obtain the final cocoa-derived products, among which drying, roasting, and conching (Barišić *et al*., 2019). However, what seems to be clear is that a certain amount of microbial fermentation products, among which acetic acid is assumed to be the most important one, is needed to accelerate the death of the cocoa bean embryo, enabling the formation of the necessary cocoa flavor precursor molecules inside the cocoa beans (De Vuyst and Weckx, 2016b).

To gain more insight into the influence of varying metabolite concentrations in the growth medium on the specific growth rate and metabolic flux distribution of a microbial cell population, dynamic FBA (dFBA) has been proposed, which couples flux balance analysis (FBA) predictions of a genome-scale metabolic model (GEM) to a set of ODEs describing the bacterial cell population and metabolite dynamics (Mahadevan *et al*., 2002). The consumption and production flux values predicted by FBA are fed into the relevant ODEs, and feedback of the metabolite dynamics on the consumption flux values is governed by a modulating equation, for example the Michaelis-Menten equation of enzyme kinetics.

In this chapter, a first step is taken to model the population and metabolite dynamics of the candidate functional starter cultures *A. pasteurianus* 386B and *A. ghanensis* LMG 23848^{T}, using experimental data from monoculture fermentation experiments performed in a cocoa pulp simulation medium. The use of phenomenological models will allow to estimate basic

microbial growth parameter values, and the use of dFBA simulations will allow to gain insight into the possible metabolic changes during the growth of the AAB cell population.

2 Materials and methods

2.1 Strains and data sets

Three strains were investigated throughout this study, namely *A. pasteurianus* NCC 316, a cheese isolate (Adler *et al*., 2014), and *A. pasteurianus* 386B and *A. ghanensis* LMG 23848^T, two candidate functional starter culture strains for cocoa bean fermentation processes.

Concerning data sets for *A. pasteurianus*, a first series of experimental data (further referred to as data set 1) were taken from a publication describing the population and extracellular metabolite dynamics of *A. pasteurianus* NCC 316 in a cocoa pulp simulation medium for AAB (CPSM-AAB; Adler *et al*., 2014). This growth medium contained lactic acid and ethanol as the main carbon sources. Similar experimental data (further referred to as data set 2) were taken from a publication describing the growth of the cocoa bean fermentation isolate *A. pasteurianus* 386B in CPSM-AAB supplemented with mannitol (Moens *et al*., 2014). Biomass data in grams of cell dry weight (g_{CDW}) units were directly obtained from data set 1. For data set 2, a conversion from measured colony forming units (CFU) into grams of cell dry weight was required, using a theoretical conversion factor of 0.28 pg_{CDW}/CFU. This conversion factor has been previously estimated based on the geometry of *Acetobacter* cells and the density of a single *E. coli* cell (Moreno-Zambrano *et al*., 2018). For initial parameter fitting, the experimental data points of the first 24 h of microbial growth were used. The mannitol dynamics were not considered for data set 2, since the consumption of mannitol was minor compared to the other carbon sources added and it was restricted to the second phase of growth (Moens *et al*., 2014).

Concerning data sets for *A. ghanensis* LMG 23848^T, a single experimental data set was taken from a publication describing the population and extracellular metabolite dynamics of *A. ghanensis* DSM 18895, which is the same strain as *A. ghanensis* LMG 23848^T, in CPSM-AAB (Adler *et al*., 2014). This experimental data set was chosen as it contained population dynamics data expressed in units of g_{CDW}.

2.2 Modeling of the population and metabolite dynamics of *Acetobacter pasteurianus* NCC 316 and *A. pasteurianus* 386B under cocoa pulp-simulating conditions

ODE-based models were constructed that could describe the population dynamics in both experimental data sets mentioned above. The models consisted of a set of equations describing the metabolite dynamics, coupled to two different growth models, namely the Verhulst growth model (Equation 1) and the Monod growth model. For the Monod growth model, two modeling concepts were explored, *i.e.*, one that had a separate, additive growth term for each carbon source (Equation 2), and one that combined growth on the different substrates into a unified

equation (Equation 3), the latter being referred to as the modified version (Supplementary File 1).

$$dX / dt = \mu * (1 - X / K) * X \quad (1)$$

$$dX/dt = (\mu_{Etoh} * Etoh/(Etoh + K_{Etoh}) + \mu_{La} * La/(La + K_{La}) + \mu_{Ac} * Ac/(Ac + K_{Ac}) * K_{inh}/(Etoh + K_{inh}) + \mu_{Aco} * Aco/(Aco + K_{Aco}) * K_{inh}/(Etoh + K_{inh}) + \mu_{Pyr} * Pyr/(Pyr + K_{Pyr}) * K_{inh}/(Etoh + K_{inh})) * X \quad (2)$$

$$dX/dt = \mu_1 * La/(La + K_{La}) * (w + (1 - w) * Etoh/(Etoh + K_{Etoh})) * X + \mu_2 * ((Ac/(Ac + K_{Ac}) + Aco/(Aco + K_{Aco}) + Pyr/(Pyr + K_{Pyr})) * K_{inh}/(Etoh + K_{inh}) * X \quad (3)$$

The parameters of the Verhulst growth model (Equation 1) included the maximum specific growth rate (μ) and the carrying capacity (K). For the Monod growth models (Equation 2), the specific growth rate parameters (μ_{Etoh}, μ_{La}, μ_{Ac}, $\mu_{Aco,}$ and μ_{Pyr}) were defined for each substrate separately, as well as the Monod constants (K_{Etoh}, K_{La}, K_{Ac}, K_{Aco}, and K_{Pyr}). A Monod inhibition constant (K_{inh}) was defined to simulate the switch in between the phases of growth. For the modified Monod growth model (Equation 3), a maximum specific growth rate parameter was defined for each phase of growth separately (μ_1 and μ_2). The basal growth weight parameter (w) defined the importance of the 'boosting' resource ethanol on the specific growth rate (Kettle et al., 2018).

The product yields were defined by assuming separated metabolite conversion pathways of ethanol into acetate and of lactate into pyruvate and acetoin (Adler *et al*., 2014). The ODE models were solved using the SciPy integrate package with the VODE numerical integrator (Jones *et al*., 2001). The time-step was set to 0.1 h. Parameter fitting was performed combining a manual and computational fitting process, using the SciPy fmin method. Parameters were first fitted to the experimental data of the first 24 h of growth, after which the whole growth curve was fitted.

Subsequently, the fitted parameter values of the ODE models were used to define the constraints applied to a GEM of *A. pasteurianus* 386B (Pelicaen *et al*., 2019). For dFBA modeling of *A. pasteurianus* NCC 316, the GEM of *A. pasteurianus* 386B (iAp386B454) was used with the assumption that the reactions present in this GEM would be present in *A. pasteurianus* NCC 316 as well, for which no genome sequence was publicly available. The dFBA model included a set of ODEs describing the population and metabolite dynamics, of which the parameter values, namely specific growth rate, consumption fluxes, and production fluxes, were predicted by the GEM (Figure 1A, Supplementary File 2). Numerical integration of the ODEs in the dFBA model was performed using the Euler integration method with a time-step of 0.1 h and non-negative concentrations of extracellular metabolites were enforced. Intermediary FBA results were saved for further analysis. A constraint screening analysis was

performed to identify a suitable set of constraints needed to describe the population and metabolite dynamics in both phases of growth.

For the ODE models, a Michaelis-Menten inhibition term dependent on the ethanol concentration was defined to model the whole growth curve using the Monod growth model. This enabled to activate the substrate consumption terms of the second phase of growth, namely acetate, acetoin, and pyruvate, when ethanol was consumed. A similar strategy was applied for dFBA, for which an ethanol Michaelis-Menten inhibition term was used to increase the range of the constraint bounds of the first phase of growth, and an ethanol Michaelis-Menten activation term was used to restrain the range of the constraint bounds of the second phase of growth. This allowed to simulate the metabolic switch of co-consumption of lactate and ethanol with the production of acetate, acetoin, and pyruvate in the first phase of growth to the consumption of the metabolites produced in the second phase of growth.

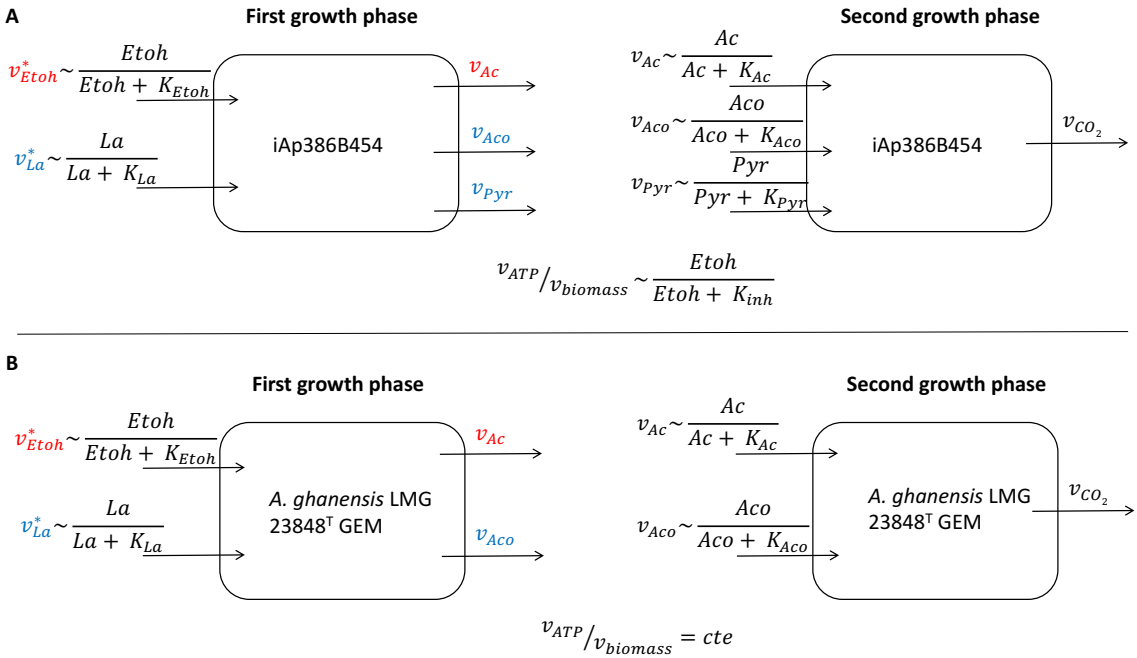


Figure 1. Schematic representation of the dynamic flux balance analysis (dFBA) simulation for the (A) *Acetobacter pasteurianus* 386B genome-scale metabolic model (iAp386B454) and (B) *Acetobacter ghanensis* LMG 23848^T genome-scale metabolic model (Chapter 4). The genome-scale metabolic model constraints are updated at each time-step of the simulation by Michaelis-Menten terms that are dependent on the metabolite concentrations in the *in silico* growth medium. Flux constraints are shown for metabolites for which experimental data points were available. The active substrate consumption flux constraints are shown for each phase of growth. Production flux constraints dependent on substrate consumption flux constraints are shown in color and substrate consumption flux constraints that are dependent on each other are shown with an asterisk (*). The bounds of these constraints are controlled by Michaelis-Menten equations depending on ethanol, allowing to simulate the switch between the two phases of growth (Supplementary Files 2 and 4). The growth-associated maintenance constraint is shown beneath each sub-figure where v_{ATP} represents an ATP hydrolysis flux.

2.3 Modeling of the population and metabolite dynamics of *Acetobacter ghanensis* LMG 23848^T under cocoa pulp-simulating conditions

ODE modeling was performed using the Verhulst growth model, the Monod growth model, and the modified version of the Monod model as described above for *A. pasteurianus* (see also Supplementary File 3). The time-step was set to 0.1 h. Parameter fitting was performed manually. Parameters were first fitted to the experimental data of the first 24 h of growth, after which the whole growth curve was fitted.

For dFBA modeling of *A. ghanensis* LMG 23848^T, the fitted parameter values of the ODE models were used to define the constraints applied to the *A. ghanensis* LMG 23848^T GEM. The same set of constraints as for *A. pasteurianus* was used to model the first growth phase, except for the pyruvate production constraint (Figure 1B). Numerical integration of the ODEs in the dFBA model was performed using the Euler integration method with a time-step of 0.1 h non-negative concentrations of extracellular metabolites were enforced. The intermediary FBA results were saved for further analysis. The metabolic switch was simulated in a similar way as for *A. pasteurianus*, using a Michaelis-Menten equation depending on ethanol, to control the constraint bounds of both phases of growth (Supplementary File 4).

3 Results

3.1 Modeling of the population and metabolite dynamics of *Acetobacter pasteurianus* NCC 316 and *A. pasteurianus* 386B under cocoa pulp-simulating conditions using a genome-scale metabolic model

The computational fitting process provided estimates of the specific growth rates but failed to fit the biomass yields, to which negative values were assigned frequently (Table 1). Thus, manual fitting was required to obtain a better fit of the experimental data sets (Figures 2 and 3).

The carrying capacity parameter estimated from the first 24 h of growth of data set 1 (*A. pasteurianus* NCC 316) was almost tenfold higher than that for data set 2 (*A. pasteurianus* 386B). This means that the amount of biomass formed, for a similar concentration of carbon source provided, seemed to be much higher in data set 1 than that in data set 2. The conversion of colony forming units to biomass for data set 2 probably explained this difference, indicating an underestimation of the amount of biomass corresponding to the CFU counts. Product yield parameters were the same for all growth models used, thus reflecting a fixed substrate-to-product stoichiometry. The fact that these parameter values were below 1.0 indicated that an amount of carbon was used for biomass formation or lost in the form of carbon dioxide. Biomass yields (inverse of substrate-to-biomass yields) were much lower for data set 2 than for data set 1, following directly from the difference in carrying capacities, since the total substrate consumption was similar for both data sets.

The biomass yield parameter values were altered for the modified Monod growth model to prevent lactate from being totally consumed prior to ethanol. Moreover, the specific growth rate was lower in the second growth phase. The Monod growth constants were kept deliberately low to allow linking the parameter values of ODE and dFBA models, and to prevent the growth constants to interfere with the fitted specific growth rates and biomass yield parameters, the latter contributing to the necessary information to constrain the GEM. A low Monod inhibition constant value described best the metabolic shift after the first phase of growth. Converting the biomass yield coefficients of *A. pasteurianus* NCC 316 for ethanol and lactate (Table 1) by multiplication with their molar masses led to a value of 6.91 g/g_{CDW} and 3.60 g/g_{CDW}, respectively.

Table 1. Fitted ODE model parameters of the AAB strains investigated (modified Monod growth model). The carrying capacity was obtained by fitting the Verhulst growth model. NA, not available.

Model parameter	*A. pasteurianus*		*A. ghanensis*
	NCC 316 (data set 1)	386B (data set 2)	LMG 23848^T
Carrying capacity (mg/l)	2000.000	235.800	1000.000
Maximum specific growth rate phase 1 (h^{-1})	0.600	0.270	0.600
Maximum specific growth rate phase 2 (h^{-1})	0.004	0.005	0.006
Basal growth weight	0.400	0.400	0.020
Ethanol-to-biomass yield (mmol/mg)	0.150	1.600	0.250
Lactate-to-biomass yield (mmol/mg)	0.040	0.600	0.050
Acetate-to-biomass yield (mmol/mg)	0.070	1.600	0.100
Acetoin-to-biomass yield (mmol/mg)	0.010	0.600	0.020
Pyruvate-to-biomass yield (mmol/mg)	0.010	NA	NA
Acetate-from-ethanol yield (mmol/mmol)	0.700	0.900	0.700
Acetoin-from-lactate yield (mmol/mmol)	0.300	0.600	0.300
Pyruvate-from-lactate yield (mmol/mmol)	0.150	NA	NA
Monod constant of ethanol (mM)	0.100	0.100	0.010
Monod constant of lactate (mM)	0.100	0.100	10.000
Monod constant of acetate (mM)	0.100	0.100	0.100
Monod constant of acetoin (mM)	0.100	0.100	0.100
Monod constant of pyruvate (mM)	0.100	NA	NA
Monod inhibition constant of ethanol (mM)	0.100	0.100	0.100

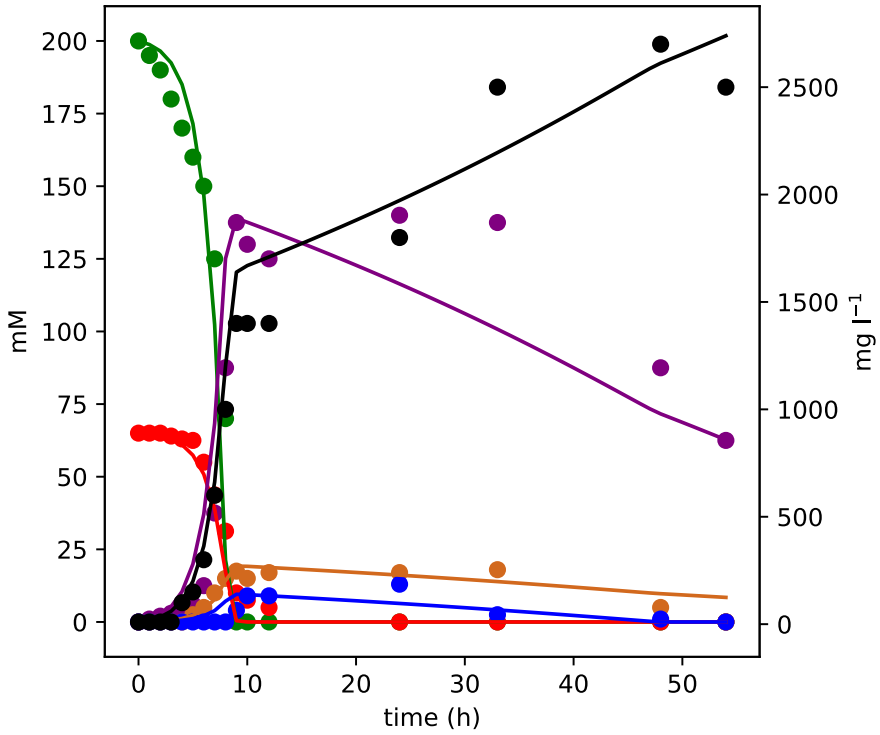


Figure 2. Population (biomass in black) and metabolite dynamics fit (ethanol in green, lactate in red, acetate in purple, acetoin in brown, and pyruvate in blue) of *Acetobacter pasteurianus* NCC 316 (data set 1) under cocoa pulp-simulating conditions with the modified Monod growth model. Dots represent experimental data points, lines represent dynamic model fittings.

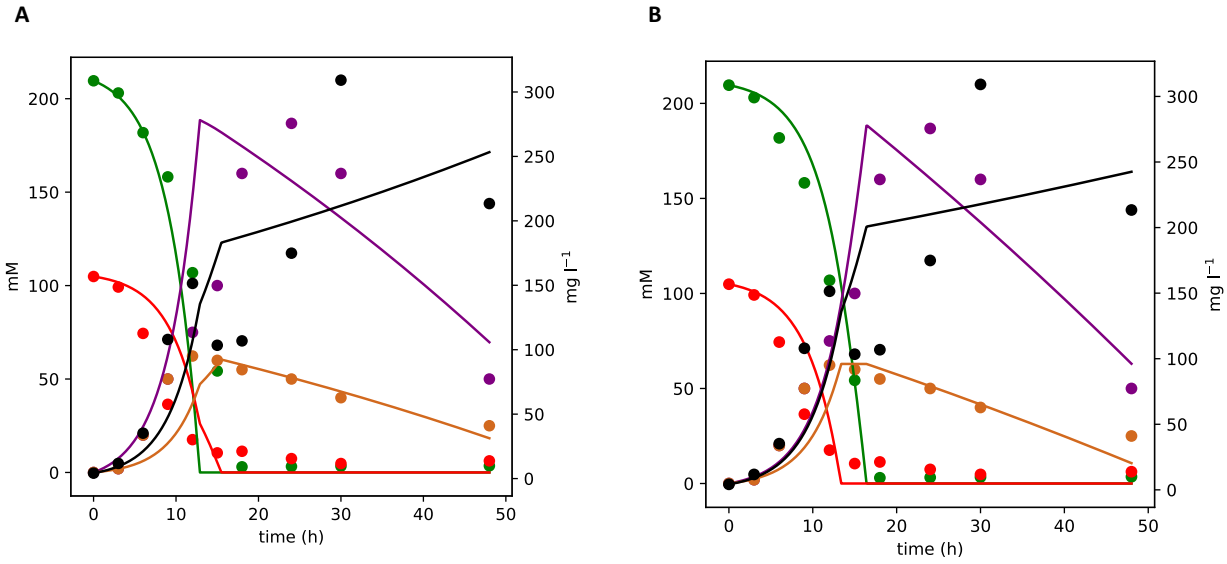


Figure 3. Population (biomass in black) and metabolite dynamics fit (ethanol in green, lactate in red, acetate in purple, and acetoin in brown) of *Acetobacter pasteurianus* 386B (data set 2) under cocoa pulp-simulating conditions with the modified Monod growth model (A) and the Monod growth model (B). Dots represent experimental data points, lines represent dynamic model fittings.

Different constraints were screened to identify a suitable set for which FBA predicted results that were concurrent with the experimentally obtained population and metabolite dynamics in the first and second phase of growth. For the first phase of growth, the production yield constraints on acetoin and pyruvate were essential for their production. This, in turn, revealed the need to add a constraint defining a higher ethanol than lactate uptake flux. Dynamic FBA resulted in a too high acetate production. Adding a production yield constraint on acetate led to a slowdown in metabolite consumption dynamics because of a lowered specific growth rate. Thus, in addition to a production yield constraint on acetate, the consumption of an additional substrate had to be assumed to fit the dFBA results with the experimental data.

Given the fact that, next to their biosynthesis, also a small consumption of L-aspartate, L-alanine, and L-glutamate was reported in the study of Adler *et al*. (2014), for dFBA, L-aspartate was included in the set of consumable substrates, as it is produced from oxaloacetate, an anaplerotic source of the TCA cycle. For the second phase of growth, it was essential to constrain the maximum uptake flux of acetate, acetoin, and pyruvate. Additionally, an ATP maintenance constraint had to be added to decrease the specific growth rate to obtain the experimentally obtained population dynamics in the second phase of growth.

Modeling the whole growth curve of *A. pasteurianus* NCC 316 (data set 1) with dFBA, using the iAp386B454 GEM, resulted in a predicted average specific growth rate of 0.592 h^{-1} for the first phase of growth stage and 0.004 h^{-1} for the second phase of growth (Figure 4). These values corresponded best to the fitted maximum specific growth rates of the ODE model using the modified Monod growth model. The average ethanol and lactate consumption fluxes corresponded to 57.40 and 17.96 mmol/g_{CDW}/h, respectively. These values were between 2 and 3 times higher as those used to constrain the model in Chapter 5, and corresponded to a total consumption flux of 168.74 C-mmol/g_{CDW}/h, which was almost 3 times higher as the total consumption flux of 60 C-mmol/g_{CDW}/h assumed in Chapter 5, based on the glucose consumption flux of *E. coli* (Varma et al., 1993). The resulting oxygen consumption flux was constrained to 90 mmol/g_{CDW}/h, allowing a good fit to the experimental data. Ethanol was oxidized in the periplasm. The metabolic switch delineating both growth phases occurred after 9.3 h of fermentation, when ethanol was completely consumed. Thus, in this time range, the average biomass yields were 10.28 g_{CDW}/mol for ethanol and 32.85 g_{CDW}/mol for lactate. The average consumption rates of ethanol and lactate were 21.51 mM/h and 6.73 mM/h, respectively. In the second phase of growth, the substrates lactate, acetate, acetoin, pyruvate, and L-aspartate were consumed slowly. Parsimonious dFBA predicted a switch in respiratory chain enzyme usage from NADH dehydrogenase (NDH) type II and ubiquinol oxidase (UOX) *bd* activity for the first phase of growth to NDH type I and UOX ba_3 for the second phase of growth. Especially for the second phase of growth, these results differed from the results obtained in Chapter 5, and this was due to the addition of the growth-associated maintenance constraint to simulate the population dynamics of the second phase of growth with dFBA.

3.2 Modeling of the population and metabolite dynamics of *Acetobacter ghanensis* LMG 23848^T under cocoa pulp-simulating conditions using a genome-scale metabolic model

The development of a GEM of *A. ghanensis* LMG 23848^T indicated the need to co-consume L-asparagine and L-proline to allow *in silico* growth with glucose as the sole carbon source (Chapter 4). However, an FBA simulation under cocoa pulp-simulating conditions with lactate and ethanol as the sole carbon sources did not lead to *in silico* growth. Debugging the GEM was possible by making the reaction catalyzed by quinol-cytochrome-c reductase (cytochrome bc_1 complex; EC 7.1.1.8) reversible. The reason for this lies in the metabolic reconstruction of the *A. ghanensis* LMG 23848^T GEM, in which the cytochrome bc_1 complex was the sole reaction that potentially could re-oxidize reduced cytochrome *c*.

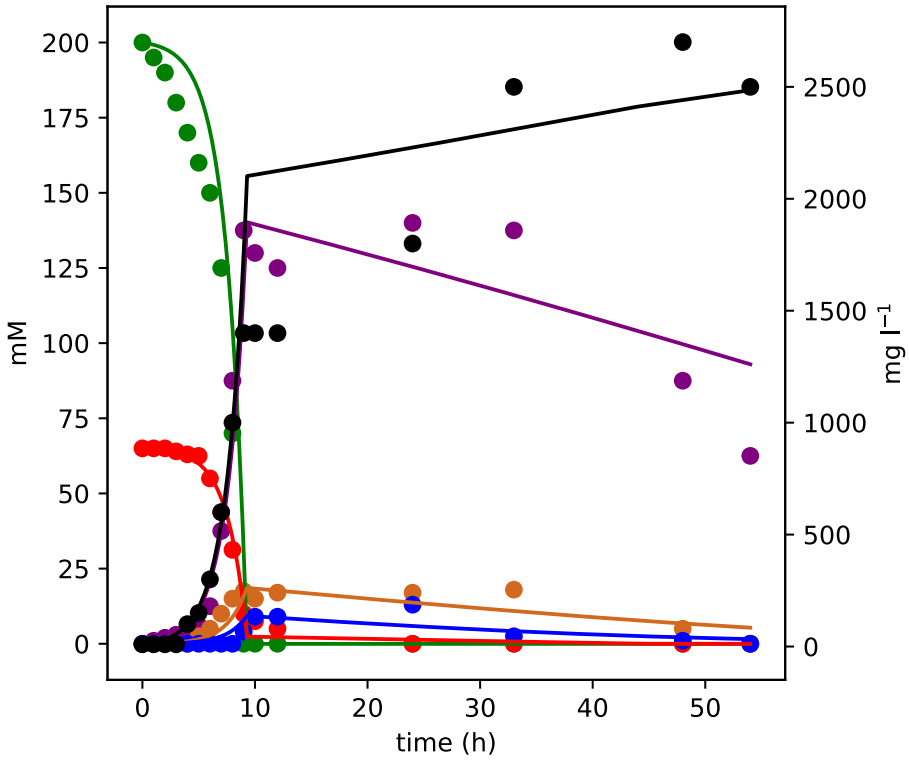


Figure 4. Population (biomass in black) and metabolite dynamics fit (ethanol in green, lactate in red, acetate in purple, acetoin in brown, and pyruvate in blue) of *Acetobacter pasteurianus* NCC 316 (data set 1) under cocoa pulp-simulating conditions with dynamic flux balance analysis. Dots represent experimental data points, lines represent dynamic model fittings.

The *A. ghanensis* LMG 23848^T genome only encoded a cytochrome *c*-dependent D-lactate dehydrogenase (EC 1.1.2.4), whereas a quinone-dependent (EC 1.1.5.12) enzyme was absent. Similarly to the modeling of the two *A. pasteurianus* strains, the population and metabolite dynamics of *A. ghanensis* LMG 23848^T were modeled for the two phases of growth separately. For the first phase of growth, fitting the experimental data with the Verhulst model indicated a

carrying capacity of 1000 mg_{CDW}/l and an ethanol-to-biomass yield of 0.25 mmol/mg_{CDW}. This indicated that *A. ghanensis* LMG 23848^T formed less biomass but had a higher ethanol consumption flux, since the fitted maximal specific growth rate was the same for both species. The use of the Monod growth model did not give a good fit to the experimental data. In this modeling framework, two separate growth terms were assumed, each depending on lactate or ethanol consumption separately. The dynamics obtained with this model did not correspond to the experimentally obtained co-consumption of lactate and ethanol and the abrupt decrease in population growth rate after ethanol was consumed. Alternatively, the modified Monod growth model assuming an unified growth term that took into account the metabolite concentrations of both lactate and ethanol allowed to fit the experimental data by putting a high weight to ethanol as boosting resource for population growth (Table 1 and Figure 5). Converting the biomass yield coefficients for ethanol and lactate (Table 1) by multiplication with their molar masses, led to a value of 11.51 g/g_{CDW} and 4.50 g/g_{CDW}, respectively.

The fitted ethanol-to-biomass yield parameter in the modified Monod growth model was used in the dFBA simulation and allowed to simulate the slower lactate consumption rate as found for *A. ghanensis* LMG 23848^T, compared to *A. pasteurianus* NCC 316 and *A. pasteurianus* 386B. A growth-associated maintenance constraint had to be added to decrease the predicted specific growth rate and to be able to fit the population dynamics data. The assumptions taken to fit the simulated dFBA data revealed an excess of energy produced by *A. ghanensis* LMG 23848^T under cocoa pulp-simulating conditions. For the second phase of growth, a different maximal lactate consumption flux had to be assumed compared to the first phase of growth, allowing the slow consumption of the acetate and acetoin produced. Fitting the acetate consumption flux and including the same growth-associated maintenance constraint as for the first phase of growth were sufficient to fit the experimental data of *A. ghanensis* LMG 23848^T in the second phase of growth.

Modeling the whole growth curve of *A. ghanensis* LMG 23848^T with dFBA resulted in a metabolic switch after 9.7 h of fermentation, when ethanol was completely consumed (Figure 6). In this time range, the average specific growth rate was 0.50 h^{-1}. The average ethanol consumption flux value was 90.45 mmol/g_{CDW}/h and the average lactate consumption flux value was 18.16 mmol/g_{CDW}/h. Thus, the average biomass yields from ethanol and lactate were 5.53 g_{CDW}/mol ethanol and 27.53 g_{CDW}/mol lactate, respectively. The average consumption rates of ethanol and lactate were 20.62 mM/h and 4.19 mM/h, respectively. Ethanol was oxidized in the cytoplasm to increase the NADH production flux and uphold the necessary energy requirements imposed by the growth-associated maintenance constraint. The resulting average oxygen consumption flux was 150.50 mmol/g_{CDW}/h. The predicted specific growth rate dropped from an average value of 0.50 h^{-1} in the first phase of growth to 0.012 h^{-1} in the second phase of growth. In this phase, the lactate consumption flux became the growth-limiting constraint. Finally, no switch in respiratory chain was predicted.

4 Discussion

Modeling the population and metabolite dynamics of monoculture fermentations of *A. pasteurianus* NCC 316 and *A. pasteurianus* 386B, as well as *A. ghanensis* LMG 23848^T, using a dynamic modeling strategy combining experimental data and phenomenological models, allowed to estimate their microbial growth parameter values and metabolic flux distributions under cocoa fermentation conditions, making use of a cocoa pulp simulation medium for AAB. In contrast to FBA, dynamic modeling allowed to quantitatively estimate the microbial growth parameter values of candidate functional starter cultures and to compare these values to the ones of related microbial strains.

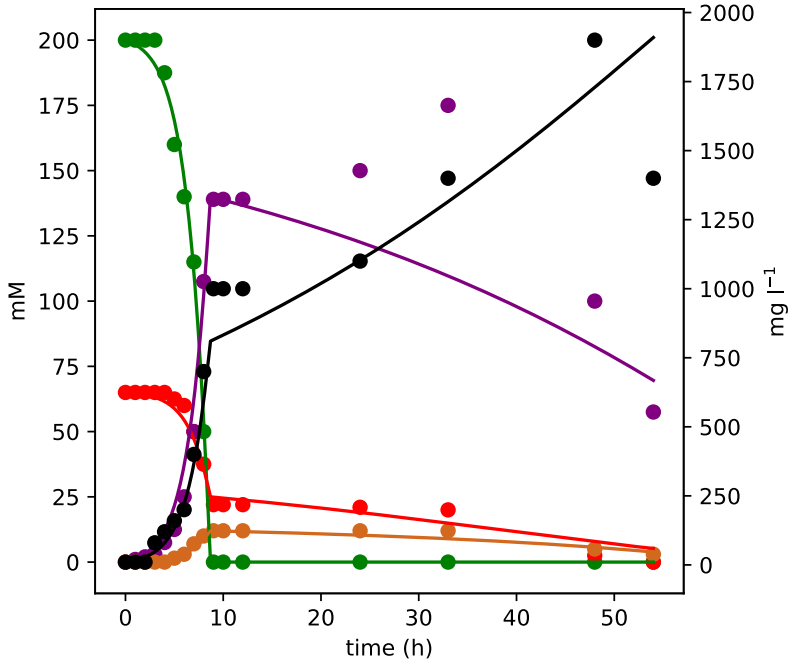


Figure 5. Population (biomass in black) and metabolite dynamics fit (ethanol in green, lactate in red, acetate in purple, and acetoin in brown) of *Acetobacter ghanensis* LMG 23848^T under cocoa pulp-simulating conditions with the modified Monod growth model. Dots represent experimental data points, lines represent dynamic model fittings.

Analysis of the microbial dynamics of *A. pasteurianus* NCC 316, using the iAp386B454 GEM, showed a low average biomass yield per mole ethanol. For *A. ghanensis* LMG 23848^T, this value was almost halved. Both values were lower than for *Acetobacter peroxydans* LMG 1635, growing in a chemostat with ethanol as the sole carbon source and combining a high substrate uptake rate with a low specific growth rate, resulting in a higher but still relatively low biomass yield of 13.1 g_{CDW}/mol ethanol and a P/O ratio of 0.5 (Luttik *et al.*, 1997). These authors argued that the low P/O ratio of *A. peroxydans* LMG 1635 could explain its low biomass yield on ethanol, since theoretical estimation of the ATP requirement for growth gave a similar outcome as for other microorganisms (*e.g.*, *E. coli* and *Candida utilis*).

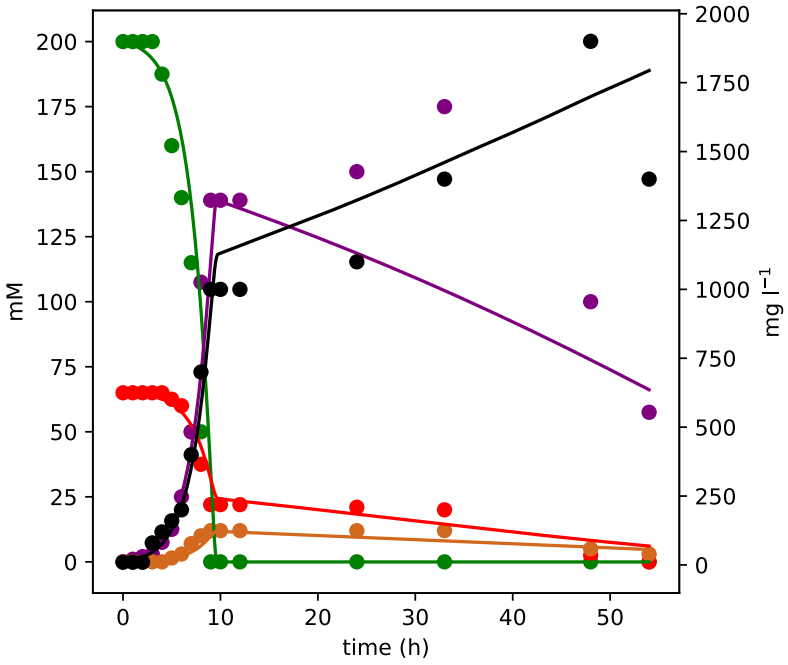


Figure 6. Population (biomass in black) and metabolite dynamics fit (ethanol in green, lactate in red, acetate in purple, and acetoin in brown) of *Acetobacter ghanensis* LMG 23848^T under cocoa pulp-simulating conditions with dynamic flux balance analysis. Dots represent experimental data points, lines represent dynamic model fittings.

For the two *A. pasteurianus* strains considered in the present study, the co-consumption of ethanol with lactate, whereby only a minor part of ethanol was used in biomass assimilation pathways and a significant overflow of acetoin occurred, may explain the resulting low biomass yields. Accordingly, to compare the overall biomass yield of *A. pasteurianus* with other microorganisms, a biomass yield defined per amount of C-mol substrate consumed seemed to be more adequate. Co-consumption of ethanol and lactate led to a biomass yield of 3.50 g_{CDW}/C-mol for *A. pasteurianus* NCC 316, and 2.12 g_{CDW}/C-mol for *A. ghanensis* LMG 23848^T. For *A. peroxydans* LMG 1635, this is 6.55 g_{CDW}/C-mol ethanol (Luttik *et al.*, 1997), and for *E. coli* this is 15.67 g_{CDW}/C-mol glucose based on simulated data (Varma *et al.*, 1993). This indicated the considerable lower biomass yields for the *Acetobacter* strains under study.

Although at first sight the inclusion of a reversible cytochrome bc_1 complex in the *A. ghanensis* LMG 23848^T GEM seemed to be merely a technical solution to obtain steady-state in the FBA simulation with lactate and ethanol as the sole carbon sources, it might also reveal biological insights. The absence of a quinol-dependent D-lactate dehydrogenase in the *A. ghanensis* LMG 23848^T genome raised the question as to how this species generated its energy with a respiratory chain that relies on quinol-dependent terminal oxidases. The reversibility of the cytochrome bc_1 complex provided a simple solution for this problem, as reduced cytochrome *c* generated by the cytochrome *c*-dependent D-lactate dehydrogenase could be used to reduce quinone to quinol, with concomitant usage of a proton motive force. The reversibility of the cytochrome bc_1 complex has been shown previously (Sarewicz and Osyczka, 2015). The

physiological function of the cytochrome bc_1 complex in AAB is currently unknown and has been a long-standing question in the AAB research community (Sakurai *et al*., 2011; Hanke *et al*., 2012).

The dynamics of *A. ghanensis* LMG 23848^T mainly differ from *A. pasteurianus* 386B in their slower consumption of lactate, namely 2.35 mM/h compared to 8.43 mM/h, respectively (Moens *et al*., 2014). However, the experimental data collected by Adler *et al*. (2014) shed a slightly different perspective, whereby the lactate consumption of *A. ghanensis* LMG 23848^T drops abruptly only after ethanol is rapidly consumed. Fitting population and metabolite dynamics data with dFBA allowed to obtain additional insights into the benefits of slowing down the lactate consumption rate, or, alternatively, increasing the ethanol consumption rate. Based on the FBA simulations presented here and in Chapters 4 and 5, a C3 compound was necessary for the growth of *A. pasteurianus* 386B and *A. ghanensis* LMG 23848^T, as both *Acetobacter* species did not possess the glyoxylate cycle. Thus, from an evolutionary perspective, since no pyruvate was found in their growth medium, the slower consumption of lactate by *A. ghanensis* LMG 23848^T may be a survival strategy to allow the consumption of acetic acid after ethanol depletion. This strategy is different from the overflow of pyruvate by *A. pasteurianus* NCC 316 found under the same cocoa pulp-simulating conditions (Adler *et al*., 2014).

The higher average ethanol consumption flux value for *A. ghanensis* LMG 23848^T compared to *A. pasteurianus* NCC 316 indicated that, in theory, an increase of the biomass of *A. ghanensis* LMG 23848^T could lead to superior ethanol consumption rates. As ethanol fermentation by yeast species during cocoa bean fermentation is a major driver to rapidly consume the carbohydrates available in the cocoa pulp (De Vuyst and Weckx, 2016a), a similar increase in acetic acid productivity is of interest to accelerate the cocoa bean fermentation process by the use of appropriate starter cultures.

ODE modeling allowed to estimate the biomass yield coefficients for ethanol and lactate for *A. pasteurianus* NCC 316, *A. pasteurianus* 386B, and *A. ghanensis* LMG 23848^T. Previously, these parameters have been fitted using experimental data of *in vivo* cocoa bean fermentation processes, leading to high values, ranging between 170 and 1298 g/g_{CDW} for ethanol, and 63 and 1929 g/g_{CDW} for lactate, where in each case the biomass represented the total AAB population (Moreno-Zambrano *et al*., 2018). *In vitro* monoculture fermentation studies of *Acetobacter* species reported biomass yields for ethanol between 7.55 and 166.67 g/g_{CDW}, and for lactate 7.90 g/g_{CDW} (Bar *et al*., 1987; Salgueiro Machado *et al*., 1995; Gómez and Cantero, 1998). The discrepancy between the *in vitro* and *in vivo* parameter estimations has been attributed to the noise in the experimental data and to the possible contribution of evaporation during the cocoa bean fermentation processes, inflating the substrate-to-biomass yields, which was not included in the model (Moreno-Zambrano *et al*., 2018). However, the inflation of these yield parameters could alternatively be due to the conversion of the CFU counts measured to g_{CDW} values, using the theoretically estimated conversion factor. This was exemplified in this chapter, in which the fitted biomass yield parameters for data set 2 (population dynamics in CFU/ml) for *A. pasteurianus* 386B were much lower than in the case of data set 1 (population

dynamics in g_{CDW}/ml) for *A. pasteurianus* NCC 316. Therefore, theoretical conversion of CFU counts to biomass values could have led to biomass estimates that were too low in the study of Moreno-Zambrano et al. (2018). Inflated substrate-to-biomass yield parameters would then be a logical consequence to explain the metabolite dynamics found in the cocoa bean fermentation process.

These findings indicated that selective plating, which leads to population abundances expressed in CFU per mass or volume of sample, used to discriminate between the different subpopulations in mixed-culture fermentation processes, may represent a significant experimental hurdle to obtain accurate biomass (expressed in g_{CDW}) estimations from these subpopulations. These estimations are primordial for dFBA simulations, since FBA simulates population growth in terms of biomass production (De Jong *et al*., 2017). The scientific community is currently advancing the dFBA framework toward microbial community models (Gottstein *et al*., 2016; Ang *et al*., 2018). Also for the cocoa bean fermentation process, this approach seems to be an attractive framework to better understand the ecological interactions between microbial species and the influence of species diversity, and to explore possible biochemical interactions between microorganisms and the cocoa pulp-bean mass (Moreno-Zambrano *et al*., 2018). However, to become predictive, such models need first to be validated with experimental data. The currently available experimental data of *in vivo* cocoa bean fermentation processes regarding population dynamics are mostly expressed in CFU counts, as biomass determinations are difficult to carry out. Therefore, these data sources do not seem adequate to validate community dFBA simulations.

5 Conclusion

Dynamic FBA can be used to estimate the consumption and production flux values of a microbial cell population in monoculture using a data set experimentally obtained, to describe the population and metabolite dynamics. Since dFBA combines elements from ODE, capturing the environmental population and metabolite dynamics, and FBA modeling, capturing the metabolic potential and constraints of the microorganism considered, it is better positioned to model *in vivo* fermentation processes than each of those frameworks separately. However, this also means that it still suffers from the inherent limitations from both frameworks, including the assumption of a pseudo steady-state of the metabolic network at each time-step considered. The *in silico* data generated contain – sometimes indirectly – important basic microbial growth parameters, of which the values can be compared to other microorganisms in the literature. However, to obtain a good fit to the data, assumptions need to be made as to which constraints have to be added to the GEMs. Moreover, the many parameter values that are present in the dynamic equations pose an additional challenge for the modeling process. In this chapter, a combination of ODE and dFBA modeling was proposed as well as the separation of the growth curves of *A. pasteurianus* NCC 316 and *A. ghanensis* LMG 23848^T in two phases of growth. This allowed to better organize the data and simplify the fitting procedure, thereby allowing a better overall fit for both species. Dynamic FBA allowed to gain more insights into the monoculture fermentation processes of *A. pasteurianus* NCC 316 and *A. ghanensis* LMG 23848^T under cocoa pulp-simulating conditions. For *A. pasteurianus* NCC 316 in particular, it

revealed a better estimation of the *in vivo* ethanol and lactate consumption fluxes. For *A. ghanensis* LMG 23848^T, a new physiological role of the cytochrome bc_1 complex was proposed. Although the metabolic phenotype of the strains belonging to these *Acetobacter* species is similar, they seem to employ different metabolic strategies to continue their growth after ethanol depletion. Finally, the simulations performed provided mechanistic hypotheses to explain the metabolic phenotypes of *A. pasteurianus* NCC 316 and *A. ghanensis* LMG 23848^T, which could be tested further using other experimental techniques, for instance, transcriptomics and enzyme assays. At the same time, the simulations highlighted the importance of accurate experimental data and meta-data to be able to simulate the population and metabolite dynamics of growing microbial cell populations in the most accurate way.

Chapter 7

Genome-scale metabolic modeling of the candidate functional starter cultures *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 for cocoa bean fermentation processes

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

Supplementary material for this Chapter is available in the Annexes.

Summary

Successful cocoa bean fermentation processes are characterized by the presence of lactic acid bacteria (LAB), typically species of the genus *Lactobacillus,* besides yeasts and acetic acid bacteria. Two LAB candidate functional starter cultures have been proposed previously, namely *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80, because of their favorable and complementary substrate consumption and product formation profiles. Here, core genome-scale metabolic models were developed of both strains focusing on their carbon core metabolism. The metabolic reconstruction allowed a system-level analysis of the core metabolism of these strains and specification of the putative functions of different metabolic enzymes and transporters. The models were subsequently used to contextualize previously reported experimental data. This allowed to gain more insight into the possible underlying metabolic flux distributions that could explain these findings. Dynamic flux balance analysis revealed the necessary addition of a degree of freedom to the models, in particular regarding replenishment of the intracellular (2S)-2-acetolactate and pyruvate pools, to fit the experimental data. This generated the hypothesis that when growing the strains in a cocoa pulp simulation medium, the strains were most likely able to additionally consume so far undefined medium components. Finally, the necessary addition of both consumption and production constraints revealed a deviation of the growth yield optimization assumption of flux balance analysis, making an optimization of the reaction rates by these strains more probable.

1 Introduction

Lactic acid bacteria (LAB) constitute an important bacterial group occurring in spontaneous cocoa bean fermentation processes (De Vuyst and Weckx, 2016a). In contrast to acetic acid bacteria (AAB), LAB are a dominant microbial population at the earlier stage of the fermentation process, when the fermenting cocoa pulp-bean mass becomes less anaerobic because of the pectinolytic activity of the yeasts and when there are still plenty of carbohydrates available. LAB consume these carbohydrates rapidly, mainly as an energy source, producing microbial by-products such as lactic acid and acetic acid.

The presence of LAB seems ubiquitous in many spontaneous cocoa bean fermentation processes performed in different cocoa-producing countries and the dominating LAB are often (facultatively) heterofermentative species of the genus *Lactobacillus* (Camu *et al*., 2007; Meersman *et al*., 2013; Papalexandratou *et al*., 2013; Miescher Schwenninger *et al*., 2016; Visintin *et al*., 2016; Ozturk and Young, 2017). Their dominance has been ascribed to the consumption of citric acid, present in considerable amounts in the cocoa pulp (Pettipher, 1986; De Vuyst and Weckx, 2016b). By consuming citric acid, *Lactobacillus* species evade substrate competition with yeast species and its conversion into lactic acid increases the pH of the fermenting cocoa pulp-bean mass. The production of lactic acid however allows to control this pH rise, minimizing the risk of spoilage by the background microbiota (De Vuyst and Weckx, 2016b). The essentiality of the presence of LAB in successful spontaneous cocoa bean fermentation processes has however been questioned (Ho *et al*., 2015, 2018). For example, inhibition of the growth of LAB in an experimental spontaneous cocoa bean fermentation process did lead to fully fermented cocoa beans, and acceptable chocolates produced thereof (Ho *et al*., 2015). Similarly, controlled cocoa bean fermentations under aseptic conditions using a starter culture have led to good-quality chocolate when only yeasts are used to inoculate the cocoa pulp-bean mass (Ho *et al*., 2018). Both experimental studies included a controlled temperature increase under laboratory conditions.

The strains *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 have been isolated from spontaneous cocoa bean fermentation processes performed in Ghana (Camu *et al*., 2007). Subsequently, these strains have been tested and selected among a pool of isolates as candidate functional starter cultures, since they can rapidly consume glucose, fructose, and citric acid and achieve high *in vitro* cell densities. Moreover, both strains have different metabolite profiles, since *L. fermentum* 222 co-produces lactic acid, acetic acid, and mannitol, whereas *L. plantarum* 80 produces mainly lactic acid (Lefeber *et al*., 2010, 2011b). The genomes of these strains have been sequenced and functionally annotated, allowing the reconstruction of their core metabolism (Illeghems *et al*., 2015). This has confirmed the status of *L. fermentum* 222 as being strictly heterofermentative, solely encoding the genes of the phosphoketolase pathway, and *L. plantarum* 80 as being facultatively heterofermentative, encoding the genes of the Embden-Meyerhof-Parnas (EMP) pathway as well. Moreover, both genomes contain the genes necessary for the consumption of citric acid, but *L. fermentum* 22 consumed more citric acid in a cocoa pulp simulation medium (CPSM) for LAB (Lefeber *et al*. 2010, 2011b; Illeghems *et al*., 2015).

Genome-scale metabolic models (GEMs) provide a framework to obtain a system-level view of the metabolic potential of microorganisms based on their genome annotation (Thiele & Palsson, 2010). With GEMs, flux balance analysis (FBA) simulations can be performed that allow to gain more insight into which metabolic phenotypes are feasible for a microorganism, and which are not (Orth *et al*., 2010). For *L. plantarum*, a GEM has been reconstructed for the *L. plantarum* WCFS1 strain, an isolate from human saliva (Kleerebezem *et al*., 2003; Teusink *et al*., 2005, 2006). For *L. fermentum*, no GEM has been reported yet. Population and metabolite dynamics of *in vitro* fermentation experiments using a CPSM are available for both *L. fermentum* 222 and *L. plantarum* 80 (Lefeber *et al*., 2010). Additionally, *in vitro* fermentation experiments and metabolic flux analysis of two related strains, *L. fermentum* NCC 575 and *L. fermentum* NCC 2829, have been performed using the same growth medium (Adler *et al*., 2013).

The current study aimed to develop core genome-scale metabolic models (core GEMs) focusing on the carbon core metabolism of *L. fermentum* 222 and *L. plantarum* 80; and to use these models to contextualize previously reported experimental data, by estimating the metabolic flux distributions underlying these data. Additional insight into their metabolic properties will lead to better use of these strains as functional starter cultures for cocoa bean fermentation processes.

2 Materials and methods

2.1 Strains and data sets

Four strains were investigated throughout this study, namely *L. fermentum* 222 and *L. plantarum* 80, two candidate functional starter culture strains for cocoa bean fermentation processes (Lefeber *et al*., 2010, 2011b, 2012), and *L. fermentum* NCC 575 and *L. plantarum* NCC 2829, from the Nestlé Culture Collection (Lausanne, Switzerland) (Adler *et al*., 2013).

Data sets that describe the population and metabolite dynamics of *L. fermentum* 222, *L. plantarum* 80, *L. fermentum* NCC 575, and *L. plantarum* NCC 2829 were used for dynamic modeling (Lefeber *et al*., 2010; Adler *et al*., 2013). Biomass data of the first two strains that are expressed in colony forming units were converted into units of g_{CDW}, using a theoretically estimated conversion factor of 1.25 pg_{CDW}/CFU (Moreno-Zambrano *et al*., 2018). This allowed comparison between the modeling results.

2.2 Development of core genome-scale metabolic models of *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80

The protein-encoding genes of the *L. fermentum* 222 and *L. plantarum* 80 genomes, annotated using GenDB v.2.2 before (Illeghems *et al*., 2015), were re-annotated using an in-house developed bioinformatics pipeline (Pelicaen *et al*., 2019). Hereto, the National Collection for Bioinformatic Information (NCBI) RefSeq genome annotation version was taken as a reference for the re-annotation of *L. fermentum* 222 (release of March 2017, 1759 protein-encoding

genes) and *L. plantarum* 80 (release of April 2017, 2947 protein-encoding genes). The amino acid sequences of each set of protein-encoding genes (further referred to as protein sequences) were annotated using a combination of tools, namely BlastKOALA from the KEGG database (Kanehisa *et al*., 2016, 2017), the TransAAP tool to predict transport proteins (Elbourne *et al*., 2017), the subcellular localization predictor CELLO (Yu *et al*., 2004), eggNOG-mapper (Huerta-Cepas *et al*., 2017), the enzyme annotation tool PRIAM (Claudel-Renard, 2003), the RAST annotation pipeline available in the KBase platform (Overbeek *et al.*, 2014; Arkin *et al.*, 2018), and the tools embedded in InterProScan 5.22-61.0 (Jones *et al*., 2014). Furthermore, since the initial publication and release of the genome sequences of both strains, these genomes have been re-annotated by several genome annotation platforms from which additional functional annotations were retrieved. These included the Integrated Microbial Genomes platform of the Joint Genome Institute (JGI IMG) (Markowitz *et al*., 2012) and the Pathosystems Resource Integration Center (PATRIC) database (Wattam *et al*., 2017). For the Universal Protein Resource (UniProt) database (The UniProt Consortium, 2019), only functional annotation data were available for *L. fermentum* 222. Since the genomes of both strains were not present in the Carbohydrate-Active enZYmes (CAZy) database (Lombard *et al*., 2014), they were re-annotated in the current study using the dbCAN2 webserver (Zhang *et al*., 2018), of which the annotation tools rely on the CAZy database.

The genome annotation pipelines used by GenDB (Meyer, 2003), the NCBI prokaryotic genome annotation pipeline linked with RefSeq (O'Leary *et al*., 2016; Tatusova *et al*., 2016)), and PATRIC (RASTtk; Brettin *et al*., 2015) represent independent methods combining gene prediction and gene annotation. Functional annotation data of these sources were combined in in-house relational databases (MySQL), specifically designed for *L. fermentum* 222 and *L. plantarum* 80.

Comparative genomics was performed using OrthoFinder, allowing to predict orthogroups from protein-encoding genes in whole genomes (Emms and Kelly, 2015). Protein sequences of a selection of bacterial species obtained from NCBI RefSeq were compared, namely *L. fermentum* 222, *L. plantarum* 80, *L. plantarum* WCFS1, *Lactobacillus paracasei* LC2W, *Lactobacillus reuteri* JCM1112, *Bacillus subtilis* 168, and *Escherichia coli* str. K-12 substr. MG1655 (further referred to as *E. coli*). These bacterial strains, except for *L. fermentum* 222 and *L. plantarum* 80, were specifically chosen since GEMs have been reconstructed for these strains before, namely for *L. plantarum* WCFS1 (Teusink *et al*., 2005), *L. paracasei* LC2W (Xu *et al*., 2015; Wuyts *et al*., 2017), *L. reuteri* JCM1112 (Kristjansdottir *et al*., 2019), *B. subtilis* 168 (Oh *et al*., 2007; Henry *et al*., 2009), and *E. coli* (Monk *et al*., 2017). Additionally, a data set of previously functionally annotated and curated protein sequences of LAB was included in the analysis (Zheng *et al*., 2015). Finally, targeted alignment searches using blastp (Altschul *et al*., 1990) were performed with sequences from identified proteins available in the literature (Supplementary File 1).

Core GEMs were developed of *L. fermentum* 222 and *L. plantarum* 80, based on their metabolic pathways described before (Illeghems *et al*., 2015). Manual curation was performed to assign

the gene-protein-reaction associations, based on the information available in the species-specific relational databases and the comparative genomics analyses.

2.3 Genome-scale metabolic modeling of *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80

The GEMs were analyzed using the COBRApy package version 0.11.3 (Ebrahim *et al*., 2013). Three constraint-based modeling methods were used to explore the optimal solution space of the core GEMs, namely FBA to find the optimal value of an objective function taking into account a set of constraints, parsimonious FBA (pFBA) to minimize the cumulative sum of reaction fluxes using the optimal value of the objective function as an additional constraint, and flux variability analysis (FVA) to find the maximal and minimal flux values of each reaction, using a fraction of the optimal value of the objective function as an additional constraint. Regarding the latter, the loopless variant of FVA was used, which omits reaction flux cycles from the analysis, and the fraction of optimum was set to 99 %. Optimization of the core GEMs was performed using the CPLEX solver from IBM (Armonk, New York, USA). For FBA simulations, transport reactions for intracellular metabolites were included, based on available experimental and genomic evidence (Lefeber *et al*., 2010; Adler *et al*., 2013; Illeghems *et al*., 2015). Also, to simulate biomass formation, a universal biomass reaction was added (Edirisinghe *et al*., 2016; Supplementary Figure 1). Alternatively, an ATP hydrolysis reaction was used as the objective function to simulate the influence on the flux distribution when the substrates are assumed to be energy sources only. The GEMs' consumption of phosphate as phosphate source was allowed without constraint. In the case that biomass was used as the objective function, 2-oxoglutarate consumption was allowed without constraints, as no reactions producing this metabolite were included in the core GEMs. The influence of oxygen consumption was investigated by alternatively allowing and disallowing it. A constraint was also alternatively added to the transaldolase reaction (EC 2.2.1.2) to investigate the influence of blocking its flux on the resulting flux distribution. Indeed, the flux carried by the enzymes of the non-oxidative branch of the pentose phosphate pathway (PPP) could lead to a flux cycle involving metabolites from the EMP and the phosphoketolase pathway. The mapping of simulated metabolic flux data onto reconstructed metabolic pathways was visualized using Escher (King *et al*., 2015). The carbon recovery was calculated as the ratio of the sum of C-moles of product metabolites to the sum of C-moles of substrate metabolites.

Both for the *L. fermentum* 222 and the *L. plantarum* 80 core GEMs, a first series of simulations were performed whereby all possible (co-)consumptions of the substrates glucose, fructose, and citric acid were investigated. For each substrate, the maximal consumption flux was set to 10 mmol/g_{CDW}/h. In addition, the influence of oxygen consumption and the PPP constraint were assessed. A second series of simulations were performed whereby published metabolic flux analysis data of *L. fermentum* NCC 575 and *L. plantarum* NCC 2829 were used as a reference for the *L. fermentum* 222 and *L. plantarum* 80 core GEMs, respectively, since all strains co-consume glucose, fructose, and citric acid (Lefeber *et al*., 2010; Adler *et al*., 2013). For both *L. fermentum* NCC 575 and *L. plantarum* NCC 2829, lactic acid and acetic acid production data are available, whereas mannitol is only produced by the *L. fermentum* strain

only (Adler *et al*., 2013). No other secreted metabolites have been detected for both strains (Adler *et al*., 2013). The normalized flux data were classified as one set of consumption flux constraints and one set of production flux constraints. This allowed to investigate the influence of its separate usage onto the flux distribution. Also, the influence of alternative secreted metabolites that are not part of the original data sets were assessed, as well as the influence of the PPP constraint. The simulations were performed both under aerobic and anaerobic conditions.

2.4 Modeling of the bacterial cell population and metabolite dynamics of *Lactobacillus fermentum* and *Lactobacillus plantarum* under cocoa-pulp simulating conditions using a genome-scale metabolic model

Dynamic ordinary differential equation (ODE)-based models were built that could describe the experimental data sets of both species (Supplementary Files 2 - 5). The models were based on a set of equations describing the metabolite dynamics, coupled to two different growth models, namely the Verhulst growth model (Equation 1) and the Monod growth model (Equation 2). For both growth models, the metabolite dynamics were dependent on the biomass yield parameters.

$$dX\ /\ dt = \mu * (1 - X\ /\ K) * X \tag{1}$$

$$dX/_{dt} = \sum_{S_i} \mu_{S_i} * S_i/(S_i + K_{S_i}^M) * X \tag{2}$$

The parameters of the Verhulst growth model (Equation 1) include the maximum specific growth rate (μ) and the carrying capacity (K). For the Monod growth model (Equation 2), the maximum specific growth rate parameters (μ_{S_i}) are defined for each substrate (S_i) separately, as well as the Monod constants ($K_{S_i}^M$). For both species, the substrates consisted of glucose, fructose, and citrate.

For *L. fermentum* strains, mannitol production was only made possible when fructose was present (Supplementary Files 2 and 3). Growth term equations for the Monod growth model of *L. plantarum* strains were modified by the addition of a Hill inhibition term to the Monod growth equation, depending on the presence of lactate (Supplementary Files 4 and 5). The growth terms were summed to obtain the total specific growth rate, assuming substitutable substrates (Kettle *et al*., 2018). Additionally, a separate ODE was constructed, based on the Monod growth model but including consumption and production flux parameters to describe the metabolite dynamics. The flux parameter values were controlled by Hill equations which, depending on the metabolite concentrations, increased or decreased the flux parameter values. The ODE models were solved using the SciPy integrate package with the VODE numerical integrator (Jones *et al*., 2001), except for the Monod growth model with flux parameters, which was solved with the Euler method. The latter allowed to save the intermediary flux parameter values. The time-step was set to 0.1 h and non-negative concentrations of extracellular metabolites were enforced. Parameter values of the dynamic equations were manually fitted to the experimental data points.

Subsequently, the fitted parameter values of the ODE models were used as guidelines to define parameter values of the dFBA models (Supplementary Files 6 and 7). Here, data sets that describe the population and metabolite dynamics of *L. fermentum* NCC 575 and *L. plantarum* NCC 2829 were used as a reference for *L. fermentum* 222 and *L. plantarum* 80, respectively, since they include the biomass data in units of g_{CDW}. The specific growth rate was predicted by the *L. fermentum* 222 and *L. plantarum* 80 core GEMs. Secretion reactions for metabolites not measured in the original data sets were blocked. The PPP was also constrained by blocking the flux through the transaldolase reaction. Additionally, a growth-associated maintenance constraint was added, forcing the model to overproduce ATP at a certain specific growth rate. For *L. fermentum* 222, the fitted product yield of fructose to mannitol conversion was added as a constraint to the core GEM. Additionally, after citrate was fully consumed, an exchange reaction for oxaloacetate was activated, allowing further growth to be predicted. At the same time, the flux through oxaloacetate decarboxylase (EC 1.1.1.38, 1.1.1.40) was blocked. For *L. plantarum* 80, a decrease in lactate productivity was simulated by applying a lactate-dependent inhibition term on the upper bound of the lactate production flux of the *L. plantarum* 80 core GEM. Activation of the growth-associated maintenance constraint and change of the allowable flux bounds was controlled using modulating Hill equations, which were depending on the citrate concentration for *L. fermentum* 222 and the lactate concentration for *L. plantarum* 80. Anaerobic conditions were assumed for the dFBA simulations, since no aeration was performed under the experimental conditions. Parameter values of the dynamic equations were manually fitted to the experimental data points.

3 Results

3.1 Development of core genome-scale metabolic models of *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80

Functional annotation information of the two cocoa bean fermentation-related *Lactobacillus* species examined in the present study, *L. fermentum* 222 and *L. plantarum* 80, was stored in their respective relational databases. For *L. fermentum* 222, the functional annotation comprised 1864 protein-encoding genes originally annotated with GenDB, extended with 135 protein-encoding genes unique for NCBI RefSeq, and 221 protein-encoding genes unique for PATRIC.

For *L. plantarum* 80, the functional annotation comprised 3123 protein-encoding genes originally annotated with GenDB, extended with 145 protein-encoding genes unique for NCBI RefSeq, and 227 protein-encoding genes unique for PATRIC.

Manual curation resulted in core GEMs of *L. fermentum* 222, containing 48 genes, 42 reactions, and 50 metabolites, and *L. plantarum* 80, containing 80 genes, 52 reactions, and 52 metabolites. These models did not contain transport reactions nor a biomass production reaction.

Both models contained only one compartment. This has as a consequence that all transport reactions were implicit, meaning that an exchange reaction, without gene-protein-reaction

association, was only included for an intracellular metabolite if needed for the FBA simulation performed. Furthermore, no proton-translocating enzymes were included, such as the H^+-transporting ATP synthase (EC 7.1.2.2). Thus, monosaccharide permease systems were assumed to be simple exchange reactions, without usage of a proton motive force. This allowed to estimate total ATP production by substrate-level phosphorylation when ATP hydrolysis was used as the objective function. Substrate exchange reactions were only applied for glucose, fructose, and citrate, as their consumption has been found experimentally (Lefeber *et al*., 2010). For both *Lactobacillus* species, candidate permease transporters and phosphotransferase systems (PTS) could be retrieved for D-glucose and D-fructose. The compartmentalization strategy entailed that the PTS were simulated as being an intracellular phosphorylation reaction of the monosaccharide with phosphoenolpyruvate as phosphate/energy donor, thus neglecting the phosphorylation cascade and cytoplasmic transport. This simplification is justified by the different functional roles of the monosaccharides and their phosphorylated derivatives in the metabolic network. Monosaccharide PTS have been typically associated to homofermentative LAB, but also in heterofermentative LAB phospoenolpyruvate-dependent PTS have been found (Wisselink *et al*., 2002; Zheng *et al*., 2015). Experimentally characterized fructose PTS that phosphorylate fructose on the C1 position, giving fructose 1-phosphate, have been described in *E. coli* and *Streptomyces coelicolor* (Kornberg, 1986; Nothaft *et al*., 2003). In *Bifidobacterium breve,* fructose is phosphorylated to fructose 6-phosphate (Maze *et al*., 2007). The reaction catalyzed by 1-phosphofructokinase (EC 2.7.1.56), namely the phosphorylation of fructose 1-phosphate to fructose 1,6-biphosphate, an intermediary of the EMP pathway, has been proposed as a necessary link between the fructose PTS-dependent transport and catabolic pathways in *L. brevis* (Saier *et al*., 1996), and has been encoded as such in the GEM of *L. plantarum* WCFS1 (Teusink *et al*., 2006). Only for *L. plantarum* 80, a 1-phosphofructokinase-encoding gene (LP80_903) could be identified. Since no such gene was present in the genome of *L. fermentum* 222, the assumption was made that fructose PTS phosphorylates fructose to fructose 1-phosphate in *L. plantarum* 80 and fructose to fructose 6-phosphate in *L. fermentum* 222. Only the Enzyme II (EII) subunits were included in the gene-protein-reaction associations of the glucose and fructose PTS in both *Lactobacillus* species, since the phosphocarrier HPr protein (LFER_1497, LP80_2597) and Enzyme I (EI, LFER_1498 or LFER_1548, LP80_2596) are shared between these systems. The reaction catalyzed by citrate lyase (EC 4.1.3.6) was added to both GEMs, as it has been found in both *Lactobacillus* genomes previously (Illeghems *et al*., 2015).

Previously, no mannitol 2-dehydrogenase-encoding gene could be identified in the genome of *L. fermentum* 222, although fermentation profiles of *L. fermentum* 222 are characterized by the formation of D-mannitol from D-fructose (Lefeber *et al*., 2010, 2011b). The genome re-annotation process could however identify a candidate mannitol 2-dehydrogenase-encoding gene, previously annotated as a zinc-containing alcohol dehydrogenase (LFER_58; Supplementary File 1), by combining orthogroup information from the curated mannitol dehydrogenase protein sequence of *Lactobacillus sanfransiscensis* TMW 1.1304 and a good alignment (100 % coverage, 89 % identity) with the experimentally characterized NADP-dependent mannitol 2-dehydrogenase of *L. reuteri* ATCC 53608 (Sasaki *et al*., 2005). Similarly, no previous candidate enzyme-encoding genes were found in *L. fermentum* 222 for

fructokinase (EC 2.7.1.4) and glucose-6-phosphate isomerase (EC 5.3.1.9), key enzymes of the EMP pathway. By re-annotating the genome, a gene encoding a candidate fructokinase was found (LFER_856; Supplementary File 1) as well as two genes encoding a candidate glucose-6-phosphate isomerase (LFER_RS04265 and LFER_RS03170). However, the latter candidates seemed truncated genes, having only 12 % alignment coverage to an identified glucose-6-phosphate isomerase in *L. fermentum* NRRL-B-1932 (Helanto *et al*., 2006). In that strain, the glucose-6-phosphate isomerase (*fruI*) and fructokinase (*fruK*) genes form an operon. A similar genomic structure was found for the genes LFER_856 and LFER_RS04265 for *L. fermentum* 222. The glucose-6-phosphate isomerase enzyme is essential to channel fructose to the phosphoketolase pathway with glucose 6-phosphate as the necessary intermediate. Without this reaction fructose could not be used as an energy source but only as an alternative electron acceptor for reduction to mannitol (Wisselink *et al*., 2002). When *L. fermentum* 222 is grown in a medium with fructose as the sole energy source, it produces considerable amounts of lactic acid, acetic acid, and ethanol, next to mannitol (Lefeber *et al*., 2010). This finding does not align with fructose, being solely used as alternative electron acceptor. Since the *L. fermentum* 222 genome has a draft status and both LFER_RS04265 or LFER_RS03170 were found at the boundaries of their respective contigs, it cannot be excluded that during genome assembly the glucose-6-phosphate isomerase-encoding sequence was lost. The integrated information led us to tentatively add the glucose-6-phosphate isomerase reaction to the *L. fermentum* 222 GEM.

Both *Lactobacillus* species encoded the bifunctional acetaldehyde dehydrogenase (EC 1.2.1.10)/alcohol dehydrogenase (EC 1.1.1.1) enzyme, necessary for the reduction of acetyl-CoA to ethanol. Additionally, for both species, candidate enzyme-encoding genes were retrieved for a NAD or NADP-dependent malate dehydrogenase (malic enzyme; EC 1.1.1.38, EC 1.1.1.40) and malolactic enzyme (EC 4.1.1.101). In addition, *L. plantarum* 80 contained a separate malate dehydrogenase (EC 1.1.1.37). Next to the enzymes previously found in pyruvate metabolism, an additional pyruvate carboxylase (EC 6.4.1.1) and phosphoenolpyruvate carboxykinase (EC 4.1.1.49) were found in the genome annotation of *L. plantarum* 80. The occurrence of fumarate reductase in both *Lactobacillus* species has been suggested (Illeghems *et al*., 2015). However, no succinate production was found for *L. fermentum* 222 and *L. plantarum* 80. Yet, it is known that some *Lactobacillus* species can produce succinate from citrate using the reductive TCA cycle (Dudley and Steele, 2005; Kang *et al*., 2013; Tsuji *et al*., 2013; Gänzle, 2015). The presence of the previously identified genes encoding a flavoprotein in *L. fermentum* 222 and *L. plantarum* 80 (Illeghems et al., 2015), raised the hypothesis that these enzymes could be involved in the use of fumarate as an alternative electron acceptor, albeit not yet found experimentally. Therefore, an unspecified flavoprotein-dependent fumarate-reducing reaction was added to both GEMs. Finally, the *L. plantarum* 80 core GEM contained different oxygen-consuming enzymes, being NADH oxidase (EC 1.6.3.3 and EC 1.6.3.4) and pyruvate oxidase (EC 1.2.3.3).

The reconstructed metabolic networks of *L. fermentum* 222 and *L. plantarum* 80 are shown in Supplementary Figures 2 and 3. The GEMs are available as excel and xml files (SBML format) as well. Metabolic schemes of the carbon core metabolism of *L. fermentum* 222 and *L. plantarum* 80 are shown in Figure 1.

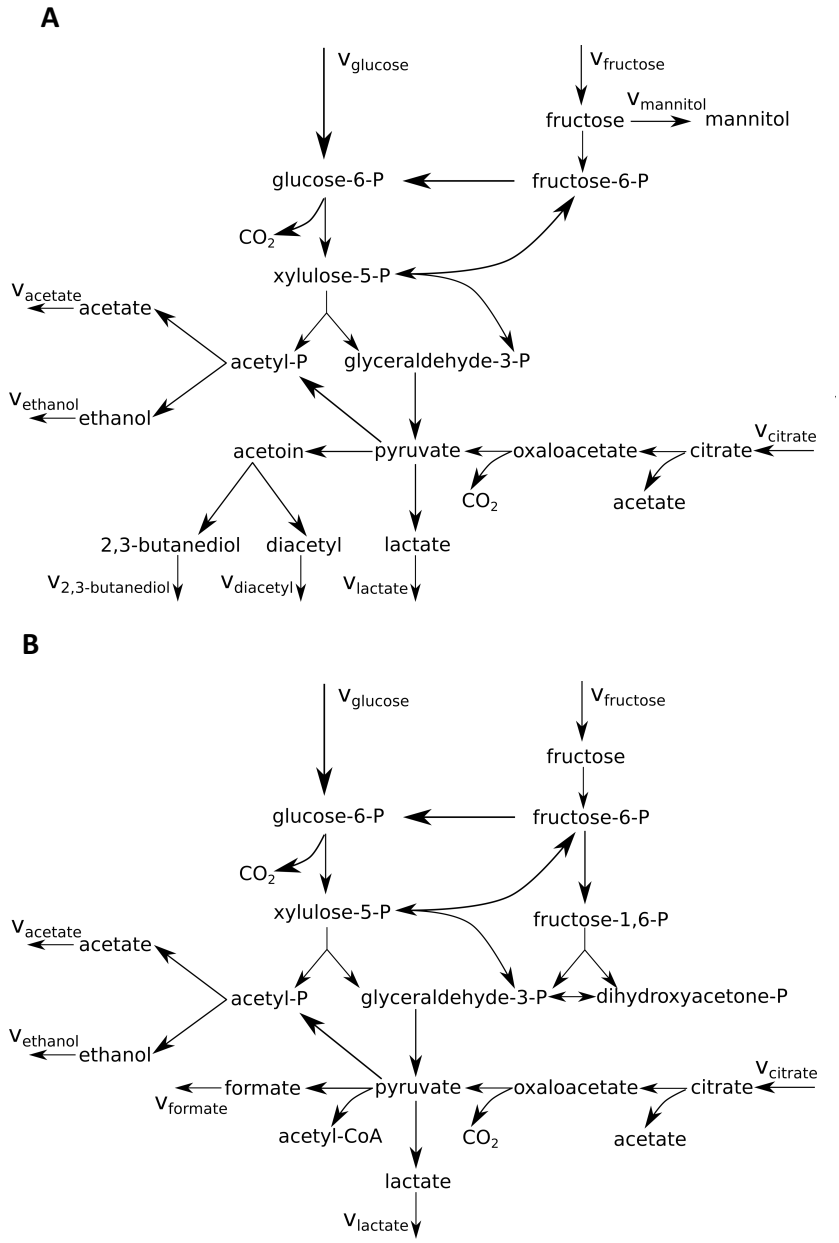


Figure 1. Metabolic scheme of the carbon core metabolism of *Lactobacillus fermentum* 222 (A) and *Lactobacillus plantarum* 80 (B). For each metabolic model, the exchange fluxes of metabolites are shown as $v_{metabolite}$. "P" stands for phosphate. Edges connecting xylulose-5-phosphate, fructose-6-phosphate, and glyceraldehyde-3-phosphate represent the non-oxidative branch of the pentose phosphate pathway.

3.2 Genome-scale metabolic modeling of *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80

For the *L. fermentum* 222 core GEM, characterized by co-consumption of glucose, fructose, and citric acid under cocoa-pulp simulating conditions, the addition of the PPP constraint was necessary to predict glucose and fructose catabolism *via* the phosphoketolase pathway. In the absence of citrate and oxygen consumption, and only when the PPP constraint was enforced, the *L. fermentum* 222 core GEM predicted mannitol formation from fructose. Using ATP formation as the objective function additionally revealed that blocking ethanol secretion

decreased the amount of ATP formed, while at the same time allowing mannitol formation. A similar simulation using fructose as the sole substrate essentially revealed the same result. When glucose was the sole substrate available, and taking into account the PPP constraint, *in silico* growth was only possible if oxygen consumption was allowed. Pyruvate oxidase (EC 1.2.3.3), being the only enzyme present in the *L. fermentum* 222 core GEM that could consume oxygen, allowed to increase the flux through the acetyl phosphate-ethanol pathway branch, ensuring redox balancing. Since pyruvate oxidase produces hydrogen peroxide, it was predicted to be secreted. None of the conditions tested allowed *in silico* growth when citrate was the sole substrate available.

For the *L. plantarum* 80 core GEM, the main flux was through the EMP pathway under cocoa-pulp simulating and anaerobic conditions. The addition of the PPP constraint did not lead to a change in the predicted specific growth rate. Although this parameter influenced the metabolic flux distribution, and the secretion fluxes of formate, ethanol, and acetate in particular, the flux variability as resulting from FVA was within the optimal solution space. Allowing oxygen consumption omitted the need to enforce the PPP constraint to predict glucose and fructose catabolism via the EMP pathway. In this case, oxygen could be used as an alternative electron acceptor, omitting the need of the formation of ethanol and lactate. Acetate, formate, and carbon dioxide were the sole carbon-containing metabolites secreted. In all cases, the EMP pathway was the dominant energy-generating pathway. Malate dehydrogenase (EC 1.1.1.37) in combination with malic enzyme (EC 1.1.1.38/40) and pyruvate-formate lyase (EC 2.3.1.54) allowed to ensure the redox balance, whereby formate secretion was always accompanied by ethanol secretion. When glucose and citrate were available as the sole substrates and oxygen consumption was not allowed, FBA predicted the formation of formate, ethanol, and acetate. No lactate formation was predicted by the model. *In silico* growth on citrate only was possible, since the *L. plantarum* 80 core GEM contained fructose-bisphosphatase (EC 3.1.3.11). In this case, no oxygen consumption was predicted.

Next, metabolic flux analysis data of *L. fermentum* NCC 575 and *L. plantarum* NCC 2829 (Adler *et al*., 2013) was used as a reference to simulate the metabolic activities of *L. fermentum* 222 and *L. plantarum* 80, respectively, using their core GEMs.

For the *L. fermentum* 222 core GEM, the use of the consumption flux constraints solely led to a maximal consumption of glucose, fructose, and citrate, together with oxygen, leading to the production of mainly acetate, next to hydrogen peroxide and ethanol. No mannitol was produced. The carbon recovery obtained was 72 % under aerobic conditions and 76 % under anaerobic conditions. When the production flux constraints were enforced and the additional secretion reactions were blocked, for example for ethanol, a metabolic flux distribution was obtained whereby a low but significant flux of 12 % was directed toward the phosphoketolase pathway. No oxygen was consumed. The carbon recovery and carbon dioxide production flux obtained were 91 % and 192 mmol/g_{CDW}/h, respectively, when the PPP constraint was enforced, and 89 % and 151 mmol/g_{CDW}/h, respectively, when this was not the case. Omitting the production flux constraints only led to mannitol formation when the additional secretion reactions were blocked. The fructose into mannitol conversion was increased by enforcing the

PPP constraint, leading to a complete conversion of fructose under aerobic conditions and 97 % conversion under anaerobic conditions.

For the *L. plantarum* 80 core GEM, the use of the consumption flux constraints solely led to a maximal consumption of glucose, fructose, and citrate, together with oxygen, leading to the production of acetate, formate, and carbon dioxide. Under anaerobic conditions, also ethanol formation was predicted, but not lactate formation. The carbon recovery obtained was 59 % under aerobic conditions and 68 % under anaerobic conditions. When the production flux constraints were enforced and the additional secretion reactions blocked, the model predicted the presence of a futile cycle, consuming ATP, both under aerobic and anaerobic conditions. This indicated that the metabolic exchange fluxes found corresponded with a metabolic flux distribution having an excess of energy produced. Under aerobic conditions with the addition of the PPP constraint, the co-utilization of the EMP and phospoketolase pathways was predicted, and the simulation resulted in a carbon recovery of 80 %. Omitting the production flux constraints led to ethanol production in all cases. If the additional secretion reactions were blocked, lactate formation was only predicted when oxygen consumption was not allowed.

3.3 Modeling of the bacterial cell population and metabolite dynamics of *Lactobacillus fermentum* and *Lactobacillus plantarum* under cocoa-pulp simulating conditions using a genome-scale metabolic model

For *L. fermentum* NCC 575, fitting the Verhulst growth model to the experimental data indicated a carrying capacity of 1.692 g_{CDW}/l and a maximum specific growth rate of 0.55 h^{-1} (Table 1). As mannitol production is known to occur from fructose only, its production yield could be specified to be 0.9 mmol/mmol.

Table 1. Fitted dynamic ordinary differential equation (ODE) model parameters of the *Lactobacillus* strains investigated.

Model parameter	*L. fermentum* 222	*L. fermentum* NCC 575	*L. plantarum* 80	*L. plantarum* NCC 2829
Carrying capacity (mg/l)	233.070	1692.450	677.240	743.150
Maximum specific growth rate (1/h)	0.440	0.550	0.230	0.160
Maximum specific growth rate on glucose (1/h)	0.100	0.700	0.160	0.120
Maximum specific growth rate on fructose (1/h)	0.200	0.300	0.160	0.120
Maximum specific growth rate on citrate (1/h)	0.600	0.600	0.050	0.050
Monod constant glucose (mM)	140.000	140.000	140.000	140.000
Monod constant fructose (mM)	140.000	140.000	140.000	140.000
Monod constant citrate (mM)	40.000	30.000	40.000	40.000
Glucose-to-biomass yield (mmol/mg)	1.000	0.050	0.100	0.050
Fructose-to-biomass yield (mmol/mg)	1.500	1.100	0.250	0.120

Table 1. Fitted dynamic ordinary differential equation (ODE) model parameters of the *Lactobacillus* strains investigated (continued).

Model parameter	*L. fermentum* 222	*L. fermentum* NCC 575	*L. plantarum* 80	*L. plantarum* NCC 2829
Citrate-to-biomass yield (mmol/mg)	2.000	0.350	0.200	0.050
Lactate-from-biomass yield (mmol/mg)	0.450	0.070	0.300	0.150
Acetate-from-biomass yield (mmol/mg)	0.550	0.077	0.050	0.020
Mannitol-from-fructose yield (mmol/mmol)	1.100	0.900	NA	NA

NA, not available

Since the metabolite dynamics were dependent on the Verhulst growth term, fitting the metabolite dynamics with the Verhulst growth model was not successful. To allow a better fit of the metabolite dynamics of *L. fermentum* 222, the Monod growth model was used, for which the substrate dynamics were described by growth terms as a function of each substrate separately (Figure 2A). Further, the experimental data of *L. fermentum* NCC 575 were better fitted using a Hill equation in the growth functions, with the Hill coefficient set to 5 (Figure 2B).

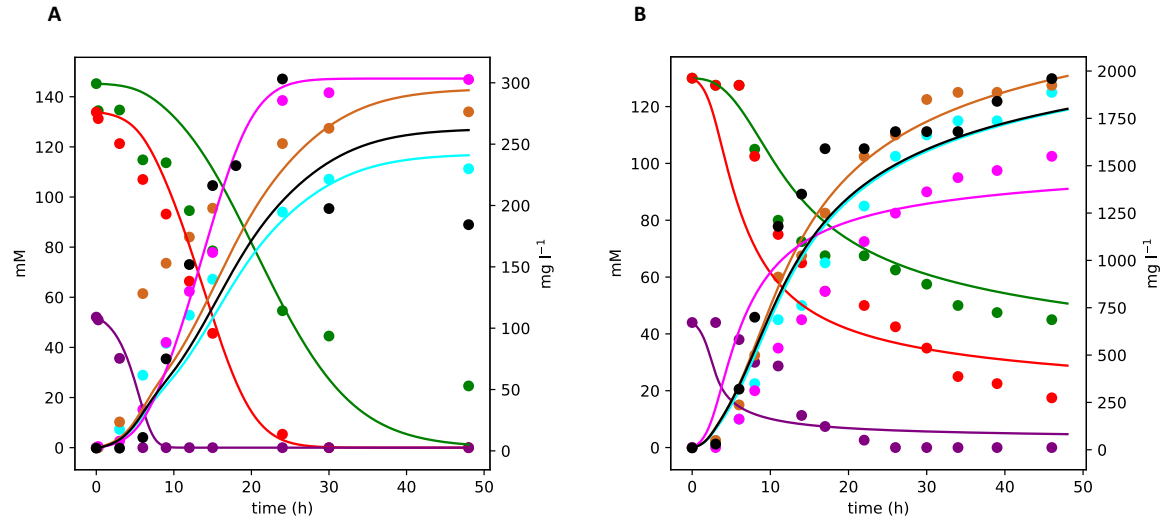


Figure 2. Population (biomass in black) and metabolite dynamics fit (glucose in green, fructose in red, citrate in purple, lactate in cyan, acetate in brown, and mannitol in magenta) of *Lactobacillus fermentum* 222 under cocoa pulp-simulating conditions with the Monod growth model (A) and *Lactobacillus fermentum* NCC 575 with the Monod growth model adapted with a Hill equation (B). Dots represent experimental data points, lines represent dynamic model fittings.

Additionally, the experimental data published before revealed that after citrate consumption, the specific growth rate decreased, leading to a slower fructose and glucose consumption.

These experimental data points were fitted using dFBA by adding a growth-associated maintenance constraint to the *L. fermentum* 222 core GEM, depending on the citrate concentration. A difficulty arose to fit glucose consumption after citrate was consumed. This was due to the loss of the role of citrate as alternative electron acceptor, decreasing the amount of glucose that could be catabolized *via* the phosphoketolase pathway. Fructose could be completely converted into mannitol, but this did not correspond with the fructose and mannitol dynamics found. Thus, additional glucose consumption seemed to be fundamentally constrained by the lack of sufficient concentrations of external electron acceptors. Allowing oxygen usage could only release this constraint if ethanol secretion was allowed, but this metabolite was not detected in the growth medium (Adler *et al.*, 2013). Thus, an additional small consumption of (2S)-2-acetolactate (max 0.8 mmol/g_{CDW}/h) was allowed, which resulted in the formation of 2,3-butanediol. The resulting final concentration of 2,3-butanediol in the *in silico* growth medium was 50 mM (Figure 3).

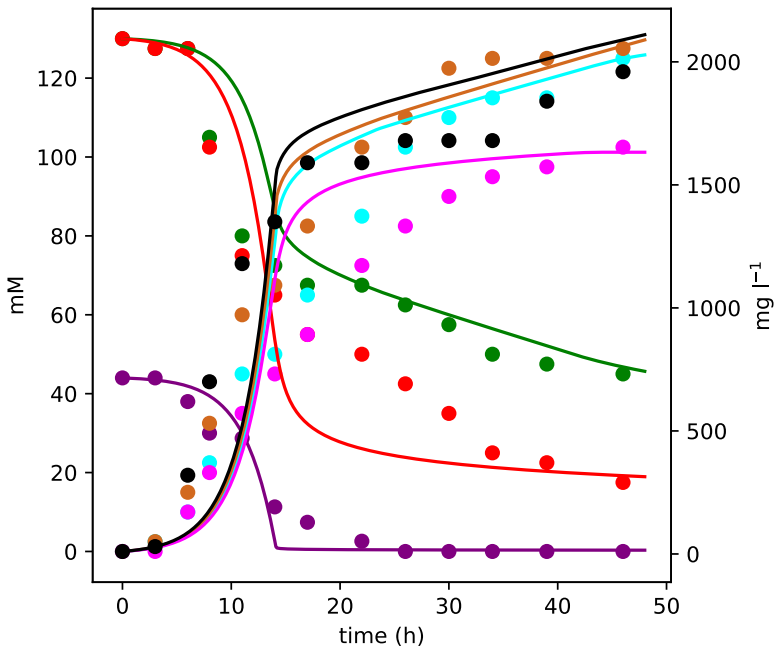


Figure 3. Population (biomass in black) and metabolite dynamics fit (glucose in green, fructose in red, citrate in purple, lactate in cyan, acetate in brown, and mannitol in magenta) of *Lactobacillus fermentum* NCC 575 under cocoa pulp-simulating conditions with dynamic flux balance analysis. Dots represent experimental data points, lines represent dynamic model fittings.

For *L. plantarum* NCC 2829, the Verhulst growth model indicated a carrying capacity of 0.743 g_{CDW}/l and a maximum specific growth rate of 0.16 h^{-1} (Table 1). The metabolite dynamics were well fitted using the Verhulst growth model, except for lactate production. The fit of *L. plantarum* 80 and *L. plantarum* NCC 2829 was optimized with the Monod growth model, including lactate-dependent growth inhibition. The Hill coefficient in the lactate-dependent inhibition term was set to 8 for *L. plantarum* 80 and to 3 for *L. plantarum* NCC 2829 (Figure 4).

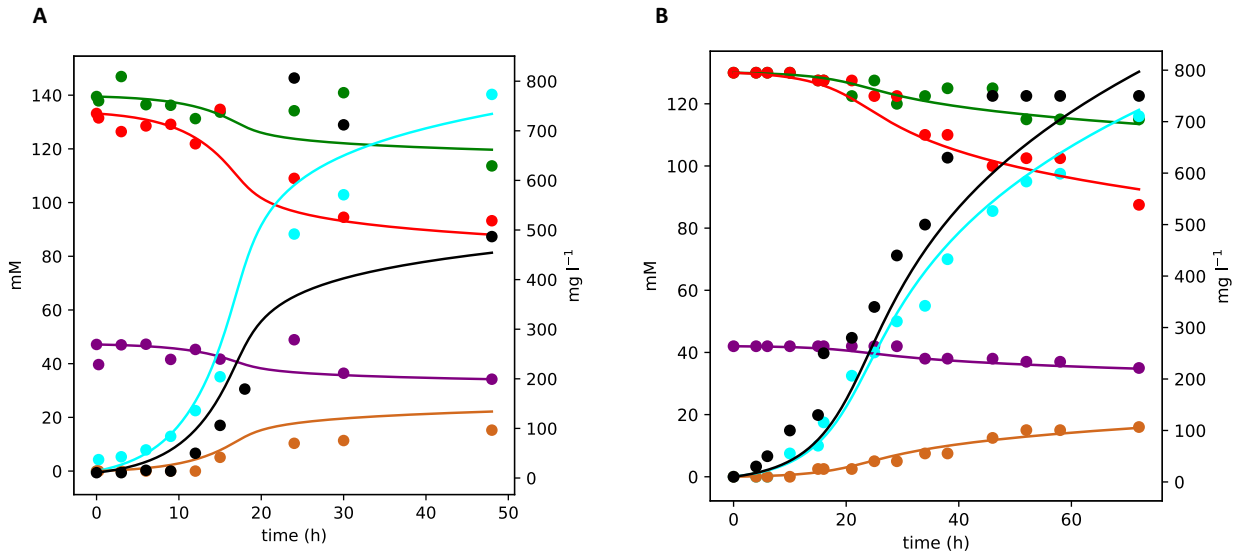


Figure 4. Population (biomass in black) and metabolite dynamics fit (glucose in green, fructose in red, citrate in purple, lactate in cyan, acetate in brown, and mannitol in magenta) of *Lactobacillus plantarum* 80 (A) and *Lactobacillus plantarum* NCC 2829 (B) under cocoa pulp-simulating conditions with the Monod growth model adapted with a lactate-dependent Hill inhibition term. Dots represent experimental data points, lines represent dynamic model fittings.

Lactate inhibition was used to fit the experimental data points using dFBA by adding a growth-associated maintenance constraint to the *L. plantarum* 80 core GEM, depending on the lactate concentration. Using the parameter values of the Monod growth model in the dFBA simulation led to a good fit of the biomass and metabolite dynamics, except for lactate production (Figure 5). Using the same Monod growth model with the flux parameters allowed to estimate the actual consumption and production fluxes of the metabolites considered, instead of their biomass yields. Closer inspection of the data revealed that although the total amount of glucose and fructose consumed theoretically led to the total amount of lactate produced, assuming the stoichiometry of the EMP pathway, the initial lactate production fluxes in the Monod growth model were at least three times higher than those predicted by FBA. This pointed to a problem whereby the product yield and the product flux did not correspond, essentially revealing that the yield parameter did not give a time indication. Artificially increasing the lactate production flux would lead to a redox imbalance, since no additional glucose and fructose is consumed according to the experimental data. Since *L. plantarum* NCC 2829 was grown in a complex medium, it is not unlikely that other metabolites were consumed that could explain this fermentation pattern.

4 Discussion

Re-annotation of the *L. fermentum* 222 and *L. plantarum* 80 genomes with the goal to develop GEMs allowed to perform a system-level analysis of the core metabolism of these strains and thereby tentatively specify the functionalities of specific metabolic enzymes, for example those involved in the transport of glucose and fructose as well as those involved in redox balancing.

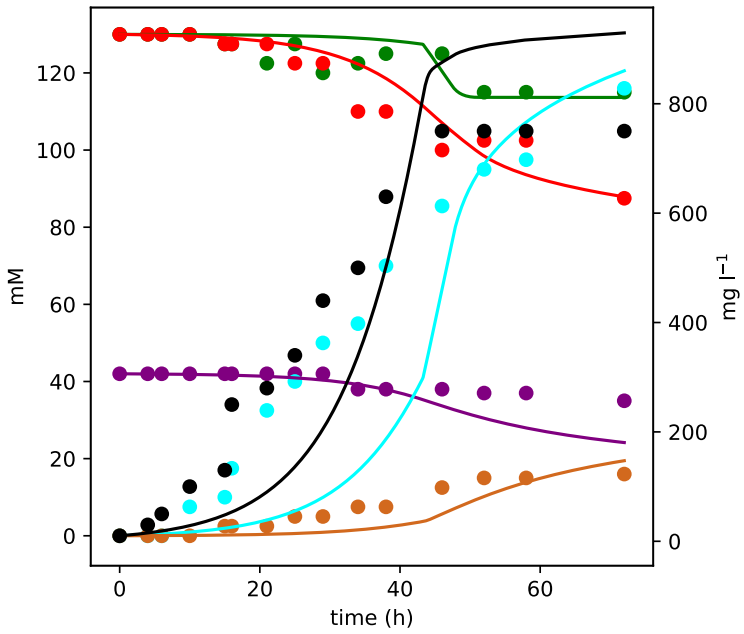


Figure 5. Population (biomass in black) and metabolite dynamics fit (glucose in green, fructose in red, citrate in purple, lactate in cyan, acetate in brown, and mannitol in magenta) of *Lactobacillus plantarum* NCC 2829 under cocoa pulp-simulating conditions with dynamic flux balance analysis. Dots represent experimental data points, lines represent dynamic model fittings.

Species of the genus *Lactobacillus fermentum* and *Lactobacillus plantarum* have recently been reclassified as *Limosilactobacillus fermentum* and *Lactiplantibacillus plantarum*, respectively (Zheng *et al*., 2020). This reclassification did not impact the development of the *L. fermentum* 222 (now *Limosilactobacillus fermentum* 222) and *L. plantarum* 80 (now *Lactiplantibacillus plantarum* 80) GEMs. Future statements related to the link with bacterial taxonomy of these models should however use the new classification genus names.

The metabolic potential for oxygen consumption seemed to be higher for *L. plantarum* 80 than for *L. fermentum* 222. This may contradict a previously used statement that the lower oxygen tolerance of *L. plantarum* strains in comparison to *L. fermentum* strains could explain the dominance of *L. fermentum* in succession of *L. plantarum* during spontaneous cocoa bean fermentation processes (De Vuyst and Weckx, 2016b). Reconstructed core GEMs without predefined transport reactions nor an objective function allowed to investigate metabolic flux distributions using different combinations of metabolites consumed and produced on the one hand, and the influence of maximization of the biomass production or ATP formation on the other hand, thus providing a flexible framework to perform FBA simulations *à la carte*.

In general, for *L. fermentum*, the addition of a PPP constraint was necessary to obtain carbohydrate catabolism via the phosphoketolase pathway, as has been found before for *L. fermentum* NCC 575 (Adler *et al*., 2013). The non-oxidative PPP is favored in terms of biomass yield predicted by FBA, since it is redox-neutral and prevents carbon loss in the form of carbon dioxide. Connecting the PPP to the lower part of the EMP pathway led to increased ATP

production, which could be further maximized by acetate production from acetyl phosphate. Investigation of a number of already published LAB GEMs, developed for *L. plantarum* WCSF1 (Teusink *et al*., 2006), *Lactobacillus casei* 12A and *L. casei* ATCC 334 (Vinay-Lara *et al*., 2014), *Leuconostoc mesenteroides* subsp. *cremoris* ATCC 19254 (Özcan *et al*., 2019), *Oenococcus oeni* PSU-1 (Mendoza *et al*., 2017), *Le. mesenteroides* ATCC 8293 (Koduru *et al*., 2017), and *Lactococcus lactis* MG1363 (Flahaut *et al*., 2013), led to the conclusion that in many cases, at least one of the reactions usually associated to the enzymes of the PPP was not present or constrained in these models. This may prevent a flux cycle involving PPP enzymes to occur in these models.

FBA simulations with the *L. fermentum* 222 core GEM showed that the phosphoketolase pathway is redox-balanced when ethanol is secreted. In terms of ATP product yield, this means that increasing the carbon flux through this pathway was more optimal than uncoupling NAD^+ recuperation by the fructose into mannitol conversion. Under cocoa pulp-simulating conditions, *L. fermentum* 222 and *L. fermentum* NCC 575, almost completely convert fructose into mannitol, meaning that fructose is used as alternative electron acceptor only (Lefeber *et al*., 2010; Adler *et al*., 2013). It has been proposed that fructose is only reduced to mannitol when fructose is present in excess in the growth medium, and that this is due to the relatively high Michaelis-Menten constant of mannitol dehydrogenase (Wisselink *et al*., 2002). Thus, the separation of glucose catabolism, to generate energy, and fructose conversion into mannitol, to ensure the redox balance, seems to be an adaptation to optimize the substrate conversion rate rather than an optimization of the biomass or energy yields.

When simulating only glucose availability under aerobic conditions, the model predicted hydrogen peroxide production by pyruvate oxidase. Hydrogen peroxide production has also been found experimentally for *L. fermentum* strains growing in a glucose-containing medium, whereby its accumulation in the medium ultimately leads to poor growth (Sakamoto and Komagata, 1996). When citrate was the sole substrate available, none of the conditions tested allowed *in silico* growth. This was due to the lack of a gluconeogenesis pathway in the *L. fermentum* 222 core GEM. *In vitro* growth of *L. fermentum* 222 in CPSM with citrate as the sole added energy source was however possible (Lefeber *et al*., 2010). This is an indication that *L. fermentum* 222 has the metabolic potential to use complex medium components required for its growth.

Simulating the metabolic flux distribution of *L. fermentum* 222 under cocoa pulp-simulating conditions, and adding consumption and production flux constraints based on experimental data obtained for *L. fermentum* NCC 575, led to a predicted fructose consumption flux of 12 % directed toward the phosphoketolase pathway, which is higher than the 4 % reported for *L. fermentum* NCC 575 before (Adler *et al*., 2013). Although this indicated a semi-quantitative agreement between the metabolic flux analysis data and FBA making use of constraints, biomass optimization still leads to an underestimation of the amount of mannitol produced. Additionally, a carbon recovery of around 90 % was predicted, which corresponded to the 90.3 % reported for *L. fermentum* NCC 575 before (Adler *et al*., 2013). This indicated an agreement

of the GEM and the previous estimation of the amount of carbon dioxide produced by *L. fermentum* NCC 575 under cocoa pulp-simulating conditions.

Under cocoa pulp-simulating conditions in the absence of oxygen consumption, the *L. plantarum* 80 core GEM predicted glucose and fructose catabolism via the EMP pathway, as has been found before for *L. plantarum* NCC 2829 (Adler *et al*., 2013). Allowing the possibility of oxygen consumption, no lactate production was predicted by the *L. plantarum* 80 core GEM. The metabolic flux distribution that followed from unlimited oxygen consumption was an extreme case of the finding that homofermentative LAB growth under aerobic conditions leads to a decrease in their lactate production and an increase in their acetate production (Sakamoto and Komagata, 1996; Zotta *et al*., 2017). Also, on glucose and citrate as the sole substrates, the *L. plantarum* 80 core GEM did not predict lactate formation. A similar result has been obtained with FBA for the *L. plantarum* WCFS1 GEM, which was however in contradiction to the formation of mainly lactate during a chemostat experiment using a chemically defined medium with glucose and citric acid as the sole substrates (Teusink *et al*., 2006). The *in silico* result of a mixed-acid fermentation pattern instead of mainly lactate formation has been attributed to the notion that *L. plantarum* WCFS1 does not optimize its biomass yield under these growth conditions. The fact that pyruvate-formate lyase expression in *Lc. lactis* is repressed by glucose and allosterically inhibited by glycolytic intermediates contributes to this idea (Teusink *et al*., 2006). Simulating the metabolic flux distribution of *L. plantarum* 80 under cocoa pulp-simulating conditions in the presence of oxygen, and adding consumption and production flux constraints based on experimental data obtained for *L. plantarum* NCC 2829, led to the co-utilization of the EMP and phospoketolase pathway, as has been found before (Adler *et al*., 2013). In this case the carbon recovery was 80 %, which was lower than that one of 109.5 % obtained before for *L. plantarum* NCC 2829 (Adler *et al*., 2013).

Additional consumption of (2S)-2-acetolactate was necessary to fit the *L. fermentum* 222 core GEM to the *L. fermentum* NCC 575 bacterial cell population and metabolite dynamics data, leading to the production of 2,3-butanediol in the growth medium *in silico*. Previously, the formation of 2,3-butanediol has been found for *L. fermentum* 222 growing in CPSM (Lefeber *et al*., 2011b). However, no 2,3-butanediol formation occurs with *L. fermentum* NCC 575 in the same growth medium (Adler *et al*., 2013). The addition of a degree of freedom to the model, in the form of (2S)-2-acetolactate, allowed a better fit of the bacterial cell population and metabolite dynamics for the whole growth curve of *L. fermentum* NCC 575. ODE modeling of the *L. plantarum* 80 and *L. plantarum* NCC 2829 dynamics included the additional assumption of lactate inhibition on the specific growth rate. Since the pH also decreases from 3.8 to 3.4 during the fermentation process (Adler *et al*., 2013), not only the concentration of lactate but its influence on the acidity of the medium could explain the resulting dynamics. Using dFBA, the lactate dynamics of *L. plantarum* NCC 2829 could not be fitted using the *L. plantarum* 80 GEM. Similarly as for *L. fermentum*, consumption of unknown complex medium components could explain the bacterial cell population and metabolite dynamics obtained. Interestingly, proteolytic systems have been identified in the genomes of *L. fermentum* 222 and *L. plantarum* 80 (Illeghems *et al*., 2015). Thus, the degradation of proteins and peptides in a complex medium by *L. fermentum* NCC 575 and *L. plantarum* NCC 2829 may lead to replenishment of

the intracellular (2S)-2-acetolactate and pyruvate pools, respectively, which could explain the extracellular metabolite dynamics of both strains.

5 Conclusion

In this chapter, the complete workflow of GEM reconstruction and analysis, in a static and dynamic way, was performed using the candidate functional starter cultures, *L. fermentum* 222 and *L. plantarum* 80, for cocoa bean fermentation processes. Since their genomes are in a draft status and no information is available regarding their nutritional requirements, core GEMs were reconstructed, focusing on their central carbon metabolic pathways only. A strategy combining genome re-annotation, comparative genomics, and targeted sequence alignments was performed to obtain an accurate metabolic network reconstruction and gene-protein-reaction associations, resulting in the *L. fermentum* 222 and *L. plantarum* 80 core GEMs, of which the *L. fermentum* 222 core GEM is the first for the species *L. fermentum*.

Both for *L. fermentum* 222 and *L. plantarum* 80, comparison of metabolic flux distributions obtained *in silico* with experimental data available in the literature for other strains of the same species, grown in a cocoa pulp simulation medium, revealed a deviation of the central assumption of FBA. The metabolic flux distributions corresponding to an optimization of the yield, be it biomass production or ATP formation, did not correspond to the metabolic phenotypes found for the *L. fermentum* and *L. plantarum* strains examined. Thus, optimization of the reaction rates by these strains may have occurred, thereby enhancing their competitiveness in the cocoa bean fermentation process. This is in line with the FBA results obtained for *Acetobacter pasteurianus* 386B (Chapter 5). Additional knowledge regarding the metabolites produced is needed, to define appropriate constraints, after which *a posteriori* FBA can quantitatively estimate the remaining unknown fluxes, such as the intracellular fluxes or the carbon dioxide production flux, which are more difficult to measure. Similarly, a number of constraints related to the consumption and production fluxes as well as growth-associated maintenance were needed to fit the dFBA models. They represented more or less mechanistic hypotheses to explain the bacterial cell population and metabolite dynamics obtained and need to be tested experimentally in the future.

Chapter 8

General conclusions

Rudy Pelicaen, Didier Gonze, Luc De Vuyst, and Stefan Weckx

An in-depth understanding of the physiology of starter cultures is necessary to apply them in the most optimal way to control food fermentation processes. The influence of different fermentation parameters, in particular the substrate availability, on the metabolic flux distribution in dedicated strains of starter cultures can be assessed using genome-scale metabolic models (GEMs). The reconstruction process of such GEMs involves a combination of genome annotation, comparative genomics, and metabolic pathway analysis to set the biological information into a metabolic context. Flux balance analysis (FBA) is an *in silico* method that allows to perform simulations of the metabolic fluxes of a starter culture strain. It can be used to validate the GEM, by comparison to available experimental data, and to predict new features of a starter culture, allowing the formulation of testable hypotheses. The simulations provided by FBA give a static view of the metabolic potential of a starter culture, but inherently the behavior of a starter culture in the food matrix is dynamic. Therefore, a dynamic extension of FBA, called dynamic FBA (dFBA), was applied to capture and predict the interplay between the population and extra-cellular metabolite dynamics with the metabolism of the bacterial cell population, represented by the GEM. By combining these *in silico* techniques, much of the data obtained *in vitro* for candidate functional starter cultures for cocoa bean fermentation processes could be (re-)analyzed, thus revealing metabolic features and systemic properties that were hitherto not yet reported. The methodology developed and knowledge acquired led to a number of deliverables that is expected to allow enhanced applications of functional starter cultures in cocoa bean fermentation processes.

- **A workflow for genome-scale metabolic reconstruction of candidate functional starter culture strains**

In this PhD thesis, a workflow for genome-scale metabolic reconstruction was developed, which was applied to different candidate functional starter culture strains (Figure 1). For *Acetobacter pasteurianus* 386B, a candidate starter culture for cocoa bean fermentation processes, the reconstruction of its metabolism was performed *de novo*, by combining genome re-annotation, comparative genomics, and Pathway/Genome Database (PGDB) curation. This resulted in the iAp386B454 GEM, containing 322 metabolic reactions related to central carbon metabolism, energy generation, and biosynthetic pathways to produce all biomass precursors present in the biomass reaction. The GEM and curated PGDB can therefore be seen as the current knowledge bases of *A. pasteurianus* 386B. They also represent the first genome-scale metabolic model reconstruction for a member of the genus *Acetobacter*. Subsequently, the information stored in the iAp386B454 GEM, in particular the gene-protein-reaction (GPR) association rules, was leveraged to reconstruct GEMs of two related candidate functional starter cultures for cocoa bean fermentation processes, namely *Acetobacter ghanensis* LMG

23848^{T} and *Acetobacter senegalensis* 108B, assuming an identical biomass composition. Finally, core GEM reconstruction of two lactic acid bacteria (LAB) candidate functional starter cultures for cocoa bean fermentation processes, one obligately heterofermentative strain (*Lactobacillus fermentum* 222) and one facultatively heterofermentative strain (*Lactobacillus plantarum* 80) was performed. The *L. fermentum* 222 core GEM is the first for the species *L. fermentum*. All GEMs were reconstructed using the MNXref 3.0 namespace of reactions and metabolites.

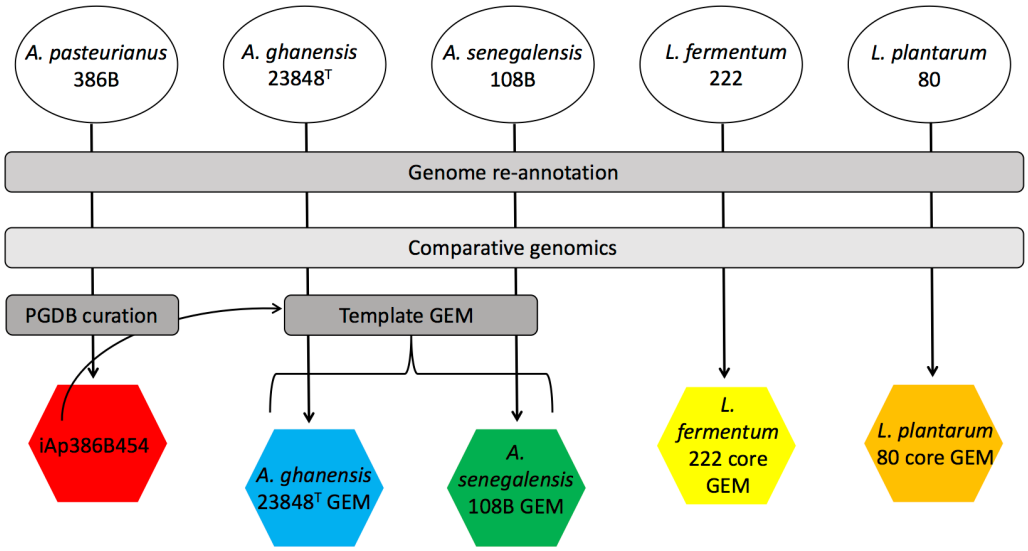


Figure 1. Genome-scale metabolic reconstruction workflow of the candidate functional starter cultures for controlled cocoa bean fermentation processes adopted in this PhD thesis.

The genome-scale metabolic network reconstruction processes showed the added value of combining multiple functional annotation sources. Information from neighbors of the query genes can be retrieved from different biological databases, each with their own biological context and curation processes. In addition, multiple functional annotation tools will look for the presence of known biological functions in the query sequences. Thus, the information obtained from these sources is complementary. Moreover, it is the task and responsibility of the human curator to choose which bioinformatics tools to use and which databases to consult, to obtain a metabolic reconstruction of the microorganism under study that is as complete as possible, given the current state of scientific knowledge. In this PhD thesis, the information collecting procedure was automated as much as possible, but the *de novo* metabolic reconstruction and curation process was an almost completely manual process. It is clear that in the future, the latter should also be automated as much as possible. Tools for automated metabolic reconstruction do exist but they currently do not allow the user to choose its namespace, potentially limiting the scope of the metabolic reconstruction. Further, they heavily rely on comparison with GEMs of microorganisms already reconstructed. Genome-scale metabolic model reconstruction of microorganisms for which there is no GEM of a close relative available is therefore hampered, and what follows is a long and cumbersome curation

process, for which the time gain of the initial automatic reconstruction may become a negligible fraction of the total time spent. Interactive and comprehensive software platforms are needed that integrate functional annotation, GPR associations, and FBA to check for consistency and report conflicts between different metabolic reconstructions. In this way, manual curation may become more targeted, and therefore more instructive in a biological sense. For example, none of the automated metabolic reconstruction tools could retrieve the specific aerobic respiratory chain of the *Acetobacter* species examined. The use of computationally encoded curation rules, taking into account information concerning different aspects of the biological neighborhood, may prove most useful toward reconstruction consistency and quantification of the level of uncertainty in the presented GPR associations.

- **Insights into the metabolic strategies adopted by the candidate functional starter culture strains in the cocoa bean fermentation process**

FBA with the GEMs obtained allowed to test multiple metabolic scenarios *in silico* and compare them to experimental data. For *A. pasteurianus* 386B, previous work indicated that its growth under cocoa fermentation conditions is characterized by two phases (Adler *et al*., 2014; Moens *et al*., 2014). Using the iAp386B454 GEM, the separation of lactate and ethanol catabolism in the first phase of growth of *A. pasteurianus* 386B was investigated in a systemic way, revealing the possible contribution of $NAD(P)^+$ transhydrogenase as a NADPH-producing enzyme. Also, the metabolic flux distribution corresponding to the acetate overoxidation phase of *A. pasteurianus* 386B in the second phase of growth was investigated for the first time. The consumption of pyruvate, accumulated in the growth medium during the first phase of growth, was essential to predict the *in silico* growth. Also, an increase in the tricarboxylic acid (TCA) cycle flux indicated a switch in the respiratory enzymes used compared to the first phase of growth. This was due to the increased NADH production flux by the TCA cycle in the second phase of growth, compared to the increased ubiquinol production flux by the periplasmic dehydrogenase enzymes in the first phase of growth. Most likely reaction kinetics and regulation of the respiratory chain also play a role, meaning that the FBA predictions related to the fluxes of the respiratory enzymes in each phase of growth need to be thoroughly tested experimentally. By performing dFBA, thus forcing the GEM to fit the population and metabolite dynamics, additional information is provided that explains the very low specific growth rate of *A. pasteurianus* 386B in the second phase of growth and its slow oxidation of acetic acid. A possible explanation for this finding is that the low specific growth rate arises from a high growth-associated maintenance energy requirement. Since acetic acid is a toxic compound that is known to induce various acetic acid resistance mechanisms in acetic acid bacteria (AAB), microbial growth in the second phase of growth may demand additional energy. Also, a direct interference of acetic acid in the respiratory chain could occur, as it is known to be a strong uncoupling agent (Matsushita *et al*., 2004; Mitchell, 2011). Thus, the uncoupling effect of acetate may be responsible for the activation of a high-efficiency respiratory chain in the second phase of growth of *A. pasteurianus* 386B, as predicted by dFBA using the iAp386B454 GEM.

The biphasic growth curve of *A. ghanensis* LMG 23848^T strongly resembles that of *A. pasteurianus* 386B, but the consumption of lactic acid by this strain is slower, overlapping for both phases of growth (Adler *et al*., 2014; Moens *et al*., 2014). Dynamic FBA revealed that the slower lactate consumption of *A. ghanensis* LMG 23848^T could be a metabolic strategy to allow subsequent acetate overoxidation. Also, a possible physiological role for the cytochrome bc_1 complex was proposed for *A. ghanensis* LMG 23848^T, as it was the only enzyme that could re-oxidize reduced cytochrome *c* produced by the cytochrome *c*-dependent D-lactate dehydrogenase. For *A. senegalensis* 108B, the presence of citrate lyase indicated the possibility of growth on citric acid. *In silico* FBA combined with *in vitro* growth experiments confirmed that this was indeed the case, linking the possible contribution of citric acid consumption during cocoa bean fermentation for the first time, next to LAB, also to AAB. *In vitro* growth of *A. senegalensis* 108B with ethanol as the sole carbon source was not found, which most probably followed from the lack of the glyoxylate pathway enzymes in the genome of this strain. However, the *in vitro* result was in contrast with the species description of *A. senegalensis*. Comparative genomics using a selection of *Acetobacter* species genomes revealed that the type strains of *A. senegalensis* and *Acetobacter cibinongensis* both missed the glyoxylate pathway. Therefore, the reported growth of these strains on ethanol as the sole carbon source may be false positives (Cleenwerck *et al*., 2008).

Flux balance analysis of the *L. fermentum* 222 and *L. plantarum* 80 GEMs revealed the influence of the Emden-Meyerhof-Parnas pathway, the phosphoketolase pathway, and the pentose phosphate pathway on the energetic and redox balances of these strains. In addition, dFBA revealed that the population and metabolite dynamics of related *L. fermentum* and *L. plantarum* strains under cocoa pulp-simulating conditions may be due to the consumption of undefined complex medium components.

The FBA presented here indicated that, for the AAB strains as well as for the LAB strains, correspondence of *in silico* metabolic flux distributions with *in vitro* experimental data necessitated the addition of production flux constraints, next to substrate consumption flux constraints, when biomass formation was taken as the objective of the simulation. Therefore, it is apparent that these bacterial species exhibited a metabolic strategy that was not optimal in terms of biomass precursor formation and energy generation. This finding highlighted the need of integration of experimental data obtained under cocoa pulp-simulating conditions for the GEMs to become predictive, or at least descriptive using FBA (Figure 2). Future modeling efforts should first benefit from omics experimental data that capture the physiological responses of the candidate functional starter culture strains in cocoa fermentation conditions to gain more mechanistic insights.

- **Use of dynamic flux balance analysis is a step closer to bioprocess modeling of the cocoa bean fermentation process**

The use of GEMs has been proposed to enhance the predictive power of bioprocess models of industrial fermentation processes (Mahadevan *et al*., 2005). Bioprocess models try to capture the space and time dependencies of the state variables that make up the fermentation process.

So far, the metabolism of microorganisms was mostly characterized by black-box models, defining net stoichiometric conversion yields (Heijnen, 2010). dFBA is a bioprocess modeling technique that, unlike black-box models, allows to predict changes in the metabolic flux distribution, resulting in altered stoichiometric yields, in response to the dynamics of the environmental state variables (Höffner *et al*., 2013).

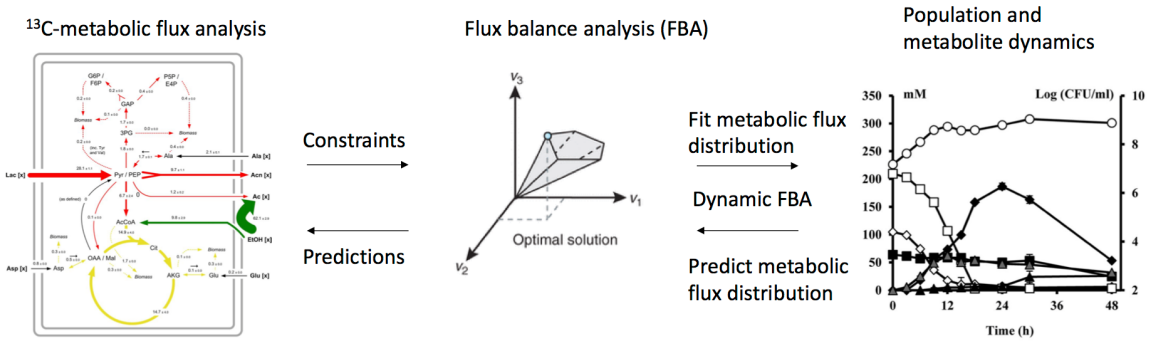


Figure 2. Strategies adopted in this PhD thesis to integrate experimental data (^{13}C-metabolic flux analysis, population and metabolite dynamics) in GEMs to perform (dynamic) FBA. Individual illustrations were obtained from other sources (from left to right: Adler *et al*., 2014; Orth *et al*., 2010; Moens *et al*., 2014).

The results obtained with GEMs provide testable hypotheses, underlining the importance of *in silico* experimentation next to *in vitro* and *in vivo* experimentation. However, to allow model assessment, a rigorous experimental set-up is needed. The use of chemically defined media has been used to test GEM predictions, as it allows to track the consumption of individual substrates in a medium. This, in combination with a targeted or untargeted metabolomics approach, would allow to investigate the dynamics of extracellular metabolites over time. However, to allow comparison to GEM predictions, it is important to keep fermentation parameters not included in the GEM framework constant, such as the temperature and pH. Also, metabolic shifts predicted by the GEM could be further investigated to see whether they correspond to changes in gene expression. Indeed, the integration of transcriptomic data in GEMs has been proposed before (Kim and Lun, 2014).

The gap from *in vitro* to *in vivo* investigation of cocoa bean fermentation processes is currently large and overcoming this gap goes with some difficulties associated to the logistics on the cocoa farms and the limited amount of experimental control. The influence of the use of starter cultures on cocoa bean fermentation processes must be further investigated. This is particularly challenging when non-aseptic conditions are applied, as is mostly the case for these processes, since the fermentations are performed *in situ* at the cocoa farm. Related to this is the question of ecosystem robustness of the autochthonous community that performs the fermentation process. Qualitative interpretation of the current experimental studies revealed a distinctive pattern. Initially, high concentrations of carbohydrates are consumed by yeasts and LAB, leading to the production of mainly ethanol and lactic acid. The AAB species subsequently rapidly consume these substrates, forming mainly acetic acid and acetoin. Especially the

exothermic oxidation of ethanol to acetic acid is thought to lead to subsequent yeast death as well as to the death of the cocoa embryo, allowing the release of the essential cocoa flavor precursors (De Vuyst and Weckx, 2016a). The success of the mere opportunistic behavior of AAB depends on its ability to attain higher reaction rates than its competitors. For example, it is well known that yeasts could perfectly respire the ethanol accumulated, but this is a slow process (Fraenkel, 2011). The fact that this pattern is found in spontaneous cocoa bean fermentation processes all over the world is an indication that these ecosystem interactions are sufficiently robust to overcome variations in the cocoa pulp and natural community composition of the inoculum. To take further advantage of the cocoa bean fermentation dynamics and to further accelerate the process, the selection of strains with rapid substrate consumption and product formation rates is therefore advisable. Thus, strains with a low biomass yield may be preferable in the selection procedure.

To increase control over cocoa bean fermentation processes, starter-culture initiated fermentation processes have been attempted by adding autochthonous or allochthonous starter cultures to the initial microbial ecosystem. Autochthonous starter cultures have the shown advantage to withstand the different environmental constraints typical for cocoa bean fermentation, such as a low pH in the initial stage of the fermentation process and an increasing temperature as the fermentation progresses. However, the contribution of adaptation to their competitiveness may be minor, since the cocoa bean fermentation process is essentially a batch fermentation process, and therefore may not lead to long-term adaptation events of the community members. Allochthonous starter cultures are used to add specific functionalities to the microbial ecosystem, such as a favorable aroma production profile or increased pectinolytic activity. In both cases, to allow the strains to increase their fitness potential, adaptive laboratory evolution could be performed under cocoa fermentation conditions. This is a technique that has been successfully used to obtain superior strains in terms of substrate utilization rate and specific growth rate (Long and Antoniewicz, 2018). In terms of traceability, the use of autochthonous starter cultures forms another experimental hurdle, as it is difficult to discern their genotypes from the related strains present in the microbial ecosystem.

The temperature of the fermentation process is an important parameter that relates both to the microbial community dynamics and cocoa flavor formation. Current experimental studies do not treat this parameter independently from the presence or absence of specific microbial community members, which may lead to difficulties in interpreting the results and doubts in the resulting claims (Ho *et al*., 2014, 2015, 2018). Similarly, the influence of air ingress on the fermentation dynamics has only been assessed in a spontaneous setting (Camu *et al*., 2008).

To sum up, future cocoa bean fermentation experiments should benefit from established experimental design principles instead of *ad hoc* experimentations. Cocoa bean fermentation is a complex fermentation process, whereby, as opposed to more conventional fermented foods, the cocoa flavor precursor molecules formed are the products of the fermentation process rather than the microbial product profiles themselves. Additionally, the downstream value chain to produce cocoa-derived products introduces a large number of additional process-specific parameters. Factors that influence cocoa bean fermentation processes include the fermentation

technique (heaps, boxes, platforms, …), the effect of the microbial communities (environmental microbiota, use of functional starter cultures, use of antimicrobials), the cocoa pulp composition, the temperature, the frequency of mixing, turning and thus air ingress, and the fermentation time. Therefore, the experimental design should treat as much as possible these factors independently and interpret and report the results in a way that treats only the fermentation factors that are to be tested. To assess the influence of cocoa bean fermentation processes on the quality of the chocolates produced thereof, standardized drying, roasting and conching protocols need to be employed too.

In the future, the use of bioprocess models of cocoa bean fermentation processes may allow to simulate their optimization. In this framework, microbial GEMs and a GEM of the cocoa seed could be combined, enabling *in silico* simulation of the metabolic interactions between microbial community members and the cocoa beans. This would be a useful tool to assess the influence of various fermentation factors *in silico*, to interpret experimental data, and to inform the set-up of new experiments. For example, a collection of GEMs coupled to a dynamic reactive-transport model of partial differential equations may describe the space and time dependency of the microbial dynamics on the oxygen gradient in a fermenting cocoa bean heap. An additional 'turning' parameter could then be defined that controls the frequency of air ingress. However, successful bioprocess modeling requires consistent measurements of all fermentation factors, leading to high-dimensional data sets. For example, meta-transcriptomics data may be integrated in the GEMs as well. Additionally, further research is needed to identify appropriate objective functions for the GEMs to accurately simulate the metabolic behavior of the microbial community members.

Due to the complexity of cocoa bean fermentation processes, it is clear that for successful, large-scale cocoa production to become possible, technological improvements will be needed as well as specialized craft workers along the whole value chain. Therefore, current cocoa farmers should empower themselves to enable the consistent production of well-fermented and well-flavored cocoa beans. The use of functional starter cultures could help them to obtain this target and in that regard the results obtained in this PhD thesis provide some guidelines to select starter cultures strains with appropriate functionalities; such as a lack of the glyoxylate pathway, which was the case for all AAB candidate functional starter culture strains investigated, leading to increased acetate production rates, increased lactate consumption rates using a quinone-dependent D-lactate dehydrogenase (EC 1.1.5.12, *A. pasteurianus* 386B), and superior ethanol consumption fluxes (*A. ghanensis* LMG 23848^{T}). Fine-tuning of process-specific parameters, such as increasing the turning frequency in the second phase of growth of the AAB strains to induce acetate overoxidation, may be further needed to adequately treat the functional starter cultures used and allow them to reach their full metabolic potential during the whole cocoa bean fermentation process.

References

Abdel-Haleem, A. M., Hefzi, H., Mineta, K., Gao, X., Gojobori, T., Palsson, B. Ø., Lewis, N. E. & Jamshidi, N. (2018). Functional interrogation of *Plasmodium* genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Computational Biology* **14**, e1005895.

Abram, F. (2015). Systems-based approaches to unravel multi-species microbial community functioning. *Computational and Structural Biotechnology Journal* **13**, 24–32.

Adachi, O., Ano, Y., Toyama, H. & Matsushita, K. (2008). A novel 3-dehydroquinate dehydratase catalyzing extracellular formation of 3-dehydroshikimate by oxidative fermentation of *Gluconobacter oxydans* IFO 3244. *Bioscience, Biotechnology, and Biochemistry* **72**, 1475-1482.

Adachi, O. & Yakushi, T. (2016). Membrane-bound dehydrogenases of acetic acid bacteria, pp. 273-297. *In* Matsushita, K., Toyama, H., Tonouchi, N., Okamoto-Kainuma, A., (Eds.), Acetic Acid Bacteria Ecology and Physiology. Springer Japan, Tokyo, Japan.

Addou, S., Rentzsch, R., Lee, D. & Orengo, C. A. (2009). Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *Journal of Molecular Biology* **387**, 416–430.

Adler, P., Bolten, C. J., Dohnt, K., Hansen, C. E. & Wittmann, C. (2013). Core fluxome and metafluxome of lactic acid bacteria under simulated cocoa pulp fermentation conditions. *Applied and Environmental Microbiology* **79**, 5670-5681.

Adler, P., Frey, L. J., Berger, A., Bolten, C. J., Hansen, C. E. & Wittmann, C. (2014). The key to acetate: metabolic fluxes of acetic acid bacteria under cocoa pulp fermentation-simulating conditions. *Applied and Environmental Microbiology* **80**, 4702–4716.

Afoakwa, E. O. (2016). *Chocolate Science and Technology*. Ames: Wiley-Blackwell.

Alber, B. E., Spanheimer, R., Ebenau-Jehle, C. & Fuchs, G. (2006). Study of an alternate glyoxylate cycle for acetate assimilation by *Rhodobacter sphaeroides*. *Molecular Microbiology* **61**, 297-309.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Ameyama, M., Matsushita, K., Ohno, Y., Shinagawa, E. & Adachi, O. (1981). Existence of a novel prosthetic group, PQQ, in membrane-bound, electron transport chain-linked, primary dehydrogenases of oxidative bacteria. *FEBS Letters* **130**, 179–183.

Ang, K. S., Lakshmanan, M., Lee, N.-R. & Lee, D.-Y. (2018). Metabolic modeling of microbial community interactions for health, environmental and biotechnological applications. *Current Genomics* **19**, 712-722.

Antoniewicz, M. R. (2013). Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks. *Current Opinion in Biotechnology* **24**, 973–978.

Aprotosoaie, A. C., Luca, S. V. & Miron, A. (2016). Flavor chemistry of cocoa and cocoa products-an overview. *Comprehensive Reviews in Food Science and Food Safety* **15**, 73–91.

Araujo, F. A., Barh, D., Silva, A., Guimarães, L. & Ramos, R. T. J. (2018). GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific Reports* **8**, 1794.

Arkin, A. P. et al. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology* **36**, 566–569.

Armenta-Medina, D., Segovia, L. & Perez-Rueda, E. (2014). Comparative genomics of nucleotide metabolism: a tour to the past of the three cellular domains of life. *BMC Genomics* **15**, 800.

Aung, H. W., Henry, S. A. & Walker, L. P. (2013). Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology* **9**, 215–228.

Axelsson, L. (2004). Lactic acid bacteria: classification and physiology, pp. 1-66. *In* Salminen, S., von Wright, A., & Ouwehand, A., (Eds.), *Lactic acid bacteria: Microbiological and Functional Aspects*, third edition. Marcel Dekker, New York, United States of America.

Aziz, R. K. et al. (2008). The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**, 75.

Aziz, R. K. et al. (2012). SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. *PLoS ONE* **7**, e48053.

Bar, R., Gainer, J. L. & Kirwan, D. J. (1987). An unusual pattern of product inhibition: batch acetic acid fermentation. *Biotechnology and Bioengeering* **29**, 796-798.

Barak, Z. & Chipman, D. M. (2012). Allosteric regulation in acetohydroxyacid synthases (AHASs) - different structures and kinetic behavior in isozymes in the same organisms. *Archives of Biochemistry and Biophysics* **519**, 167-174.

Barata A., Malfeito-Ferreira M. & Loureiro V., 2012. The microbial ecology of wine grape berries. *International Journal of Food Microbiology* **153**, 243-259.

Barišić, V., Kopjar, M., Jozinović, A., Flanjak, I., Ačkar, Đ., Miličević, B., Šubarić, D., Jokić, S. & Babić, J. (2019). The chemistry behind chocolate production. *Molecules* **24**, 3163.

Barreteau, H., Kova, A., Boniface, A., Sova, M., Gobec, S. & Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiology Reviews* **32**, 168-207.

Belly, R. T., & Claus, G. W. (1972). Effect of amino acids on the growth of *Acetobacter suboxydans*. *Archiv für Mikrobiologie* **83**, 237–245.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers E. W. (2017). GenBank. *Nucleic Acids Research* **45**, D37–D42.

Bernard, T., Bridge, A., Morgat, A., Moretti, S., Xenarios, I. & Pagni, M. (2014). Reconciliation of metabolites and biochemical reactions for metabolic networks. *Briefings in Bioinformatics* **15**, 123–135.

Bertrand, R. L. (2019). Lag phase is a dynamic, organized, adaptive, and evolvable period that prepares bacteria for cell division. *Journal of Bacteriology* **201**, e00697-18.

Bertuzzi, A. S., Walsh, A. M., Sheehan, J. J., Cotter, P. D., Crispie, F., McSweeney, P. L. H., Kilcawley, K. N. & Rea, M. C. (2018). Omics-based insights into flavor development and microbial succession within surface-ripened cheese. *mSystems* **3**, e00211-17.

Blasche, S., Kim, Y., Oliveira, A. P. & Patil, K. R. (2017). Model microbial communities for ecosystems biology. *Current Opinion in Systems Biology* **6**, 51–57.

Blount, Z. D. (2015). The unexhausted potential of *E. coli*. *eLife* **4**, e05826.

Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. (2008). LibSBML: an API library for SBML. *Bioinformatics* **24**, 880-881.

Bortolini, C., Patrone, V., Puglisi, E. & Morelli, L. (2016). Detailed analyses of the bacterial populations in processed cocoa beans of different geographic origin, subject to varied fermentation conditions. *International Journal of Food Microbiology* **236**, 98–106.

Brenner, S. E. (1999). Errors in genome annotation. *Trends in Genetics* **15**, 132–133.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olsen, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R. & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports* **5**, 8365.

Bruggeman, F. J., & Westerhoff, H. V. (2007). The nature of systems biology. *Trends in Microbiology* **15**, 45–50.

Burge, S., Kelly, E., Lonsdale, D., Mutowo-Muellenet, P., McAnulla, C., Mitchell, A., Sangrador-Vegas, A., Yong, S-Y., Mulder, N. & Hunter, S. (2012). Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation. *Database* **2012**, bar068.

Camu, N., De Winter, T., Verbrugghe, K., Cleenwerck, I., Vandamme, P., Takrama, J.S., Vancanneyt, M. & De Vuyst, L. (2007). Dynamics and biodiversity of populations of lactic acid bacteria and acetic acid bacteria involved in spontaneous heap fermentation of cocoa beans in Ghana. *Applied and Environmental Microbiology* **73**, 1809–1824.

Camu N., González A., De Winter, T., Van Schoor, A., De Bruyne, K., Vandamme, P., Takrama, J.S., Addo, S.K. & De Vuyst L. (2008). Influence of turning and environmental contamination on the dynamics of populations of lactic acid and acetic acid bacteria involved in spontaneous cocoa bean heap fermentation in Ghana. *Applied and Environmental Microbiology* **74**, 86-98.

Carr, R., & Borenstein, E. (2012). NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. *Bioinformatics* **28**, 734–735.

Caspi, R. et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleid Acids Research* **42**, D459–D471.

Caspi, R. et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleid Acids Research* **44**, D471–D480.

Caspeta, L., Shoaie, S., Agren, R., Nookaew, I. & Nielsen, J. (2012). Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and *in-silico* evaluation of their potentials. *BMC Systems Biology* **6**, 24.

Chiappino-Pepe, A., Tymoshenko, S., Ataman, M., Soldati-Favre, D. & Hatzimanikatis, V. (2017). Bioenergetics-based modeling of *Plasmodium falciparum* metabolism reveals its

essential genes, nutritional requirements, and thermodynamic bottlenecks. *PLoS Computational Biology* **13**, e1005397.

Chinnawirotpisan P., Matsushita K., Toyama H., Adachi O., Limtong S. & Theeragool G., 2003. Purification and characterization of two NAD-dependent alcohol dehydrogenases (ADHs) induced in the quinoprotein ADH-deficient mutant of *Acetobacter pasteurianus* SKU1108. *Bioscience, Biotechnology, and Biochemistry* **67**, 958-965.

Claudel-Renard, C., Chevalet, C., Faraut, T. & Kahn, D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research* **31**, 6633-6639.

Cleenwerck, I., Vandemeulebroecke, K., Janssens, D. & Swings, J. (2002). Re-examination of the genus *Acetobacter*, with descriptions of *Acetobacter cerevisiae* sp. nov. and *Acetobacter malorum* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1551-1558.

Cleenwerck, I. & De Vos, P. (2008). Polyphasic taxonomy of acetic acid bacteria: An overview of the currently applied methodology. *International Journal of Food Microbiology* **125**, 2–14.

Cleenwerck, I., Gonzalez, A., Camu, N., Engelbeen, K., De Vos, P. & De Vuyst, L. (2008). *Acetobacter fabarum* sp. nov., an acetic acid bacterium from a Ghanaian cocoa bean heap fermentation. *International Journal of Systematic and Evolutionary Microbiology* **58**, 2180-2185.

Coe, S. D. & Coe, M. D. (2013). *The True History of Chocolate*. Thames & Hudson, London, United Kingdom.

Colombo, M. L., Pinorini-Godly, M. T. & Ario, C. (2012). Botany and pharmacognosy of the cocoa tree, pp. 41-62. *In* Paoletti, R., Poli, A., Conti, A., & Visioli, F., (Eds.), *Chocolate and Health*. Springer, Milan, Italy.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

Coton, M., Pawtowski, A., Taminiau, B., Burgaud, G., Deniel, F., Coulloumme-Labarthe, L., Fall, A., Daube, G. & Coton, E. (2017). Unraveling microbial ecology of industrial-scale Kombucha fermentations by metabarcoding and culture-based methods. *FEMS Microbiology Ecology* **93**, 5.

Crafack, M., Mikkelsen, M.B., Saerens, S., Knudsen, M., Blennow, A., Lowor, S., Takrama, J., Swiegers, J. H., Petersen, G. B., Heimdal, H. & Nielsen, D. S. (2013). Influencing cocoa flavour using *Pichia kluyveri* and *Kluyveromyces marxianus* in a defined mixed starter culture for cocoa fermentation. *International Journal of Food Microbiology* **167**, 103–116.

da Veiga Moreira, I. M., de Figueiredo Vilela, L., da Cruz Pedroso Miguel, M., Santos, C., Lima, N. & Schwan, R. F. (2017). Impact of a microbial cocktail used as a starter culture on cocoa fermentation and chocolate flavor. *Molecules* **22**, 766.

da Cruz Pedrozo Miguel, M. G., Viana de Castro Reis, L., Efraim, P., Santos, C., Lima, N. & Schwan, R. F. (2017). Cocoa fermentation: microbial identification by MALDI-TOF

MS, and sensory evaluation of produced chocolate. *LWT – Food Science and Technology* **77**, 362–369.

Dale, J. M., Popescu, L. & Karp, P. D. (2010). Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics* **11**, 15.

Danchin, A (2009). Bacteria as computers making computers. *FEMS Microbiology Reviews* **33**, 3–26.

Danchin, A. & Fang, G. (2016). Unknown unknowns: essential genes in quest for function. *Microbial Biotechnology* **9**, 530–540.

Danchin, A., Ouzounis, C., Tokuyasu, T. & Zucker, J-D. (2018). No wisdom in the crowd: genome annotation in the era of big data - current status and future prospects. *Microbial Biotechnology* **11**, 588–605.

Daniel, H-M., Vrancken, G., Takrama, J. F., Camu, N., De Vos, P. & De Vuyst, L. (2009). Yeast diversity of Ghanaian cocoa bean heap fermentations. *FEMS Yeast Research* **9**, 774–783.

De Filippis, F., Parente, E. & Ercolini, D. (2017). Metagenomics insights into food fermentations. *Microbial Biotechnology* **10**, 91–102.

de Jong, H., Casagranda, S., Giordano, N., Cinquemani, E., Ropers, D., Geiselmann, J. & Gouzé, J-L. (2017). Mathematical modelling of microbes: metabolism, gene expression and growth. *Journal of The Royal Society Interface* **14**, 20170502.

De Ley, J. (1959). On the formation of acetoin by *Acetobacter*. *Journal of General Microbiology* **21**, 352–365.

de Melo Pereira, G. V., da Cruz Pedrozo Miguel, M. G., Ramos, C. L. & Schwan, R. F. (2012). Microbiological and physicochemical characterization of small-scale cocoa fermentations and screening of yeast and bacterial strains to develop a defined starter culture. *Applied and Environmental Microbiology* **78**, 5395–5405.

de Melo Pereira, G. V., Soccol, V. T. & Soccol, C. R. (2016). Current state of research on cocoa and coffee fermentations. *Current Opinion in Food Science* **7**, 50–57.

De Mey, M., De Maeseneire, S., Soetaert, W. & Vandamme, E. (2007). Minimizing acetate formation in *E. coli* fermentations. *Journal of Industrial Microbiology and Biotechnology* **34**, 689-700.

De Roos, J. & De Vuyst, L. (2018). Acetic acid bacteria in fermented foods and beverages. *Current Opinion in Biotechnology* **49**, 115-119.

De Taeye, C., Eyamo Evina, V. J., Caullet, G., Niemenak, N. & Collin, S. (2016). Fate of anthocyanins through cocoa fermentation. Emergence of new polyphenolic dimers. *Journal of Agricultural and Food Chemistry* **64**, 8876–8885.

De Vuyst, L. & Weckx, S. (2016a). The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *Journal of Applied Microbiology* **121**, 5–17.

De Vuyst, L. & Weckx, S. (2016b). The functional role of lactic acid bacteria in cocoa bean fermentation, pp. 248-278. *In* Mozzi, M., Raya, R. R., & Vignolo, M. M., (Eds.), *Biotechnology of Lactic Acid Bacteria: Novel Applications*. Wiley Blackwell, New Jersey, United States of America.

Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679.

Deppenmeier, U. & Ehrenreich, A. (2009). Physiology of acetic acid bacteria in light of the genome sequence of *Gluconobacter oxydans*. *Journal of Molecular Microbiology and Biotechnology* **16**, 69–80.

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**, 776–790.

Dong, J. M., Taylor, J. S., Latour, D. J., Iuchi, S. & Lin, E. C. (1993). Three overlapping *lct* genes involved in L-lactate utilization by *Escherichia coli*. *Journal of Bacteriology* **175**, 6671-6678.

Dudley, E. G. & Steele, J. L. (2005). Succinate production and citrate catabolism by Cheddar cheese nonstarter lactobacilli. *Journal of Applied Microbiology* **98**, 1423.

D'Souza, R. N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M. S. & Kuhnert, N. (2017). Origin-based polyphenolic fingerprinting of *Theobroma cacao* in unfermented and fermented beans. *Food Research International* **99**, 550–559.

Ebrahim, A., Lerman, J. A., Palsson, B. Ø. & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7**, 74.

Edirisinghe, J. N., Weisenhorn, P., Conrad, N., Xia, F., Overbeek, R., Stevens, R. L. & Henry, C.S. (2016). Modeling central metabolism and energy biosynthesis across microbial life. *BMC Genomics* **17**, 568.

Edwards, J. S., Ramakrishna, R. & Palsson, B. Ø. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnology and Bioengineering* **77**, 27 - 36.

Elbourne, L. D. H., Tetu, S. G., Hassan, K. A. & Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Research* **45**, D320-D324.

Emms, D. M. & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157.

Emms, D. M. & Kelly, S. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. *Molecular Biology and Evolution* **34**, 3267-3278.

Emms, D. M. & Kelly, S. (2018). STAG: Species Tree Inference from All Genes. *bioRxiv*, doi: 10.1101/267914.

Endo, A., Maeno, S., Tanizawa, Y., Kneifel, W., Arita, M., Dicks, L. & Salminen, S. (2018). Fructophilic lactic acid bacteria, a unique group of fructose-fermenting microbes. *Applied and Environmental Microbiology* **84**, e01290-18.

Engeseth, N. J. and Ac Pangan, M. F. (2018). Current context on chocolate flavor development — a review. *Current Opinion in Food Science* **21**, 84–91.

Enjalbert, B., Cocaign-Bousquet, M., Portais, J.-C. & Letisse, F. (2015). Acetate exposure determines the diauxic behavior of *Escherichia coli* during the glucose-acetate transition. *Journal of Bacteriology* **197**, 3173-3181.

Ercolini, D. (2013). High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied and Environmental Microbiology* **79**, 3148–3155.

Falk, G. & Walker, J. E. (1988). DNA sequence of a gene cluster coding for subunits of the F_0 membrane sector of ATP synthase in *Rhodospirillum rubrum*. Support for modular evolution of the F_1 and F_0 sectors. *The Biochemical Journal* **254**, 109-122.

Faust, K., Croes, D. & van Helden, J. (2011). Prediction of metabolic pathways from genome-scale metabolic networks. *Biosystems* **105**, 109–121.

Fayeulle, N., Vallverdu-Queralt, A., Meudec, E., Hue, C., Boulanger, R., Cheynier, V. & Sommerer, N. (2018). Characterization of new flavan-3-ol derivatives in fermented cocoa beans. *Food Chemistry* **259**, 207–212.

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**, 129–143.

Feist, A. M. & Palsson, B. Ø. (2010). The biomass objective function. *Current Opinion in Microbiology* **13**, 344–349.

Feist, A. M. & Palsson, B. Ø. (2016). What do cells actually want? *Genome Biology* **17**, 110.

Feng, Y. & Cronan, J. E. (2009). *Escherichia coli* unsaturated fatty acid synthesis. *Journal of Biological Chemistry* **284**, 29526-29535.

Figueroa-Hernández, C., Mota-Gutierrez, J., Ferrocino, I., Hernández-Estrada, Z. J., González-Ríos, O., Cocolin, L. & Suárez-Quiroz, M. L. (2019). The challenges and perspectives of the selection of starter cultures for fermented cocoa beans. *International Journal of Food Microbiology* **301**, 41–50.

Fishov, I., Zaritsky, A. & Grover, N. B. (1995). On microbial states of growth. *Molecular Microbiology* **15**, 789–794.

Flahaut, N. A. L., Wiersma, A., van de Bunt, B., Martens, D. E., Schaap, P. J., Sijtsma, L., Martins dos Santos, V. A. & de Vos, W. M. (2013). Genome-scale metabolic model for *Lactococcus lactis* MG1363 and its application to the analysis of flavor formation. *Applied Microbiology and Biotechnology* **97**, 8729–8739.

Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 10039–10044.

Fleischmann, A. (2004). IntEnz, the integrated relational enzyme database. *Nucleic Acids Research* **32**, D434 – D437.

Fraenkel, D. G. (2011). *Yeast Intermediary Metabolism*. Cold Spring Harbor Laboratory Press, New York, United States of America.

Francke, C., Siezen, R. J. & Teusink, B. (2005). Reconstructing the metabolic network of a bacterium from its genome. *Trends in Microbiology* **13**, 550-558.

Frébortova, J., Matsushita, K., Arata, H. & Adachi, O. (1998). Intramolecular electron transport in quinoprotein alcohol dehydrogenase of *Acetobacter methanolicus*: a redox-titration study. *Biochimica et Biophysica Acta* **1363**, 24–34.

Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R.M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290, 481-486.

Ganter, M., Bernard, T., Moretti, S., Stelling, J. & Pagni, M. (2013). MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* **29**, 815-816.

Gänzle, M. G. (2015). Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage. *Current Opinion in Food Science* **2**, 106–117.

Geiger, O., López-Lara, I. M. & Sohlenkamp, C. (2013). Phosphatidylcholine biosynthesis and function in bacteria. *Biochimica et Biophysica Acta – Molecular and Cell Biology Lipids* 1831, 503-513.

Gelfand, M. S. & Rodionov, D. A. (2008). Comparative genomics and functional annotation of bacterial transporters. *Physics of Life Reviews* **5**, 22-49.

Gill, A. (2017). The importance of bacterial culture to food microbiology in the age of genomics. *Frontiers in Microbiology* **8**, 777.

Gómez, J. M. & Cantero, D. (1998). Kinetics of substrate consumption and product formation in closed acetic fermentation systems. *Bioprocess Engineering* **18**, 439.

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

Gosselé, F., Van den Mooter, M., Verdonck, L., Swings, J., & De Ley, J. (1981). The nitrogen requirements of *Gluconobacter*, *Acetobacter* and *Frateuria*. *Antonie Van Leeuwenhoek* **47**, 289–296.

Gottstein, W., Olivier, B. G., Bruggeman, F. J. & Teusink, B. (2016). Constraint-based stoichiometric modelling from single organisms to microbial communities. *Journal of The Royal Society Interface* **13**, 20160627.

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**, 121.

Gudmundsson, S. & Thiele, I. (2010). Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489.

Gunawan, C., Satianegara, G., Chen, A. K., Breuer, M., Hauer, B., Rogers, P. L. & Rosche, B. (2007). Yeast pyruvate decarboxylases: variation in biocatalytic characteristics for (*R*)-phenylacetylcarbinol production. *FEMS Yeast Research* **7**, 33–39.

Gutiérrez, T. J. (2017). State-of-the-art chocolate manufacture: a review. *Comprehensive Reviews in Food Science and Food Safety* **16**, 1313–1344.

Guzmán, G. I., Utrilla, J., Nurk, S., Brunk, E., Monk, J. M., Ebrahim, A., Palsson, B. Ø. & Feist, A. M. (2015). Model-driven discovery of underground metabolic functions in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 929–934.

Hagman, A., Säll, T., Compagno, C. & Piskur, J. (2013). Yeast "make-accumulate-consume" life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS ONE* **8**, e68734.

Hanada, T., Kashima, Y., Kosugi, A., Koizumi, Y., Yanagida, F. & Udaka, S. (2001). A gene encoding phosphatidylethanolamine N-methyltransferase from *Acetobacter aceti* and some properties of its disruptant. *Bioscience, Biotechnology, and Biochemistry* **65**, 2741-2748.

Hanke, T., Nöh, K., Noack, S., Polen, T., Bringer, S., Sahm, H., Wiechert, W. & Bott, M. (2013). Combined fluxomics and transcriptomics analysis of glucose catabolism via a partially cyclic pentose phosphate pathway in *Gluconobacter oxydans* 621H. *Applied and Environmental Microbiology* **79**, 2336–2348.

Hanke, T., Richhardt, J., Polen, T., Sahm, H., Bringer, S. & Bott, M. (2012). Influence of oxygen limitation, absence of the cytochrome bc_1 complex and low pH on global gene expression in *Gluconobacter oxydans* 621H using DNA microarray technology. *Journal of Biotechnology* **157**, 359372.

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**, 376–385.

Hansen, E. B. (2002). Commercial bacterial starter cultures for fermented foods of the future. *International Journal of Food Microbiology* **78**, 119–131.

Hara, Y., Seki, M., Matsuoka, S., Hara, H., Yamashita, A. & Matsumoto, K. (2008). Involvement of PlsX and the acyl-phosphate dependent *sn*-glycerol-3-phosphate acyltransferase PlsY in the initial stage of glycerolipid synthesis in *Bacillus subtilis*. *Genes and Genetic Systems* **83**, 433-442.

Hastings, J. (2017). Primer on ontologies, pp. 3-13. *In* Dessimoz, C., Škunca, N., (Eds.), *The Gene Ontology Handbook*. Springer Open, New York, United States of America.

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

Heijnen, J. J. (2010). Impact of thermodynamic principles in systems biology, pp. 139–162. *In* Wittmann, C., Krull, R. (Eds.), *Biosystems Engineering II*. Springer, Berlin, Germany.

Helanto, M., Aarnikunnas, J., Palva, A., Leisola, M. & Nyyssölä, A. (2006). Characterization of genes involved in fructose utilization by *Lactobacillus fermentum*. *Archives of Microbiology* **186**, 51-59.

Henry, C. S., Zinner, J. F., Cohoon, M. P. & Stevens, R. L. (2009). iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biology* **10**, R69.

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**, 977–982.

Henry, C. S., Lerma-Ortiz, C., Gerdes, S. Y., Mullen, J. D., Colasanti, R., Zhukov, R., Niehaus, T. D., Hasnain, G., Conrad, N., Hanson, A. D. & de Crécy-Lagard, V. (2016). Systematic identification and analysis of frequent gene fusion events in metabolic pathways. *BMC Genomics* **17**, 473.

Heo, M.-S., & Son, H.-J. (2002). Development of an optimized, simple chemically defined medium for bacterial cellulose production by *Acetobacter* sp. A9 in shaking cultures. *Biotechnology and Applied Biochemistry* **36**, 41–45.

Herve-Jimenez, L., Guillouard, I., Guedon, E., Boudebbouze, S., Hols, P., Monnet, V., Maguin, E. & Rul, F. (2009). Postgenomic analysis of *Streptococcus thermophilus* cocultivated in milk with *Lactobacillus delbrueckii subsp. bulgaricus*: involvement of nitrogen, purine, and iron metabolism. *Applied and Environmental Microbiology* **75**, 2062–2073.

Ho V.T.T., Zhao J. & Fleet G. (2014). Yeasts are essential for cocoa bean fermentation. *International Journal of Food Microbiology* **174**, 72-87.

Ho, V. T. T., Zhao, J. & Fleet, G. (2015). The effect of lactic acid bacteria on cocoa bean fermentation. *International Journal of Food Microbiology* **205**, 54-67.

Ho, V. T. T., Fleet, G. H. & Zhao, J. (2018). Unravelling the contribution of lactic acid bacteria and acetic acid bacteria to cocoa fermentation using inoculated organisms. *International Journal of Food Microbiology* **279**, 43–56.

Höffner, K., Harwood, S. M. & Barton, P. I. (2013). A reliable simulator for dynamic flux balance analysis. *Biotechnology and Bioengeering* **110**, 792–802.

Hucka, M. et al. (2003). The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531.

Hucka, M. et al. (2018). The Systems Biology Markup Language (SBML): language specification for Level 3 Version 2 Core. *Journal of Integrative Bioinformatics* **15**.

Huerta-Cepas, J. et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**, D286–D293.

Huerta-Cepas, J. et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314.

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**, 2115-2122.

Huntemann, M. et al. (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Standards in Genomic Sciences* **10**, 86.

Hwang, B.-J., Yeom, H.-J., Kim, Y. & Lee, H.-S. (2002). *Corynebacterium glutamicum* utilizes both transsulfuration and direct sulfhydrylation pathways for methionine biosynthesis. *Journal of Bacteriology* **184**, 1277-1286.

Hyatt, D., Chen, G-L., LoCascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119.

Ibarra, R. U., Edwards, J. S. & Palsson, B. Ø. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189.

ICCO (International Cocoa Organization) (2017). Quarterly bulletin of cocoa statistics, Vol XLIII - No. 3 - Cocoa year 2016/2017.

Illeghems, K., De Vuyst, L., Papalexandratou, Z. & Weckx, S. (2012). Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. *PLoS ONE* **7**, e38040.

Illeghems, K., De Vuyst, L. & Weckx, S. (2013). Complete genome sequence and comparative analysis of *Acetobacter pasteurianus* 386B, a strain well-adapted to the cocoa bean fermentation ecosystem. *BMC Genomics* **14**, 526.

Illeghems, K., De Vuyst, L. & Weckx, S. (2015). Comparative genome analysis of the candidate functional starter culture strains *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 for controlled cocoa bean fermentation processes. *BMC Genomics* **16**, 766.

Illeghems, K., Pelicaen, R., De Vuyst, L. & Weckx, S. (2016). Assessment of the contribution of cocoa-derived strains of *Acetobacter ghanensis* and *Acetobacter senegalensis* to the cocoa bean fermentation process through a genomic approach. *Food Microbiology* **58**, 68-78.

Janßen, H. & Steinbüchel, A. (2014). Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels* **7**, 7.

Jensen, P.A., Lutz, K.A. & Papin, J.A. (2011). TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Systems Biology* **5,** 147.

Johansen, E. (2018). Use of natural selection and evolution to develop new starter cultures for fermented foods. *Annual Review of Food Science and Technology* **9**, 411–428.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. Available at: http://www.scipy.org/.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J, Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240.

Jun, S., Si, F., Pugatch, R. & Scott, M. (2018). Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. *Reports on Progress in Physics* **81**, 056601.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, D353–D361.

Kanehisa, M., Sato, Y. & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology* **428**, 726–731.

Kang, T. S., Korber, D. R. & Tanaka, T. (2013). Contributions of citrate in redox potential maintenance and ATP production: metabolic pathways and their regulation in *Lactobacillus panis* PM1. *Applied Microbiology and Biotechnology* **97**, 8693–8703.

Karp, P. D. et al. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* **11**, 40–79.

Karp, P. D. et al. (2016). Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics* **17**, 877–890.

Karp, P. D., Latendresse, M. & Caspi, R. (2011). The Pathway Tools pathway prediction algorithm. *Standards in Genomic Science* **5**, 424–429.

Karp, P. D., Riley, M., Paley, S. M. & Pellegrini-Toole, A. (2002). The MetaCyc Database. *Nucleic Acids Research* **30**, 59–61.

Karsch-Mizrachi, I., Takagi, T., Cochrane, G., and on behalf of the International Nucleotide Sequence Database Collaboration (2018). The international nucleotide sequence database collaboration. *Nucleic Acids Research* **46**, D48–D51.

Kavvas, E. S, Seif, Y. Yurkovich, J. T., Norsigian, C., Poudel, S., Greenwald, W. W., Ghatak, S., Palsson, B. Ø & Monk, J. M. (2018). Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Systems Biology* **12**, 25.

Keenan, T. W. & Lindsay, R. C. (1968). Diacetyl production and utilization by *Lactobacillus* species. *Journal of Dairy Science* **51**, 188–191.

Kerkhoven, E. J., Pomraning, K. R., Baker, S. E. & Nielsen, J. (2016). Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. *Nature partner journal Systems Biology and Applications* **2**, 16005.

Keseler, IM et al. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research* **45**, D543-D550.

Kettle, H., Holtrop, G., Louis, P. & Flint, H. J. (2018). microPop: Modelling microbial populations and communities in R. *Methods in Ecology and Evolution* **9**, 399-409.

Khandelwal, R. A., Olivier, B. G., Röling, W. F. M., Teusink, B. & Bruggeman, F. J. (2013). Community flux balance analysis for microbial consortia at balanced growth. *PLoS ONE* **8**, e64567.

Khomyakova, M., Bukmez, O., Thomas, L. K., Erb, T. J. & Berg, I. A. (2011). A Methylaspartate cycle in Haloarchaea. *Science* **331**, 334-337.

Kilstrup, M., Hammer, K., Ruhdaljensen, P. & Martinussen, J. (2005). Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiology Reviews* **29**, 555-590.

Kim, M. K. & Lun, D. S. (2014). Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal* **11**, 59–65.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, L. & Bolton, E. E. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research* **47**, D1102-D1109.

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.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. Ø. (2015). Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Compututational Biology* **11**, e1004321.

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. Ø. & Lewis, N. E. (2016). BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **44**, D515–D522.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W. E. J., Stiekema, W., Klein Lankhorts, R. M., Bron, P. A., Hoffer, S. M., Nierop Groot, M. N., Kerkhoven, R., de Vries, M., Ursing, B., de Vos, W. M. & Siezen R. J. (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1990-1995.

Klitgord, N. & Segrè, D. (2011). Ecosystems biology of microbial metabolism. *Current Opinion in Biotechnology* **22**, 541–546.

Koduru, L., Kim, Y., Bang, J., Lakshmanan, M., Han, N. S. & Lee, D.-Y. (2017). Genome-scale modeling and transcriptome analysis of *Leuconostoc mesenteroides* unravel the redox governed metabolic states in obligate heterofermentative lactic acid bacteria. *Scientific Reports* **7**, 15721.

Kohlstedt, M., Becker, J. & Wittmann, C. (2010). Metabolic fluxes and beyond—systems biology understanding and engineering of microbial metabolism. *Applied Microbiology and Biotechnology* **88**, 1065–1075.

Kongor, J. E., Hinneh, M., de Walle, D. V., Afoakwa, E. O., Boeckx, P. & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile — a review. *Food Research International* **82**, 44–52.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics* **39**, 309–338.

Koonin, E. V. & Galperin, M. Y. (2003). *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. Springer Nature, Berlin, Germany.

Kornberg, H. (1986). The roles of HPr and FPr in the utilization of fructose by *Escherichia coli*. *FEBS Letters* **194**, 12-15.

Kort, R., Westerik, N., Serrano, L. M., Douillard, F. P., Gottstein, W., Mukisa, I. M., Tuijn, C. J., Basten, L., Hafkamp, B., Meijer, W. C., Teusink, B., de Vos, W. M., Reid, G. & Sybesma, W. (2015). A novel consortium of *Lactobacillus rhamnosus* and *Streptococcus thermophilus* for increased access to functional fermented foods. *Microbial Cell Factories* **14**, 195.

Koskinen, P., Törönen, P., Nokso-Koivisto, J. & Holm, L. (2015). PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* **31**, 1544–1552.

Kresnowati, M. T. A. P., Suryani, L. & Affifah, M. (2013). Improvement of cocoa beans fermentation by LAB starter addition. *Journal of Medical and Bioengineering* **2**, 274–278.

Krieger, C. J. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research* **32**, D438 – D442.

Kristjansdottir, T., Bosma, E. F., Branco dos Santos, F., Özdemir, E., Herrgård, M. J., França, L., Fereira, B., Nielsen, A. T. & Gudmundsson, S. (2019). A metabolic

reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory. *Microbial Cell Factories* **18**, 186.

Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N. & Ullrich, M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of *Theobroma cacao* in non-fermented and fermented cocoa beans. *Food Research International* **111**, 137–147.

Kumari, N., Kofi, K. J., Grimbs, S., D'Souza, R. N., Kuhnert, N., Vrancken, G. & Ullrich, M. S. (2016). Biochemical fate of vicilin storage protein during fermentation and drying of cocoa beans. *Food Research International* **90**, 53–65.

Kunze, M., Pracharoenwattana, I., Smith, S. M. & Hartig, A. (2006). A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochimica et Biophysica Acta - Molecular Cell Research* **1763**, 1441–1452.

Lagunes Gálvez S., Loiseau G., Paredes J.L., Barel M. & Guiraud J-P., (2007). Study on the microflora and biochemistry of cocoa fermentation in the Dominican Republic. *International Journal of Food Microbiology* **114**, 124-130.

Lakshmanan, M., Chung, B. K-S., Liu, C., Kim, S-W. & Lee, D-Y. (2013). Cofactor modification analysis: a computational framework to identify cofactor specificity engineering targets for strain improvement. *Journal of Bioinformatics and Computational Biology* **11**, 1343006.

Latham, J. A., Iavarone, A. T., Barr, I., Juthani, P. V. & Klinman, J. P. (2015). PqqD is a novel peptide chaperone that forms a ternary complex with the radical *S* -adenosylmethionine protein PqqE in the pyrroloquinoline quinone biosynthetic pathway. *Journal of Biological Chemistry* **290**, 12908–12918.

Lee, D., Redfern, O. & Orengo, C. (2007). Predicting protein function from sequence and structure. *Nature Reviews Molecular Cell Biology* **8**, 995–1005.

Lefeber, T., Janssens, M., Camu, N. & De Vuyst, L. (2010). Kinetic analysis of strains of lactic acid bacteria and acetic acid bacteria in cocoa pulp simulation media toward development of a starter culture for cocoa bean fermentation. *Applied and Environmental Microbiology* **76**, 7708-7716.

Lefeber, T., Gobert, W., Vrancken, G., Camu, N. & De Vuyst, L. (2011a). Dynamics and species diversity of communities of lactic acid bacteria and acetic acid bacteria during spontaneous cocoa bean fermentation in vessels. *Food Microbiology* **28**, 457-464.

Lefeber, T., Janssens, M., Moens, F., Gobert, W. & De Vuyst, L. (2011b). Interesting starter culture strains for controlled cocoa bean fermentation revealed by simulated cocoa pulp fermentations of cocoa-specific lactic acid bacteria. *Applied and Environmental Microbiology* **77**, 6694-6698.

Lefeber, T., Papalexandratou, Z., Gobert, W., Camu, N. & De Vuyst, L. (2012). On-farm implementation of a starter culture for improved cocoa bean fermentation and its influence on the flavour of chocolates produced thereof. *Food Microbiology* **30**, 379–392.

Leroy, F. & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science and Technology* **15**, 67–78.

Leroy, F., Verluyten, J. & De Vuyst, L. (2006). Functional meat starter cultures for improved sausage fermentation. *International Journal of Food Microbiology* **106**, 270–285.

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., Weitz, K. K., Eils, R., König, R., Smith, R. D. & Palsson, B. Ø. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology* **6**, 390.

Lima, L. J. R., Almeida, M. H., Nout, M. J. R. & Zwietering, M. H. (2011). *Theobroma cacao* L., "the food of the gods": quality determinants of commercial cocoa beans, with particular reference to the impact of fermentation. *Critical Reviews in Food Science and Nutrition* **51**, 731–761.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42, D490-D495.

Lomsadze, A., Gemayel, K., Tang, S. & Borodovsky, M. (2018). Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Research* **28**, 1079–1089.

Long, C. P. & Antoniewicz, M. R. (2018). How adaptive evolution reshapes metabolism to improve fitness: recent advances and future outlook. *Current Opinion in Chemical Engineering* **22**, 209–215.

Lu, Z.-M., Wang, Z.-M., Zhang, X.-J., Mao, J., Shi, J.-S. & Xu, Z.-H. (2018). Microbial ecology of cereal vinegar fermentation: insights for driving the ecosystem function. *Current Opinion in Biotechnology* **49**, 88-93.

Luttik M., Van Spanning R., Schipper D., Van Dijken J.P. & Pronk J.T., (1997). The low biomass yields of the acetic acid bacterium *Acetobacter pasteurianus* are due to a low stoichiometry of respiration-coupled proton translocation. *Applied and Environmental Microbiology* **63**, 3345-3351.

Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B. & Bruggeman, F. J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal* **8**, 997–1008.

Maarleveld, T. R., Wortel, M. T., Olivier, B. G., Teusink, B. & Bruggeman, F. J. (2015). Interplay between constraints, objectives, and optimality for genome-scale stoichiometric models. *PLoS Computational Biology* **11**, e1004166.

Machado, D., Soons, Z., Patil, K. R., Ferreira, E. C. & Rocha, I. (2012). Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics* **28**, i515–i521.

Machado, D., Herrgård, M. J. & Rocha, I. (2016). Stoichiometric representation of gene-protein-reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. *PLoS Compututational Biology* **12**, e1005140.

Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research* **46**, 7542-7553.

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

Mahadevan, R. & Palsson, B. Ø. (2005). Properties of metabolic networks: structure versus function. *Biophysical Journal* **88**, L07–L09.

Mahadevan, R., Burgard, A. P., Famili, I., Van Dien, S. & Schilling, C. H. (2005). Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals. *Biotechnology and Bioprocess Engineering* **10**, 408–417.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D. M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J. H., Díaz-Muñiz, I., Dosti, B., Smeianov, V., Wechter, W. Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B. & Mills, D. (2006). Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15611–15616.

Makarova, K. S. & Koonin, E. V. (2007). Evolutionary genomics of lactic acid bacteria. *Journal of Bacteriology* **189**, 1199–1208.

Mamlouk, D. & Gullo, M. (2013). Acetic acid bacteria: physiology and carbon sources oxidation. *Indian Journal of Microbiology* **53**, 377–384.

Marco, M. L., Heeney, D., Binda, S., Cifelli, C. J., Cotter, P. D., Foligné, B., Gänzle, M., Kort, R., Pasin, G., Pihlanto, A., Smid, E. J. & Hutkins, R. (2017). Health benefits of fermented foods: microbiota and beyond. *Current Opinion in Biotechnology* **44**, 94–102.

Marcotte, E. M., Pellegrini, M., Ng, H-L., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751–753.

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N. & Kyrpides, N. C. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research* **40**, D115-D122.

Mashima, J., Kodama, Y., Fujisawa, T., Katayama, T., Okuda, Y., Kaminuma, E., Ogasawara, O., Okubo, K., Nakamura, Y. & Takagi, T. (2017). DNA Data Bank of Japan. *Nucleic Acids Research* **45**, D25–D31.

Matsushita, K., Toyama, H. & Adachi, O. (1994). Respiratory chains and bioenergetics of acetic acid bacteria. *Advances in Microbial Physiology* **36**, 247–301.

Matsushita, K., Kobayashi, Y., Mizuguchi, M., Toyama, H., Adachi, O., Sakamoto, K. & Miyoshi, H. (2008). A tightly bound quinone functions in the ubiquinone reaction sites of quinoprotein alcohol dehydrogenase of an acetic acid bacterium, *Gluconobacter suboxydans*. *Bioscience, Biotechnology, and Biochemistry* **72**, 2723–2731.

Matsushita, K., Toyama, H. & Adachi O. (2004). Respiratory chains in acetic acid bacteria: membrane-bound periplasmic sugar and alcohol respirations, p 81-99. *In* Zannoni, Davide (ed), Respiration in archaea and bacteria: diversity of prokaryotic respiratory systems. Springer, Dordrecht, The Netherlands.

Matsushita, K. & Matsutani, M. (2016). Distribution, evolution, and physiology of oxidative fermentation, pp. 159-178. *In* Matsushita, K., Toyama, H., Tonouchi, N., Okamoto-Kainuma, A., (Eds.), Acetic Acid Bacteria Ecology and Physiology. Springer Japan, Tokyo, Japan.

Matsutani, M., Fukushima, K., Kayama, C., Arimitsu, M., Hirakawa, H., Toyama, H., Adachi, O., Yakushi, T. & Matsushita, K. (2014). Replacement of a terminal cytochrome *c* oxidase by ubiquinol oxidase during the evolution of acetic acid bacteria. *Biochimica et Biophysica Acta - Bioenergetics* **1837**, 1810–1820.

Mayorga-Gross, A. L., Quirós-Guerrero, L. M., Fourny, G. & Vaillant, F. (2016). An untargeted metabolomic assessment of cocoa beans during fermentation. *Food Research International* **89**, 901–909.

Maze, A., O'Connell-Motherway, M., Fitzgerald, G. F., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Applied and Environmental Microbiology* **73**, 545-553.

McDonald, A. G., Boyce, S. & Tipton, K. F. (2009). ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Research* **37**, D593-D597.

Meersman, E., Steensels, J., Mathawan, M., Wittocx, P-J., Saels, V., Struyf, N., Bernaert, H., Vrancken, G. & Verstrepen, K. J. (2013). Detailed analysis of the microbial population in Malaysian spontaneous cocoa pulp fermentations reveals a core and variable microbiota. *PLoS ONE* **8**, e81559.

Meersman, E., Steensels, J., Paulus, T., Struyf, N., Saels, V., Mathawan, M., Koffi, J., Vrancken, G. & Verstrepen, K. J. (2015). Breeding strategy to generate robust yeast starter cultures for cocoa pulp fermentations. *Applied and Environmental Microbiology* **81**, 6166–6176.

Meersman, E., Steensels, J., Struyf, N., Paulus, T., Saels, V., Mathawan, M., Allegaert, L., Vrancken, G. & Verstrepen, K. J. (2016). Tuning chocolate flavor through development of thermotolerant *Saccharomyces cerevisiae* starter cultures with increased acetate ester production. *Applied and Environmental Microbiology* **82**, 732–746.

Meersman, E., Struyf, N., Kyomugasho, C., Kermani, Z. J., Santiago, J. S., Baert, E., Hemdane, S., Vrancken, G., Verstrepen, K. J., Courtin, C. M., Hendrickx, M. & Steensels J. (2017). Characterization and degradation of pectic polysaccharides in cocoa pulp. *Journal of Agricultural and Food Chemistry* **65**, 9726-9734.

Megchelenbrink, W., Huynen, M. & Marchiori, E. (2014). optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS ONE* **9**, e86587.

Mende, D. R., Letunic, I., Huerta-Cepas, J., Li, S. S., Forslund, K., Sunagawa, S. & Bork, P. (2017). proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Research* 45, D529-D534.

Mendoza, S. N., Cañón, P. M., Contreras, Á., Ribbeck, M. & Agosín, E. (2017). Genome-scale reconstruction of the metabolic network in *Oenococcus oeni* to assess wine malolactic fermentation. *Frontiers in Microbiology* **8**, 534.

Mendoza, S. N, Olivier, B. G., Molenaar, D. & Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology* **20**, 158.

Metzger, L. E. & Raetz, C. R. H. (2010). An alternative route for UDP-diacylglucosamine hydrolysis in bacterial lipid A biosynthesis. *Biochemistry* **49**, 6715-6726.

Meyer, F. (2003). GenDB-an open source genome annotation system for prokaryote genomes. *Nucleic Acids Research* **31**, 2187–2195.

Miescher Schwenninger, S., Freimüller Leischtfeld, S. & Gantenbein-Demarchi, C. (2016). High-throughput identification of the microbial biodiversity of cocoa bean fermentation by MALDI-TOF MS. *Letters in Applied Microbiology* **63**, 347–355.

Milo, R. & Phillips, R. (2016). Cell Biology by the Numbers. New York: Garland Science, Taylor & Francis Group.

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**, 1993–2004.

Mitchell, P. (2011). Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biochimica et Biophysica Acta - Bioenergetics* **1807**, 1507–1538.

Miura, H., Mogi, T., Ano, Y., Migita, C. T., Matsutani, M., Yakushi, T., Kita, K. & Matsushita, K. (2013). Cyanide-insensitive quinol oxidase (CIO) from *Gluconobacter oxydans* is a unique terminal oxidase subfamily of cytochrome *bd*. *The Journal of Biochemistry* **153**, 535-545.

Moens, F., Lefeber, T. & De Vuyst, L. (2014). Oxidation of metabolites highlights the microbial interactions and role of *Acetobacter pasteurianus* during cocoa bean fermentation. *Applied and Environmental Microbiology* **80**, 1848–1857.

Molenaar, D., van Berlo, R., de Ridder, D. & Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology* **5**, 323.

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M. & Palsson, B. Ø. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nature Biotechnology* **35**, 904–908.

Monod, J. (1949). The growth of bacterial cultures. *Annual Reviews in Microbiology* **3**, 371–394.

Moreno-Zambrano, M., Grimbs, S., Ullrich, M. S. & Hütt, M.-T. (2018). A mathematical model of cocoa bean fermentation. *Royal Society Open Science* **5**, 180964.

Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A. & Pagni, M. (2016). MetaNetX/MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research* **44**, D523–D526.

Morgat, A. et al. (2017). Updates in Rhea – an expert curated resource of biochemical reactions. *Nucleic Acids Research* **45**, D415–D418.

Motamayor, J. C., Lachenaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S. & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE* **3**, e3311.

Mounier, J., Monnet, C., Vallaeys, T., Arditi, R., Sarthou, A-S., Helias, A. & Irlinger, F. (2008). Microbial interactions within a cheese microbial community. *Applied and Environmental Microbiology* **74**, 172–181.

Mullins, E. A., Francois, J. A. & Kappock, T. J. (2008). A specialized citric acid cycle requiring succinyl-coenzyme A (CoA):acetate CoA-transferase (AarC) confers acetic acid resistance on the acidophile *Acetobacter aceti*. *Journal of Bacteriology* **190**, 4933–4940.

Mullins E.A. & Kappock T.J., (2013). Functional analysis of the acetic acid resistance (aar) gene cluster in *Acetobacter aceti* strain 1023. *Acetic Acid Bacteria* **2**, e3.

Núñez, M. F., Kwon, O., Wilson, T. H., Aguilar, J., Baldoma, L. & Lin, E. C. C. (2002). Transport of L-lactate, D-lactate, and glycolate by the LldP and GlcA membrane carriers of *Escherichia coli*. *Biochemical and Biophysical Research Communications* **290**, 824-829.

Nakano, S., Fukaya, M. & Horinouchi, S. (2006). Putative ABC transporter responsible for acetic acid resistance in *Acetobacter aceti*. *Applied and Environmental Microbiology* **72**, 497-505.

Ndoye, B., Cleenwerck, I., Engelbeen, K., Dubois-Dauphin, R., Guiro, A. T., Van Trappen, S., Willems, A. & Thonart, P. (2007). *Acetobacter senegalensis* sp. nov., a thermotolerant acetic acid bacterium isolated in Senegal (sub-Saharan Africa) from mango fruit (*Mangifera indica* L.). *International Journal of Systematic and Evolutionaly Microbiology* **57**, 1576-1581.

Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. (1990). *Physiology of the Bacterial Cell:a Molecular Approach*. Sinauer Associates, Inc., Sunderland, Massachusetts, United States of America.

Nielsen, D. S., Hønholt, S., Tano-Debrah, K. & Jespersen, L. (2005). Yeast populations associated with Ghanaian cocoa fermentations analysed using denaturing gradient gel electrophoresis (DGGE): DGGE analysis of yeasts associated with cocoa fermentation. *Yeast* **22**, 271–284.

Nilsson, A., Nielsen, J. & Palsson, B. Ø. (2017). Metabolic models of protein allocation call for the kinetome. *Cell Systems* **5**, 538–541.

Nothaft, H., Parche, S., Kamionka, A. & Titgemeyer, F. (2003). *In vivo* analysis of HPr reveals a fructose-specific phosphotransferase system that confers high-affinity uptake in *Streptomyces coelicolor*. *Journal of Bacteriology* **185**, 929-937.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44, D733-D745.

Oberhardt, M. A., Zarecki, R., Reshef, L., Xia, F., Duran-Frigola, M., Schreiber, R., Henry, C. S., Ben-Tal, N., Dwyer, D. J., Gophna, U. & Ruppin, E. (2016). Systems-wide prediction of enzyme promiscuity reveals a new underground alternative route for pyridoxal 5'-phosphate production in *E. coli*. *PLoS Computational Biology* **12**, e1004705.

Oh, Y.-K., Palsson, B. Ø., Park, S. M., Schilling, C. H. & Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *Journal of Biological Chemistry* **282**, 28791-28799.

Olivier, B. G., Rohwer, J. M. & Hofmeyr, J.-H. S. (2005). Modelling cellular systems with PySCeS. *Bioinformatics* **21**, 560-561.

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

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**, 535.

Ouattara, H. D., Ouattara, H. G., Droux, M., Reverchon, S., Nasser, W. & Niamke, S. L. (2017). Lactic acid bacteria involved in cocoa beans fermentation from Ivory Coast: species diversity and citrate lyase production. *International Journal of Food Microbiology* **256**, 11–19.

Overbeek, R et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* **42**, D206-D214.

Ozturk, G. & Young, G. M. (2017). Food evolution: the impact of society and science on the fermentation of cocoa Beans. *Comprehensive Reviews in Food Science and Food Safety* **16**, 431–455.

Özcan, E., Selvi, S. S., Nikerel, E., Teusink, B., Toksoy Öner, E. & Çakir, T. (2019). A genome-scale metabolic network of the aroma bacterium *Leuconostoc mesenteroides* subsp. cremoris. *Applied Microbiology and Biotechnology* **103**, 3153-3165.

Palsson, B. Ø. (2006). *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge, United Kingdom.

Panche, A. N., Diwan, A. D. & Chandra, S. R. (2016). Flavonoids: an overview. *Journal of Nutritional Science* **5**, e47.

Paoletti, L., Lu, Y.-J., Schujman, G. E., de Mendoza, D. & Rock, C. O. (2007). Coupling of fatty acid and phospholipid synthesis in *Bacillus subtilis*. *Journal of Bacteriology* **189**, 5816-5824.

Papalexandratou, Z., Falony, G., Romanens, E., Jimenez, J. C., Amores, F., Daniel, H-M. & De Vuyst, L. (2011a). Species diversity, community dynamics, and metabolite kinetics of the microbiota associated with traditional Ecuadorian spontaneous cocoa bean fermentations. *Applied and Environmental Microbiology* **77**, 7698–7714.

Papalexandratou, Z., Vrancken, G., De Bruyne, K., Vandamme, P. & De Vuyst, L. (2011b). Spontaneous organic cocoa bean box fermentations in Brazil are characterized by a restricted species diversity of lactic acid bacteria and acetic acid bacteria. *Food Microbiology* **28**, 1326–1338.

Papalexandratou, Z., Lefeber, T., Bahrim, B., Lee, O. S., Daniel, H-M. & De Vuyst, L. (2013). *Hanseniaspora opuntiae*, *Saccharomyces cerevisiae*, *Lactobacillus fermentum*, and *Acetobacter pasteurianus* predominate during well-performed Malaysian cocoa bean box

fermentations, underlining the importance of these microbial species for a successful cocoa bean fermentation process. *Food Microbiology* **35**, 73–85.

Parsons, J. B. & Rock, C. O. (2013). Bacterial lipids: metabolism and membrane homeostasis. *Progress in Lipid Research* **52**, 249-276.

Pastink, M. I., Teusink, B., Hols, P., Visser, S., de Vos, W. M. & Hugenholtz, J. (2009). Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Applied and Environmental Microbiology* **75**, 3627–3633.

Peleg, M. & Corradini, M. G. (2011). Microbial growth curves: what the models tell us and what they cannot. *Critical Reviews in Food Science and Nutrition* **51**, 917–945.

Pelicaen, R., Gonze, D., Teusink, B., De Vuyst, L. & Weckx, S. (2019). Genome-scale metabolic reconstruction of *Acetobacter pasteurianus* 386B, a candidate functional starter culture for cocoa bean fermentation. *Frontiers in Microbiology* **10**, 2801.

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings of the National Academy of Sciences* **96**, 4285–4288.

Peng, H. L., Wang, P. Y., Wu, C. M., Hwang, D. C. & Chang, H. Y. (1992). Cloning, sequencing and heterologous expression of a *Klebsiella pneumoniae* gene encoding an FAD-independent acetolactate synthase. *Gene* **117**, 125-130.

Pettipher, G. L. (1986). Analysis of cocoa pulp and the formulation of a standardised artificial cocoa pulp medium. *Journal of the Science of Food and Agriculture* **37**, 297–309.

Pfeiffer, T. & Morley, A. (2014). An evolutionary perspective on the Crabtree effect. *Frontiers in Molecular Biosciences* **1**, 17.

Pfeiffer, T., Schuster, S. & Bonhoeffer, S. (2001). Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504–507.

Pitkänen, E., Rousu, J. & Ukkonen, E. (2010). Computational methods for metabolic reconstruction. *Current Opinion in Biotechnology* **21**, 70-77.

Pot, B., Salvetti, E., Mattarelli, P. & Felis, G.E. (2019). The potential impact of the *Lactobacillus* name change: The results of an expert meeting organised by the Lactic Acid Bacteria Industrial Platform (LABIP). *Trends in Food Science and Technology* **94**, 105-113.

Pothakos, V., De Vuyst, L., Zhang, S.J., De Bruyn, F., Verce, M., Torres, J., Callanan, M., Moccand, C. & Weckx, S. (2020). Temporal shotgun metagenomic dissection of an Ecuadorian coffee fermentation process highlights the predominance of lactic acid bacteria. *Current Research in Biotechnology* **2**, 1-15.

Presta, L., Bosi, E., Mansouri, L., Dijkshoorn, L., Fani, R. & Fondi, M. (2017). Constraint-based modeling identifies new putative targets to fight colistin-resistant *A. baumannii* infections. *Scientific Reports* **7**, 3706.

Prigent, S., Nielsen, J. C., Frisvad, J. C. & Nielsen, J. (2018). Reconstruction of 24 *Penicillium* genome-scale metabolic models shows diversity based on their secondary metabolism. *Biotechnology and Bioengineering* **115**, 2604–2612.

Prust, C., Hoffmeister, M., Liesegang, H., Wiezer, A., Fricke, W. F., Ehrenreich, A., Gottschalk, G. & Deppenmeier, U. (2005). Complete genome sequence of the acetic acid bacterium *Gluconobacter oxydans*. *Nature Biotechnology* **23**, 195–200.

Puehringer, S., Metlitzky, M. & Schwarzenbacher, R. (2008). The pyrroloquinoline quinone biosynthesis pathway revisited: a structural approach. *BMC Biochemistry* **9**, 8.

Quek, L-E., Wittmann, C., Nielsen, L. K. & Krömer, J. O. (2009). OpenFLUX: efficient modelling software for ^{13}C-based metabolic flux analysis. *Microbial Cell Factories* **8**, 25.

Raetz, C. R. H. & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Reviews in Biochemistry* **71**, 635-700.

Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. (2007). Lipid A modification systems in Gram-negative bacteria. *Annual Reviews in Biochemistry* **76**, 295-329.

Rainbow, C., & Mitson, G. W. (1953). Nutritional requirements of acetic acid bacteria. *Journal of General Microbiology* **9**, 371–375.

Rau, M. H. & Zeidan, A. A. (2018). Constraint-based modeling in microbial food biotechnology. *Biochemical Society Transactions* **46**, 249–260.

Rauch, B., Pahlke, J., Schweiger, P. & Deppenmeier, U. (2010). Characterization of enzymes involved in the central metabolism of *Gluconobacter oxydans*. *Applied Microbiology and Biotechnology* **88**, 711–718.

Rawel, H., Huschek, G., Sagu, S. & Homann, T. (2019). Cocoa bean proteins—characterization, changes and modifications due to ripening and post-harvest processing. *Nutrients* **11**, 428.

Reddy, M. C. M., Kuppan, G., Shetty, N. D., Owen, J. L., Ioerger, T. R. & Sacchettini, J. C. (2008). Crystal structures of *Mycobacterium tuberculosis* S-adenosyl-L-homocysteine hydrolase in ternary complex with substrate and inhibitors. *Protein Science* **17**, 2134-2144.

Reimers, A-M. & Reimers, A. C. (2016). The steady-state assumption in oscillating and growing systems. *Journal of Theoretical Biology* **406**, 176–186.

Rezac, S., Kok, C. R., Heermann, M. & Hutkins, R. (2018). Fermented foods as a dietary source of live organisms. *Frontiers in Microbiology* **9**, 1785.

Reznik, E., Mehta, P. & Segrè, D. (2013). Flux imbalance analysis and the sensitivity of cellular growth to changes in metabolite pools. *PLoS Computational Biology* **9**, e1003195.

Richards, M. A., Cassen, V., Heavner, B. D., Ajami, N. E., Herrmann, A., Simeonidis, E. & Price, D. (2014). MediaDB: a database of microbial growth conditions in defined media. *PLoS ONE* 9, e103548.

Richardson, E. J. & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics* **14**, 1–12.

Richhardt J., Luchterhand B., Bringer S., Buchs J. & Bott M., (2013). Evidence for a key role of cytochrome bo_3 oxidase in respiratory energy metabolism of *Gluconobacter oxydans*. *Journal of Bacteriology* **195**, 4210-4220.

Röling, W. F. M. & van Bodegom, P. M. (2014). Toward quantitative understanding on microbial community structure and functioning: a modeling-centered approach using degradation of marine oil spills as example. *Frontiers in Microbiology* **5**, 125.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering* **12**, 85–94.

Sá-Pessoa, J., Paiva, S., Ribas, D., Silva, I. J., Viegas, S. C., Arraiano, C. M. & Casal, M. (2013). SATP (YaaH), a succinate-acetate transporter protein in *Escherichia coli*. *Biochemical Journal* **454**, 585-595.

Saeki, A., Taniguchi, M., Matsushita, K., Toyama, H., Theeragool, G., Lotong, N. & Adachi, O. (1997). Microbiological aspects of acetate oxidation by acetic acid bacteria, unfavorable phenomena in vinegar fermentation. *Bioscience, Biotechnology, and Biochemistry* **61**, 317-323.

Saeki, A., Matsushita, K., Takeno, S., Taniguchi, M., Toyama, H., Theeragool, G., Lotong, N. & Adachi, O. (1999). Enzymes responsible for acetate oxidation by acetic acid bacteria. *Bioscience, Biotechnology, and Biochemistry* **63**, 2102-2109.

Saier, M. H., Ye, J. J., Klinke, S. & Nino, E. (1996). Identification of an anaerobically induced phosphoenolpyruvate-dependent fructose-specific phosphotransferase system and evidence for the Embden-Meyerhof glycolytic pathway in the heterofermentative bacterium *Lactobacillus brevis*. *Journal of Bacteriology* **178**, 314-316.

Saier, 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**, D372-D379.

Sakamoto, M. & Komagata, K. (1996). Aerobic growth of and activities of NADH oxidase and NADH peroxidase in lactic acid bacteria. *Journal of Fermentation and Bioengineering* **82**, 210-216.

Sakurai, K., Arai, H., Ishii, M. & Igarashi, Y. (2011). Transcriptome response to different carbon sources in *Acetobacter aceti*. *Microbiology* **157**, 899–910.

Sakurai, K., Arai, H., Ishii, M. & Igarashi, Y. (2012). Changes in the gene expression profile of *Acetobacter aceti* during growth on ethanol. *Journal of Bioscience and Bioengineering* **113**, 343–348.

Sakurai, K., Yamazaki, S., Ishii, M., Igarashi, Y. & Arai, H. (2013). Role of the glyoxylate pathway in acetic acid production by *Acetobacter aceti*. *Journal of Bioscience and Bioengineering* **115**, 32–36.

Salgueiro Machado, S., Luttik, M. A. H., van Dijken, J. P., Jongejan, J. A. & Pronk, J. T. (1995). Regulation of alcohol-oxidizing capacity in chemostat cultures of *Acetobacter pasteurianus*. *Applied Microbiology and Biotechnology* **43**, 1061-1066.

Sandhya, M. V. S., Yallappa, B. S., Varadaraj, M. C., Puranaik, J., Rao, L. J., Janardhan, P. & Murthy, P. S. (2016). Inoculum of the starter consortia and interactive metabolic process in enhancing quality of cocoa bean (*Theobroma cacao*) fermentation. *Food Science and Technology* **65**, 731–738.

Santander Muñoz, M., Rodríguez Cortina, J., Vaillant, F. E. & Escobar Parra, S. (2019). An overview of the physical and biochemical transformation of cocoa seeds to beans and to chocolate: flavor formation. *Critical Reviews in Food Science and Nutrition*, 1–21.

Santos, F., Boele, J. & Teusink, B. (2011). A practical guide to genome-scale metabolic models and their analysis. *Methods in Enzymology* **500**, 509–532.

Sarewicz, M. & Osyczka, A. (2015). Electronic connection between the quinone and cytochrome *c* redox pools and its role in regulation of mitochondrial electron transport and redox signaling. *Physiological Reviews* **95**, 219-243.

Sasaki, Y., Laivenieks, M. & Zeikus, J. G. (2005). *Lactobacillus reuteri* ATCC 53608 *mdh* gene cloning and recombinant mannitol dehydrogenase characterization. *Applied Microbiology and Biotechnology* **68**, 36-41.

Sauer U., Canonaco F., Heri S., Perrenoud A. & Fischer E., (2004). The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *Journal of Biological Chemistry* **279**, 6613-6619.

Saulnier, D. M., Santos, F., Roos, S., Mistretta, T.-A., Spinler, J. K., Molenaar, D., Teusink, B. & Versalovic, J. (2011). Exploring metabolic pathway reconstruction and genome-wide expression profiling in *Lactobacillus reuteri* to define functional probiotic features. *PLoS ONE* **6**, e18783.

Schaechter, M (2015). A brief history of bacterial growth physiology. *Frontiers in Microbiology* **6**, 289.

Schuetz, R., Kuepfer, L. & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology* **3**, 119.

Schuster, S., Pfeiffer, T. & Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology* **252**, 497–504.

Schwan, R. F. (1998). Cocoa fermentations conducted with a defined microbial cocktail inoculum. *Applied and Environmental Microbiology* **64**, 1477–1483.

Schwan, R. F. & Wheals, A. E. (2004). The microbiology of cocoa fermentation and its role in chocolate quality. *Critical Reviews in Food Science and Nutrition* **44**, 205–221.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.

Seif, Y., Kavvas, E., Lachance, J.-C., Yurkovich, J.T., Nuccio, S.-P., Fang, X., Catoiu, E., Raffatellu, M., Palsson, B. Ø. & Monk, J. M. (2018). Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nature Communications* 9, 3771.

Serres, M. H. & Riley, M. (2006). Genomic analysis of carbon source metabolism of *Shewanella oneidensis* MR-1: predictions versus experiments. *Journal of Bacteriology* **188**, 4601-4609.

Shenoy, S. R. & Jayaram, B. (2010). Proteins: sequence to structure and function – current status. *CPPS* **11**, 498–514.

Shinjoh, M. & Toyama, H. (2016). Industrial applications of acetic acid bacteria (vitamin C and others), pp. 321-338. *In* Matsushita, K., Toyama, H., Tonouchi, N., Okamoto-Kainuma, A., (Eds.), Acetic Acid Bacteria Ecology and Physiology. Springer Japan, Tokyo, Japan.

Siegal, M. L. (2015). Shifting sugars and shifting paradigms. *PLoS Biology* **13**, e1002068.

Snoep, J. L., Teixeira de Mattos, M. J., Starrenburg, M. J. & Hugenholtz, J. (1992). Isolation, characterization, and physiological role of the pyruvate dehydrogenase complex and α-acetolactate synthase of *Lactococcus lactis* subsp. *lactis* bv. diacetylactis. *Journal of Bacteriology* **174**, 4838-4841.

Sohlenkamp, C. & Geiger, O. (2016). Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol Reviews* **40**, 133-159.

Sohn, S., Kim, T., Lee, J. H. & Lee, S. (2012). Genome-scale metabolic model of the fission yeast *Schizosaccharomyces pombe* and the reconciliation of *in silico*/*in vivo* mutant growth. *BMC Systems Biology* **6**, 49.

Son, H.-J., Kim, H.-G., Kim, K.-K., Kim, H.-S., Kim, Y.-G., & Lee, S.-J. (2003). Increased production of bacterial cellulose by *Acetobacter* sp. V6 in synthetic media under shaking culture conditions. *Bioresource Technology* **86**, 215–219.

Spaans S.K., Weusthuis R.A., van der Oost, J. & Kengen S.W.M. (2015). NADPH-generating systems in bacteria and archaea. *Frontiers in Microbiology* **6**, 742.

Starrenburg, M. J. & Hugenholtz, J. (1991). Citrate fermentation by *Lactococcus* and *Leuconostoc* spp. *Applied and Environmental Microbiology* **57**, 3535–3540.

Steigmiller S., Turina P. & Graber P., (2008). The thermodynamic H^+/ATP ratios of the H^+-ATP synthases from chloroplasts and *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3745-3750.

Stettner, A. I. & Segrè, D. (2013). The cost of efficiency in energy metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9629–9630.

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**, 92.

Succurro, A., Segrè, D. & Ebenhöh, O. (2019). Emergent subpopulation behavior uncovered with a community dynamic metabolic model of *Escherichia coli* diauxic growth. *mSystems* **4**, e00230-18.

Taffs, R., Aston, J. E., Brileya, K., Jay, Z., Klatt, C. G., McGlynn, S., Mallette, N., Montross, S., Gerlach, R., Inskeep, W. P., Ward, D. M. & Carlson, R. P. (2009). *In silico* approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Systems Biology* **3**, 114.

Tamang, J. P., Watanabe, K. & Holzapfel, W. H. (2016). Review: diversity of microorganisms in global fermented foods and beverages. *Frontiers in Microbiology* **7**, 337.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vadusevan, S., Wolf, Y. I., Yin, J. J. & Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**, 33–36.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M. & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research* **44**, 6614–6624.

Teusink, B., van Enckevort, F. H. J., Francke, C., Wiersma, A., Wegkamp, A., Smid, E. J. & Siezen, R. J. (2005). *In silico* reconstruction of the metabolic pathways of *Lactobacillus plantarum*: comparing predictions of nutrient requirements with those from growth experiments. *Applied and Environmental Microbiology* **71**, 7253–7262.

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**, 40041–40048.

Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R. A. & Smid, E. J. (2009). Understanding the adaptive growth strategy of *Lactobacillus plantarum* by *in silico* optimisation. *PLoS Computational Biology* **5**, e1000410.

Thakur, S. & Guttman, D. S. (2016). A De-Novo Genome Analysis Pipeline (DeNoGAP) for large-scale comparative prokaryotic genomics studies. *BMC Bioinformatics* **17**, 260.

The UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506–D515.

Thiele, I. & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols* **5**, 93–121.

Tomàs-Gamisans, M., Ferrer, P. & Albiol, J. (2018). Fine-tuning the *P. pastoris* iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. *Microbial Biotechnology* **11**, 224–237.

Toribio, A. L., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., et al. (2017). European Nucleotide Archive in 2016. *Nucleic Acids Research* **45**, D32–D36.

Toyama, H., Mathews, F. S., Adachi, O. & Matsushita, K. (2004). Quinohemoprotein alcohol dehydrogenases: structure, function, and physiology. *Archives of Biochemistry and Biophysics* **428**, 10–21.

Trinh, C. T., Wlaschin, A. & Srienc, F. (2009). Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology* **81**, 813–826.

Tsao, R. (2010). Chemistry and biochemistry of dietary polyphenols. *Nutrients* **2**, 1231–1246.

Tsuji, A., Okada, S., Hols, P. & Satoh, E. (2013). Metabolic engineering of *Lactobacillus plantarum* for succinic acid production through activation of the reductive branch of the tricarboxylic acid cycle. *Enzyme and Microbial Technology* **53**, 97–103.

Typas, A., Banzhaf, M., Gross, C. A. & Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature Reviews in Microbiology* **10**, 123-136.

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L. & Vilu, R. (2010). Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Systems Biology* **4**, 166.

Vallenet, D. et al. (2017). MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Research* **45**, D517–D528.

Van Domselaar, G. H., Stothard, P., Shrivastava, S., Cruz, J. A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R. & Wishart, D. S. (2005). BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research* **33**, W455–W459.

van Heeswijk, W. C., Westerhoff, H. V. & Boogerd, F. C. (2013). Nitrogen assimilation in *Escherichia coli*: putting molecular data into a systems perspective. *Microbiology and Molecular Biology Reviews* **77**, 628-695.

Van Impe, J., Poschet, F., Geeraerd, A. & Vereecken, K. (2005). Towards a novel class of predictive microbial growth models. *International Journal of Food Microbiology* **100**, 97–105.

van Niel, E. W. J. & Hahn-Hägerdal, B. (1999). Nutrient requirements of lactococci in defined growth media. *Applied Microbiology and Biotechnology* **52**, 617–627.

Varma, A., Boesch, B. W. & Palsson, B. Ø. (1993). Biochemical production capabilities of *Escherichia coli*. *Biotechnology and Bioengineering* **42**, 59-73.

Veith, N., Solheim, M., van Grinsven, K. W. A., Olivier, B. G., Levering, J., Grosseholz, R., Hugenholtz, J., Holo, H., Nes, I., Teusink, B. & Kummer, U. (2015). Using a genome-scale metabolic model of *Enterococcus faecalis* V583 to assess amino acid uptake and its impact on central metabolism. *Applied and Environmental Microbiology* **81**, 1622–1633.

Verce, M., De Vuyst, L. & Weckx, S. (2019). Shotgun metagenomics of a water kefir fermentation ecosystem reveals a novel *Oenococcus* species. *Frontiers in Microbiology* **10**, 479.

Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**, 501-517.

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**, e110785.

Visintin, S., Alessandria, V., Valente, A., Dolci, P. & Cocolin, L. (2016). Molecular identification and physiological characterization of yeasts, lactic acid bacteria and acetic acid bacteria isolated from heap and box cocoa bean fermentations in West Africa. *International Journal of Food Microbiology* **216**, 69-78.

Voigt, J. & Biehl, B. (1995). Precursors of the cocoa-specific aroma components are derived from the vicilin-class (7S) globulin of the cocoa seeds by proteolytic processing. *Botanica Acta* **108**, 283–289.

Voigt, J., Biehl, B. & Wazir, S. K. S. (1993). The major seed proteins of *Theobroma cacao* L. *Food Chemistry* **47**, 145–151.

Vollmer, W., Blanot, D. & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews* **32**, 149-167.

Wang, L., Nägele, T., Doerfler, H., Fragner, L., Chaturvedi, P., Nukarinen, E., Bellaire, A., Huber, W., Weiszman, J., Engelmeier, D., Ramsak, Z., Gruden, K. & Weckwerth, W. (2016). System level analysis of cacao seed ripening reveals a sequential interplay of primary

and secondary metabolism leading to polyphenol accumulation and preparation of stress resistance. *The Plant Journal* **87**, 318–332.

Wang, N.-C. & Lee, C.-Y. (2007). Enhanced transaminase activity of a bifunctional L-aspartate 4-decarboxylase. *Biochemical and Biophysical Research Communications* **356**, 368-373.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Research* **45**, D535-D542.

Wegkamp, A., Teusink, B., de Vos, W. M. & Smid, E. J. (2010). Development of a minimal growth medium for *Lactobacillus plantarum*. *Letters in Applied Microbiology* **50**, 57–64.

Westerhoff, H. V. & Palsson, B. Ø. (2004). The evolution of molecular biology into systems biology. *Nature Biotechnology* **22**, 1249–1252.

Whitfield, C., & Trent, M. S. (2014). Biosynthesis and export of bacterial lipopolysaccharides. *Annual Review of Biochemistry* **83**, 99–128.

Wisselink, H. W., Weusthuis, R. A., Eggink, G., Hugenholtz, J. & Grobben, G. J. (2002). Mannitol production by lactic acid bacteria: a review. *International Dairy Journal* **12**, 151–161.

Wolfe, B. E., Button, J. E., Santarelli, M. & Dutton, R. J. (2014). Cheese rind communities provide tractable systems for *in situ* and *in vitro* studies of microbial diversity. *Cell* **158**, 422–433.

Wolfe, B. E. & Dutton, R. J. (2015). Fermented foods as experimentally tractable microbial ecosystems. *Cell* **161**, 49–55.

Wu, Y-C., Rasmussen, M. D., Bansal, M. S. & Kellis, M. (2014a). Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Research* **24**, 475–486.

Wu, X., Wang, X. & Lu, W. (2014b). Genome-scale reconstruction of a metabolic network for *Gluconobacter oxydans* 621H. *Biosystems* **117**, 10–14.

Wuyts, S., Wittouck, S., De Boeck, I., Allonsius, C. N., Pasolli, E., Segata, N. & Lebeer, S. (2017). Large-scale phylogenomics of the *Lactobacillus casei* group highlights taxonomic inconsistencies and reveals novel clade-associated features. *mSystems* **2**, e00061-17.

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**, 205–216.

Xu, N., Liu, J., Ai, L. & Liu, L. (2015). Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus casei* LC2W. *Gene* **554**, 140–147.

Xu, Y., Holic, R. & Hua, Q. (2020). Comparison and analysis of published genome-scale metabolic models of *Yarrowia lipolytica*. *Biotechnology and Bioprocess Engineering* **25**, 53-61.

Yakushi, T. & Matsushita, K. (2010). Alcohol dehydrogenase of acetic acid bacteria: structure, mode of action, and applications in biotechnology. *Applied Microbiology and Biotechnology* **86**, 1257–1265.

Yamada, Y., Nunoda, M., Ishikawa, T. & Tahara, Y. (1981). The cellular fatty acid composition in acetic acid bacteria. *The Journal of General and Applied Microbiology* **27**, 405-417.

Yamada, Y. (2016). Systematics of acetic acid bacteria, pp. 1-50. *In* Matsushita, K., Toyama, H., Tonouchi, N., Okamoto-Kainuma, A., (Eds.), Acetic Acid Bacteria Ecology and Physiology. Springer Japan, Tokyo, Japan.

Yang, J. E., Park, S. J., Kim, W. J., Kim, H. J., Kim, B. J., Lee, H., Shin, J. & Lee, S.Y. (2018). One-step fermentative production of aromatic polyesters from glucose by metabolically engineered *Escherichia coli* strains. *Nature Communications* **9**, 79.

Yang, J. & Pittard, J. (2008). Biosynthesis of the aromatic amino acids, in EcoSal Plus Domain 3, ed. V. Stewart.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. & Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40**, W445–W451.

Yoshimura, M., Oshima, T. & Ogasawara, N. (2007). Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*. *BMC Microbiology* **7**, 69.

Yu, C.-S., Lin, C.-J. & Hwang, J.-K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science*. **13**, 1402-1406.

Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Solorzano, R. L., Bouchez, O., Vignes, H., Severts, P., Hurtado, J., Yepez, A., Grivetti, L., Blake, M. & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology and Evolution* **2**, 1879–1888.

Zaunmüller, T., Eichert, M., Richter, H. & Unden, G. (2006). Variations in the energy metabolism of biotechnologically relevant heterofermentative lactic acid bacteria during growth on sugars and organic acids. *Applied Microbiology and Biotechnology* **72**, 421-429.

Zhang, H., Ye, C., Xu, N., Chen, C., Chen, X., Yuan, F., Xu, Y., Yang, J. & Sun, D. (2017). Reconstruction of a genome-scale metabolic network of *Komagataeibacter nataicola* RZS01 for cellulose production. *Scientific Reports* **7**, 7911.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y. & Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46**, W95-W101.

Zhao, Q., Stettner, A. I., Reznik, E., Paschalidis, I. Ch. & Segrè, D. (2016). Mapping the landscape of metabolic goals of a cell. *Genome Biology* **17**, 109.

Zheng, J., Ruan, L., Sun, M. & Gänzle, M. (2015). A genomic view of lactobacilli and pediococci demonstrates that phylogeny matches ecology and physiology. *Applied and Environmental Microbiology* **81**, 7233-7243.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C. M. A. P., Harris, H. M. B., Mattarelli, P., O'Toole, P. W., Pot, B., Vandamme, P., Walter, J., Watanabe, K., Wuyts, S., Felis, G. E., Gänzle, M. G. & Lebeer, S. (2020). A taxonomic note on the genus *Lactobacillus*: description

of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *International Journal of Systematic and Evolutionary Microbiology*.

Zhu, X., Gerstein, M. & Snyder, M. (2007). Getting connected: analysis and principles of biological networks. *Genes & Development* **21**, 1010–1024.

Zotta, T., Parente, E. & Ricciardi, A. (2017). Aerobic metabolism in the genus *Lactobacillus*: impact on stress response and potential applications in the food industry. *Journal of Applied Microbiology* **122**, 857-869.

Summary

Cocoa bean fermentation is an essential fermentation process to obtain the necessary raw material for the production of cocoa-derived products, among which chocolate. Despite its huge economic importance, cocoa bean fermentation is still a spontaneous fermentation process, relying completely on the variability of the environmental conditions and microbiota. To gain a better control over the latter, the use of functional starter cultures has been proposed. These starter cultures are selected based on specific functional properties that are thought to be beneficial for the fermentation process. Successful cocoa bean fermentation processes are typically dominated by three microbial groups, namely yeasts, lactic acid bacteria, and acetic acid bacteria. Previously, a number of candidate functional starter cultures have been proposed for the lactic acid bacteria, namely *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80, and for the acetic acid bacteria, namely *Acetobacter pasteurianus* 386B, *Acetobacter ghanensis* LMG 23848^{T}, and *Acetobacter senegalensis* 108B. These strains were selected because they combine high substrate consumption rates and relatively high cell densities, allowing them to dominate the fermentation process. However, these observations were merely true under standardized conditions and lack of knowledge regarding the strains' physiology currently hinders their optimal usage in controlled cocoa bean fermentation processes. The metabolism of bacteria determines an important part of their physiology, and this is recently being investigated by using computational models. The development of such models of the candidate functional starter cultures for cocoa bean fermentation processes and the computational analysis derived thereof was the aim of this PhD thesis.

The computational models developed in the present work were genome-scale metabolic models, which constitute a comprehensive repertoire of metabolic enzymes with their concomitant reactions, and this at genome-scale. Thus, the reconstruction of the genome-scale metabolic network of a bacterial strain is fundamental. It requires a combination of high-quality genome re-annotation, comparative genomics, manual curation, and experimental validation. Genome-scale metabolic modeling together with the use of previously published experimental data under cocoa fermentation conditions allowed to contextualize the experimental data and to gain new insights into the metabolic properties of the candidate functional starter cultures. For *A. pasteurianus* 386B, it was shown that this strain could grow *in vitro* in a defined medium containing lactic acid as the sole carbon source. Simulations with the *A. pasteurianus* 386B genome-scale metabolic model allowed to decipher the metabolic roles of lactate and ethanol under cocoa fermentation conditions and the energetic properties of the strains' aerobic respiratory chain, and it shed light on the possible functional role of a NAD(P)$^{+}$ transhydrogenase. Modeling the metabolite dynamics of *A. ghanensis* LMG 23848^{T} under cocoa fermentation conditions and its relation to growth revealed an alternative strategy for its diauxic growth, compared with *A. pasteurianus* 386B under the same conditions, which was related to a difference in lactate consumption rate and pyruvate overflow. Additionally, a hypothesis was formulated for the functional role of the cytochrome bc_1 complex in *A. ghanensis* LMG 23848^{T}. For *A. senegalensis* 108B, it was shown that, next to lactic acid, also citric acid could sustain its growth *in vitro* as the sole carbon source. Furthermore, the absence

of the glyoxylate cycle predicted from its genome did not correspond with its species description that reports growth on ethanol as the sole carbon source. Comparative genomics of *Acetobacter* species revealed that other species also showed this discrepancy, pointing to the need to carefully check the reported data and compare them with genomic data. For *L. fermentum* 222 and *L. plantarum* 80, core genome-scale metabolic models allowed to gain insight into the possible metabolic flux distributions as a function of environmental conditions, most notably substrate usage. The modeling also indicated a current lack in knowledge; for example, concerning the presence and consumption of undefined substrates in the complex medium used, which could influence the resulting population and metabolite dynamics.

In summary, genome-scale metabolic modeling of candidate functional starter cultures for cocoa bean fermentation processes provided useful *in silico* tools to gain insight into their metabolic properties at a systemic level. As more experimental data will become available, these models will greatly help to integrate and interpret multi-dimensional data sets. The research performed highlighted the strengths of genome-scale metabolic modeling, allowing system-level analysis of metabolism, while at the same time revealing its limitations, since multiple constraints were needed to obtain biologically relevant solutions with the computational methods at hand. However, it is clear that the selection and further development of functional starter cultures for improved cocoa bean fermentation processes will benefit from *in silico* experimentation, next to *in vitro* and *in vivo* experiments. Ultimately, bioprocess modeling of cocoa bean fermentation processes, using genome-scale metabolic models of microbial strains and if possible a genome-scale metabolic model of the cocoa bean itself, will lead to predictive models for cocoa bean fermentation processes, providing the necessary *in silico* tools to further optimize the use of process-technological parameters during fermentation. Thus, the work performed in this PhD research has laid the basic fundaments toward this line of research and development to improve cocoa bean fermentation processes.

Samenvatting

Het cacaoboonfermentatieproces is een essentieel proces dat nodig is om de noodzakelijke grondstof, met name de gefermenteerde cacaobonen, voor de productie van cacao- afgeleide producten, waaronder chocolade, te bekomen. Alhoewel dit proces veel economisch belang heeft, is het cacaoboonfermentatieproces tot op heden een spontaan fermentatieproces. Het is hierbij volledig afhankelijk van de variërende omgevingscondities en –microbiota. Om meer controle te verkrijgen over deze laatste, is het gebruik van functionele starterculturen aangewezen. Deze starterculturen worden geselecteerd op basis van specifieke functionele karakteristieken die gunstig kunnen zijn voor het cacaoboonfermentatieproces. Succesvolle cacaoboonfermentatieprocessen worden typisch gedomineerd door drie microbiële groepen, met name gisten, melkzuurbacteriën en azijnzuurbacteriën. In vorig werk werd reeds een reeks kandidaat-functionele starterculturen voorgesteld. Voor de melkzuurbacteriën zijn dit *Lactobacillus fermentum* 222 en *Lactobacillus plantarum* 80 en voor de azijnzuurbacteriën zijn dit *Acetobacter pasteurianus* 386B, *Acetobacter ghanensis* LMG 23848^T en *Acetobacter senegalensis* 108B. Deze stammen werden geselecteerd omdat zij hoge substraatconsumptiesnelheden combineren met relatief hoge celdensiteiten, dewelke hen toelaten het fermentatieproces te domineren. Belangrijk hierbij is dat deze bevindingen enkel waar zijn onder gestandaardiseerde condities. Beperkte kennis omtrent de fysiologie van deze stammen verhindert op dit moment hun optimaal gebruik in het controleren van cacaoboonfermentatieprocessen. Het metabolisme van bacteriën bepaalt in grote mate hun fysiologie, en dit wordt recent onderzocht door middel van computationele modellen. Het ontwikkelen en analyseren van zulke modellen voor de voorgestelde kandidaat-functionele starterculturen vormde het onderwerp van deze doctoraatsthesis.

De computationele modellen waarvan sprake waren genoomwijde metabolische modellen (GEMs), dewelke het repertoire aan metabole enzymen en de biochemische reacties die zij katalyseren in de bacteriële cellen omvatten. De reconstructie van het metabole netwerk op genoomschaal is dus fundamenteel en vroeg een gecombineerde aanpak van hoge-kwaliteit genoomherannotatie, comparatieve genomica en experimentele validatie. De GEMs werden gebruikt om reeds gepubliceerde experimentele data onder cacaofermentatiecondities te contextualiseren en nieuwe inzichten te verkrijgen in de metabolische karakteristieken van de kandidaat-functionele starterculturen. Voor *A. pasteurianus* 386B werd aangetoond dat deze stam *in vitro* kon groeien in een gedefinieerd medium met melkzuur als enige koolstofbron. Door middel van simulaties met het *A. pasteurianus* 386B GEM konden de metabolische rol van zowel melkzuur als ethanol onder cacaofermentatiecondities en de energetische karakteristieken van de aerobe respiratieketen van deze stam aangetoond worden, alsook de mogelijke metabole functie van een NAD(P)$^+$ transhydrogenase. Het modelleren van de microbiële dynamica van *A. ghanensis* LMG 23848^T onder cacaofermentatiecondities en de relatie ervan met de groei van deze bacteriële cultuur wees op een alternatieve strategie voor de tweevoudige groei van deze stam, ten opzichte van de tweevoudige groei van *A. pasteurianus* 386B onder dezelfde condities, en dit omwille van een verschil in melkzuurconsumptiesnelheid en pyruvaatsecretie. Er werd ook een hypothese geformuleerd

omtrent de functionele rol van het cytochroom bc_1-complex in *A. ghanensis* LMG 23848^T. Voor *A. senegalensis* 108B werd aangetoond dat deze stam, naast melkzuur, ook op citroenzuur als enige koolstofbron kon groeien. De afwezigheid van de glyoxylaatcyclus, voorspeld op basis van het in het genoom, bij *A. senegalensis* 108B is in tegenstelling tot de soortbeschrijving, dewelke stipuleert dat deze azijnzuurbacteriesoort in staat is tot groei op ethanol als enige koolstofbron. Comparatieve genomica van verschillende soorten van het *Acetobacter*-genus toonde aan dat ook andere soorten binnen dit genus deze discrepantie vertonen. Deze vaststelling vraagt om nader onderzoek van de gerapporteerde data en een vergelijking met genoomdata. Voor *L. fermentum* 222 en *L. plantarum* 80 leidde de ontwikkeling van GEMs tot nieuwe inzichten in de mogelijke metabole fluxverdelingen, voornamelijk ten aanzien van substraatverbruik. Het modelleren van de microbiële dynamica wees ook op een tekortkoming aan huidige kennis over deze stammen, bijvoorbeeld met betrekking tot het gebruik van ongedefinieerde substraten in een rijk groeimedium en de invloed hiervan op de resulterende microbiële en metabolietdynamica.

Samenvattend werden door middel van de ontwikkelde GEMs van de kandidaat-functionele starterculturen voor cacaoboonfermentatieprocessen nieuwe inzichten verkregen in hun metabolisme en dit op systeemniveau. Wanneer nieuwe experimentele data voorhanden zullen zijn, zullen deze computationele modellen een grote hulp zijn voor de integratie en interpretatie van multidimensionele data. Het verrichte onderzoek heeft de sterkten van metabolisch modelleren op genoomschaal aangetoond, vermits een metabolische analyse op systeemniveau nu mogelijk wordt, maar ook de zwakten van dit type modellen, vermits vele randvoorwaarden opgelegd dienden te worden aan de modellen om biologisch relevante oplossingen te verkrijgen. Dit doctoraatsonderzoek heeft evenwel duidelijk gemaakt dat verbeterde selectie en ontwikkeling van functionele starterculturen voor cacaoboonfermentatieprocessen een computationeel luik nodig heeft, naast de klassieke veld- en laboratoriumexperimenten. Uiteindelijk zal modelleren van cacaoboonfermentatieprocessen mogelijk worden, en dit met behulp van GEMs van de microbiota en indien mogelijk van de cacaobonen zelf. Deze modellen kunnen dan predicties doen dewelke een optimalisatie van het gebruik van procestechnologische parameters voor het fermentatieproces mogelijk maken. Dit doctoraatsonderzoek heeft hiermee de fundamenten gelegd om deze lijn van onderzoek en ontwikkeling voor verbetering van cacaoboonfermentatieprocessen mogelijk te maken.

Résumé

La fermentation du cacao est un processus essentiel pour obtenir la matière première nécessaire pour la production de produits dérivés du cacao, comme par exemple le chocolat. Malgré l'impact économique important, la fermentation du cacao est encore réalisée de manière spontanée. Ceci entraîne une variabilité importante qui dépend des conditions environnementales et du microbiote présent naturellement. Pour obtenir un meilleur contrôle sur ce dernier, l'utilisation de cultures "starter" fonctionnelles a été proposée. Ces cultures de départ sont sélectionnées sur base de caractéristiques fonctionnelles spécifiques, importantes pour le processus de fermentation. Une fermentation de cacao favorable est caractérisée par la domination de trois groupes de microorganismes : les levures, les bactéries lactiques, et les bactéries acétiques. En ce qui concerne les bactéries, de nombreuses cultures starter ont été proposées, à savoir *Lactobacillus fermentum* 222 et *Lactobacillus plantarum* 80 pour les bactéries lactiques et *Acetobacter pasteurianus* 386B, *Acetobacter ghanensis* LMG 23848^T, et *Acetobacter senegalensis* 108B pour les bactéries acétiques. Ces souches de bactéries ont été sélectionnée car elles combinent un taux de consommation de substrat et des densités de population cellulaires relativement élevées. Cependant, ces observations ont été faites uniquement dans des conditions de laboratoire standardisées et un manque de connaissance en ce qui concerne la physiologie des souches limite leur utilisation optimale pour mieux contrôler la fermentation du cacao. Le métabolisme des bactéries constitue une partie importante de leur physiologie et la recherche actuelle se concentre de plus en plus sur la modélisation du métabolisme et la simulation des flux métaboliques par ordinateur. Le développement de modèles computationnels du métabolisme des cultures starter proposés pour la fermentation du cacao a constitué le but de cette thèse de doctorat.

Les modèles computationnels qui ont été développés dans cette thèse sont des modèles métaboliques à l'échelle du génome. Ceux-ci représentent le répertoire enzymatique dont les bactéries disposent et qui est encodé dans le génome bactérien. Par conséquence, la reconstruction du réseau métabolique est fondamentale, et a entraîné la ré-annotation du génome, une étude de génomique comparative, la curation manuelle des annotations et la validation du modèle par des expériences *in vitro*. La modélisation nous a permis de contextualiser des données expérimentales déjà publiées pour en obtenir de nouvelles informations concernant les propriétés métaboliques des cultures starter. Pour *A. pasteurianus* 386B, les expériences ont démontré que cette souche peut croître dans un milieu défini qui contient de l'acide lactique comme seule source de carbone. Des simulations utilisant le modèle métabolique de *A. pasteurianus* 386B ont clarifié les rôles métaboliques de l'acide lactique et de l'éthanol, les propriétés énergétiques de sa chaîne respiratoire, et ont permis d'assigner un rôle possible à une NAD(P)$^+$ transhydrogénase. La modélisation de la dynamique des métabolites provenant d'un milieu de croissance de *A. ghanensis* LMG 23848^T dans des conditions simulant la fermentation du cacao, a mis en évidence une stratégie alternative de croissance biphasique comparé à *A. pasteurianus* 386B. Ceci est dû à une différence dans le taux de consommation de l'acide lactique et à l'éventuelle production de pyruvate. De plus, une hypothèse a été formulée pour décrire le rôle du cytochrome bc_1 chez *A. ghanensis* LMG

23848^{T}. Pour *A. senegalensis* 108B, les expériences ont démontré, tant pour l'acide lactique que pour l'acide citrique, que ces sources de carbone permettaient, à elles seules, la croissance de cette bactérie. L'absence du cycle du glyoxylate chez *A. senegalensis* 108B ne correspondait pas à la description de cette espèce, laquelle pouvant croître sur l'éthanol comme seule source de carbone. La génomique comparative d'espèces du genus *Acetobacter* a révélé que d'autres espèces aussi montrent cette divergence, ce qui démontre la nécessité de ré-examiner les données publiées. Pour *L. fermentum* 222 et *L. plantarum* 80, la modélisation de leur métabolisme du carbone a permis d'explorer les distributions de flux métaboliques en fonction des substrats consommés. Les simulations ont aussi révélé le manque de connaissance que nous avons sur ces bactéries lactiques, puisque la consommation de substrats non identifiés venant du milieu de croissance pourrait influencer leur dynamique de croissance.

En résumé, la modélisation métabolique à l'échelle du génome des cultures starter proposées pour la fermentation du cacao a permis le développement d'outils *in silico* qui peuvent être utilisés pour mieux comprendre le métabolisme de ces souches, et donc leur physiologie. En attente de nouvelles données expérimentales, ces modèles seront très utiles pour intégrer et interpréter des données multidimensionnelles, comme ceux provenant de la transcriptomique et la métabolomique.

Cette recherche a motivé le développement et l'utilisation de modèles métaboliques permettant d'analyser le métabolisme au niveau systémique. D'autre part, la nécessité d'ajouter des contraintes multiples pour obtenir des simulations probantes dévoile ces limitations. Toutefois, la sélection et l'amélioration des cultures starter pour la fermentation du cacao peuvent bénéficier d'analyse par ordinateur, en parallèle avec des expériences *in vitro* et *in vivo*. Finalement, la modélisation du bioprocédé de la fermentation du cacao, en utilisation des modèles métaboliques à l'échelle du génome des microorganismes et éventuellement de la fève de cacao, pourrait conduire à des modèles prédictifs concernant la fermentation du cacao. Ces outils peuvent ensuite être utilisés pour optimiser l'utilisation des paramètres de régulation de la fermentation. Les résultats acquis dans cette thèse de doctorat apportent les bases nécessaires pour de futurs projets de modélisation dans la fermentation du cacao avec comme but d'améliorer la fermentation du cacao.

Curriculum vitae

Rudy Pelicaen was born on April 2, 1990, in Etterbeek, Belgium. He graduated from the Koninklijk Atheneum Grimbergen in 2008. He obtained a Master of Science in Bioengineering Sciences, specialization Molecular Biotechnology, at the Vrije Universiteit Brussel (VUB), Belgium, in 2013. In 2014, he obtained a Master of Science in Bioinformatics and Modeling at the Université libre de Bruxelles (ULB). That same year, he started a joint PhD, both at the Research Group of Industrial Microbiology and Food Biotechnology of the VUB, under the supervision of Prof. Dr. Stefan Weckx, and at the Unit of Theoretical Chronobiology of the ULB, under the supervision of Prof. Dr. Didier Gonze. Initially financed by the Research Council of the VUB in the framework of the Interuniversity Institute of Bioinformatics in Brussels, he obtained a PhD Fellowship strategic basic research of the Research Foundation Flanders (FWO-Vlaanderen), starting January 2016. From November 2017 until March 2018, he stayed at the Systems Bioinformatics group headed by Prof. Dr. Bas Teusink at the Vrije Universiteit Amsterdam, The Netherlands, for which he obtained a grant for a long stay abroad from FWO-Vlaanderen. His research dealt with the development of genome-scale metabolic models of candidate functional starter cultures for cocoa bean fermentation processes. Rudy Pelicaen is co-author of two scientific papers published in peer-reviewed international journals. He attended seven international scientific conferences where he presented his research results four times as posters and four times as oral presentation.

Publications in journals with peer review related to the PhD research

Pelicaen, R., Gonze, D., Teusink, B., De Vuyst, L. & Weckx, S. (2019). Genome-scale metabolic reconstruction of *Acetobacter pasteurianus* 386B, a candidate functional starter culture for cocoa bean fermentation. *Frontiers in Microbiology* **10**, 2801.

Other publications in journals with peer review

Illeghems, K., **Pelicaen, R.**, De Vuyst, L. & Weckx, S. (2016). Assessment of the contribution of cocoa-derived strains of *Acetobacter ghanensis* and *Acetobacter senegalensis* to the cocoa bean fermentation process through a genomic approach. *Food Microbiology* **58**, 68–78.

Oral presentations (*presenting author)

Pelicaen*, R., Gonze, D., De Vuyst, L. & Weckx, S. (2017). The reconstruction of a genome-scale metabolic model (GSMM) of the carbon core metabolism of *Acetobacter pasteurianus* 386B. International Society for Computational Biology Regional Student Group Belgium meeting. October 13, VUB, Belgium.

Pelicaen*, R., Gonze, D., De Vuyst, L. & Weckx, S (2018). Genome-scale metabolic modeling of *Acetobacter pasteurianus* 386B. 5th International Conference on Acetic Acid Bacteria, September 4 - 7, Weihenstephan, Germany.

Pelicaen*, R., Gonze, D., De Vuyst, L. & Weckx, S (2019). System-level metabolic properties of *Acetobacter pasteurianus* 386B revealed by genome-scale metabolic modeling. International Society for Computational Biology Regional Student Group Belgium meeting, May 3, Leuven, Belgium.

Faust, K. & **Pelicaen*, R.** (2019). Introductory lecture to metabolic community models. Summer School on Microbial Community Modeling, September 17, Leuven, Belgium.

Poster presentations (*presenting author)

Pelicaen*, R., Illeghems, K., De Vuyst, L. & Weckx, S (2015). Assessment of the contribution of cocoa-derived strains of *Acetobacter ghanensis* and *Acetobacter senegalensis* to the cocoa bean fermentation process through a genomic approach. 10th Benelux Bioinformatics Conference (BBC 2015). December 7-8, Antwerp, Belgium.

Pelicaen*, R., Illeghems, K., De Vuyst, L. & Weckx, S (2016). Comparative genomic analysis reveals adaptations of *Acetobacter ghanensis* and *Acetobacter senegalensis* to the cocoa bean fermentation process. 15th European Conference on Computational Biology, September 3 - 7, The Hague, The Netherlands.

Pelicaen*, R., Gonze, D., De Vuyst, L. & Weckx, S. (2018). Genome-scale metabolic modeling of *Acetobacter pasteurianus* 386B. Solvay workshop dynamics of biological systems, May 2 - 4, ULB La Plaine, Belgium.

Pelicaen*, R., Gonze, D., De Vuyst, L. & Weckx, S. (2019). Genome-scale metabolic modeling of *Acetobacter pasteurianus* 386B reveals system-level properties of its metabolism. VIB Conference 'Emerging applications of microbes', June 3 - 4, Leuven, Belgium.

Annexes

Annex Chapter 5

Table S1.

Data	Experimental data of Adler *et al.*, 2014				FBA data (mmol/g_{CDW}/h)		
Fermentation parameters	Stoichiometric yield (mmol/g_{CDW})	SD	C number	Fraction	C-mmol	mmol	SD
Lactate	41.3	1.5	3	0.28	22.11	7.370	0.268
Ethanol	106.0	2.2	2	0.72	37.88	18.940	0.393
Acetate	94.0	3.0	2	0.64	33.58	16.790	0.536
Acetoin	14.3	1.3	4	0.10	10.22	2.554	0.232
Total carbon consumption	147.3					26.310	
Total carbon consumption (C-mmol/g_{CDW}/h)	60.00				59.99		

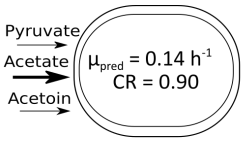
Table S2.

Data	Experimental data of Adler *et al.*, 2014			FBA data (mmol/g_{CDW}/h)	
Fermentation parameters	Peak concentration (mM)	C number	Fraction	C-mmol	mmol
Acetate	150	2	0.83	44.00	22.00
Acetoin	20	4	0.11	12.00	3.00
Pyruvate	10	3	0.06	4.50	1.50
Total carbon consumption	180				26.50
Total carbon consumption (C-mmol/g_{CDW}/h)	60.00			60.50	

Table S3.

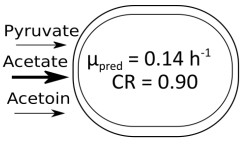
Mode of metabolism	Reactions blocked
Ethanol - acetic acid (intracellular)	Periplasmic ethanol oxidation, citrate synthase
Ethanol - acetic acid (periplasmatic)	Intracellular ethanol oxidation
Ethanol – carbon dioxide	Periplasmic ethanol oxidation
Acetic acid – carbon dioxide	None
Lactic acid - acetic acid	Citrate synthase, acetate kinase
Lactic acid – carbon dioxide	None
Lactic acid - acetoin	Acetoin dehydrogenase, citrate synthase
Acetoin – carbon dioxide	None

Figure S1.



iAp386B454

IN	pFBA	FVA max	FVA min
Pyruvate*	-1.50	-1.49	-1.50
Acetate*	22.00	22.00	22.00
Acetoin*	3.00	3.00	3.00
Oxygen	56.19	56.26	56.19
H^+	22.29	22.30	22.29
Phosphate	0.10	0.10	0.10
Ammonium	1.32	1.32	1.30
Sulfate	0.03	0.03	0.03
OUT			
CO_2	54.45	54.51	54.45
H_2O	56.24	56.27	56.24



iAp386Bcore

IN	pFBA	FVA max	FVA min
Pyruvate*	-1.50	-1.49	-1.50
Acetate*	22.00	22.00	22.00
Acetoin*	3.00	3.00	3.00
Oxygen	57.66	57.71	57.66
H^+	21.58	21.60	21.58
Phosphate	0.57	0.57	0.57
Ammonium	NA	NA	NA
Sulfate	NA	NA	NA
OUT			
CO_2	54.45	54.51	54.45
H_2O	55.49	55.53	55.49

Annex Chapter 6

Supplementary File 1

Monoculture ODE model: *Acetobacter pasteurianus*

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Etoh	Ethanol concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM
Aco	Acetoin concentration	mM
Pyr	Pyruvate concentration	mM

Parameters

symbol	name	unit
μ_1	Maximum specific growth rate phase 1	h^-1
μ_2	Maximum specific growth rate phase 2	h^-1
$Y_{Etoh/X}$	Ethanol-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-to-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-to-biomass yield	mmol/mg
$Y_{Aco/X}$	Acetoin-to-biomass yield	mmol/mg
$Y_{Pyr/X}$	Pyruvate-to-biomass yield	mmol/mg
$Y_{Ac/Etoh}$	Acetate-from-ethanol yield	mmol/mmol
$Y_{Aco/La}$	Acetoin-from-lactate yield	mmol/mmol
$Y_{Pyr/La}$	Pyruvate-from-lactate yield	mmol/mmol
K_{Etoh}	Monod constant of ethanol	mM
K_{La}	Monod constant of lactate	mM
K_{Ac}	Monod constant of acetate	mM
K_{Aco}	Monod constant of acetoin	mM
K_{Pyr}	Monod constant of pyruvate	mM
K_{inh}	Monod inhibition constant of ethanol	mM
w	Basal growth weight	

Modified Monod model

Growth equations

$$f_1 = \mu_1 * \frac{La}{La + K_{La}} * (w + (1 - w) * \frac{Etoh}{Etoh + K_{Etoh}}) \quad (1)$$

$$f_2 = \mu_2 * (\frac{Ac}{Ac + K_{Ac}} + \frac{Aco}{Aco + K_{Aco}} + \frac{Pyr}{Pyr + K_{Pyr}}) * \frac{K_{inh}}{Etoh + K_{inh}} \quad (2)$$

Population dynamics

$$\frac{dX}{dt} = (f_1 + f_2) * X \quad (3)$$

Metabolite dynamics

$$\frac{dEtoh}{dt} = -Y_{Etoh/X} * f_1 * X \tag{4}$$

$$\frac{dLa}{dt} = -Y_{La/X} * f_1 * X \tag{5}$$

$$\frac{dAc}{dt} = -Y_{Ac/Etoh} * \frac{dEtoh}{dt} - Y_{Ac/X} * f_2 * X \tag{6}$$

$$\frac{dAco}{dt} = -Y_{Aco/La} * \frac{dLa}{dt} - Y_{Aco/X} * f_2 * X \tag{7}$$

$$\frac{dPyr}{dt} = -Y_{Pyr/La} * \frac{dLa}{dt} - Y_{Pyr/X} * f_2 * X \tag{8}$$

Remarks
All biomass units assume cell dry weight.

Supplementary File 2

Monoculture dFBA model: *Acetobacter pasteurianus*

Model variables

symbol	name	unit	initial value
X	Biomass concentration	mg/l	10.
Etoh	Ethanol concentration	mM	200.
La	Lactate concentration	mM	65.
Ac	Acetate concentration	mM	0.
Aco	Acetoin concentration	mM	0.
Pyr	Pyruvate concentration	mM	0.

FBA variables

symbol	name	unit
$v_{biomass}$	Biomass production flux	mmol/g*h
v_{Etoh}	Ethanol flux	mmol/g*h
v_{La}	Lactate flux	mmol/g*h
v_{Ac}	Acetate flux	mmol/g*h
v_{Aco}	Acetoin flux	mmol/g*h
v_{Pyr}	Pyruvate flux	mmol/g*h
v_{Ox}	Oxygen flux	mmol/g*h
v_{Asp}	L-aspartate flux	mmol/g*h
v_{ATP}	ATP hydrolysis flux	mmol/g*h

Parameters

symbol	name	unit	value
$Y_{Etoh/X}$	Ethanol-to-biomass yield	mmol/mg	.16
$Y_{La/X}$	Lactate-to-biomass yield	mmol/mg	.05
$Y_{Ac/Etoh}$	Acetate-from-ethanol yield	mmol/mmol	.7
$Y_{Aco/La}$	Acetoin-from-lactate yield	mmol/mmol	.3
$Y_{Pyr/La}$	Pyruvate-from-lactate yield	mmol/mmol	.15
$v_{max\ Etoh}$	Maximum ethanol consumption flux	mmol/g*h	100.
$v_{max\ La}$	Maximum lactate consumption flux	mmol/g*h	100.
$v_{max\ Ac}$	Maximum acetate consumption flux	mmol/g*h	.5
$v_{max\ Aco}$	Maximum acetoin consumption flux	mmol/g*h	.25
$v_{max\ Pyr}$	Maximum pyruvate consumption flux	mmol/g*h	.25
K_{Etoh}	Michaelis-Menten constant of ethanol	mM	10.
K_{La}	Michaelis-Menten constant of lactate	mM	10.
K_{Ac}	Michaelis-Menten constant of acetate	mM	10.
K_{Aco}	Michaelis-Menten constant of acetoin	mM	10.
K_{Pyr}	Michaelis-Menten constant of pyruvate	mM	10.
K_{inh}	Michaelis-Menten inhibition constant of ethanol	mM	.01

Dynamic FBA
Growth predicted by FBA

$$max\ v_{biomass}$$
$$subject\ to:$$

$$S * v = 0 \tag{1}$$
$$-1000 \leq v_{Etoh} \leq v_{max\ Etoh} * \frac{Etoh}{Etoh + K_{Etoh}} \tag{2}$$
$$-1000 \leq v_{La} \leq v_{max\ La} * \frac{La}{La + K_{La}} \tag{3}$$
$$-1000 \leq v_{Ac} \leq v_{max\ Ac} * \frac{Ac}{Ac + K_{Ac}} \tag{4}$$
$$-1000 \leq v_{Aco} \leq v_{max\ Aco} * \frac{Aco}{Aco + K_{Aco}} \tag{5}$$
$$-1000 \leq v_{Pyr} \leq v_{max\ Pyr} * \frac{Pyr}{Pyr + K_{Pyr}} \tag{6}$$
$$0 \leq v_{Ox} \leq 90 \tag{7}$$
$$-1.7 \leq v_{Asp} \leq 0 \tag{8}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{Etoh} + \frac{1}{Y_{Ac/Etoh}} * v_{Ac} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{9}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{La} + \frac{1}{Y_{Aco/La}} * v_{Aco} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{10}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{La} + \frac{1}{Y_{Pyr/La}} * v_{Pyr} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{11}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{Etoh} + \frac{Y_{Etoh/X}}{Y_{La/X}} * v_{La} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{12}$$
$$-1000 * \frac{Etoh}{Etoh + K_{inh}} \leq v_{ATP} - 1500 * v_{biomass} \leq 1000 * \frac{Etoh}{Etoh + K_{inh}} \tag{13}$$

Population dynamics

$$\frac{dX}{dt} = v_{biomass} * X \tag{14}$$

Metabolite dynamics

$$\frac{dEtoh}{dt} = -v_{Etoh} * X * \frac{1g}{1000mg} \quad (15)$$

$$\frac{dLa}{dt} = -v_{La} * X * \frac{1g}{1000mg} \quad (16)$$

$$(17)$$

$$\frac{dAc}{dt} = -v_{Ac} * X * \frac{1g}{1000mg} \quad (18)$$

$$\frac{dAco}{dt} = -v_{Aco} * X * \frac{1g}{1000mg} \quad (19)$$

$$\frac{dPyr}{dt} = -v_{Pyr} * X * \frac{1g}{1000mg} \quad (20)$$

Remarks
All biomass units assume cell dry weight.
All metabolite consumption fluxes are positive and production fluxes negative, except for L-aspartate, for which the consumption flux is negative.

Supplementary File 3

Monoculture ODE model: *Acetobacter ghanensis*

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Etoh	Ethanol concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM
Aco	Acetoin concentration	mM

Parameters

symbol	name	unit
μ_1	Maximum specific growth rate phase 1	h^-1
μ_2	Maximum specific growth rate phase 2	h^-1
$Y_{Etoh/X}$	Ethanol-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-to-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-to-biomass yield	mmol/mg
$Y_{Aco/X}$	Acetoin-to-biomass yield	mmol/mg
$Y_{Ac/Etoh}$	Acetate-from-ethanol yield	mmol/mmol
$Y_{Aco/La}$	Acetoin-from-lactate yield	mmol/mmol
K_{Etoh}	Monod constant of ethanol	mM
K_{La}	Monod constant of lactate	mM
K_{Ac}	Monod constant of acetate	mM
K_{Aco}	Monod constant of acetoin	mM
K_{inh}	Monod inhibition constant of ethanol	mM
w	Basal growth weight	

Modified Monod model

Growth equations

$$f_1 = \mu_1 * \frac{La}{La + K_{La}} * (w + (1 - w) * \frac{Etoh}{Etoh + K_{Etoh}}) \tag{1}$$

$$f_2 = \mu_2 * (\frac{Ac}{Ac + K_{Ac}} + \frac{Aco}{Aco + K_{Aco}}) * \frac{K_{inh}}{Etoh + K_{inh}} \tag{2}$$

Population dynamics

$$\frac{dX}{dt} = (f_1 + f_2) * X \tag{3}$$

Metabolite dynamics

$$\frac{dEtoh}{dt} = -Y_{Etoh/X} * f_1 * X \tag{4}$$

$$\frac{dLa}{dt} = -Y_{La/X} * f_1 * X \tag{5}$$

$$\frac{dAc}{dt} = -Y_{Ac/Etoh} * \frac{dEtoh}{dt} - Y_{Ac/X} * f_2 * X \tag{6}$$

$$\frac{dAco}{dt} = -Y_{Aco/La} * \frac{dLa}{dt} - Y_{Aco/X} * f_2 * X \tag{7}$$

Remarks
All biomass units assume cell dry weight.

Supplementary File 4

Monoculture dFBA model: *Acetobacter ghanensis*

Model variables

symbol	name	unit	initial value
X	Biomass concentration	mg/l	10.
Etoh	Ethanol concentration	mM	200.
La	Lactate concentration	mM	65.
Ac	Acetate concentration	mM	0.
Aco	Acetoin concentration	mM	0.

FBA variables

symbol	name	unit
$v_{biomass}$	Biomass production flux	mmol/g*h
v_{Etoh}	Ethanol flux	mmol/g*h
v_{La}	Lactate flux	mmol/g*h
v_{Ac}	Acetate flux	mmol/g*h
v_{Aco}	Acetoin flux	mmol/g*h
v_{Ox}	Oxygen flux	mmol/g*h
v_{Asn}	L-asparagine flux	mmol/g*h
v_{Pro}	L-proline flux	mmol/g*h
v_{ATP}	ATP hydrolysis flux	mmol/g*h

Parameters

symbol	name	unit	value
$Y_{Etoh/X}$	Ethanol-to-biomass yield	mmol/mg	.25
$Y_{La/X}$	Lactate-to-biomass yield	mmol/mg	.05
$Y_{Ac/Etoh}$	Acetate-from-ethanol yield	mmol/mmol	.7
$Y_{Aco/La}$	Acetoin-from-lactate yield	mmol/mmol	.3
$v_{max\ Etoh}$	Maximum ethanol consumption flux	mmol/g*h	100.
$v_{max1\ La}$	Maximum lactate consumption flux phase 1	mmol/g*h	100.
$v_{max2\ La}$	Maximum lactate consumption flux phase 2	mmol/g*h	.5
$v_{max\ Ac}$	Maximum acetate consumption flux	mmol/g*h	1.25
$v_{max\ Aco}$	Maximum acetoin consumption flux	mmol/g*h	.25
K_{Etoh}	Michaelis-Menten constant of ethanol	mM	10.
K_{La}	Michaelis-Menten constant of lactate	mM	10.
K_{Ac}	Michaelis-Menten constant of acetate	mM	10.
K_{Aco}	Michaelis-Menten constant of acetoin	mM	10.
K_{inh}	Michaelis-Menten inhibition constant of ethanol	mM	.01

Dynamic FBA

Growth predicted by FBA

$$max\ v_{biomass}$$
$$subject\ to:$$

$$S * v = 0 \tag{1}$$
$$-1000 \leq v_{Etoh} \leq v_{max\ Etoh} * \frac{Etoh}{Etoh + K_{Etoh}} \tag{2}$$
$$-1000 \leq v_{La} \leq v_{max\ La} * \frac{La}{La + K_{La}} \tag{3}$$
$$-1000 \leq v_{Ac} \leq v_{max\ Ac} * \frac{Ac}{Ac + K_{Ac}} \tag{4}$$
$$-1000 \leq v_{Aco} \leq v_{max\ Aco} * \frac{Aco}{Aco + K_{Aco}} \tag{5}$$
$$0 \leq v_{Ox} \leq 180 \tag{6}$$
$$-1000 \leq v_{Asn} \leq 0 \tag{7}$$
$$-1000 \leq v_{Pro} \leq 0 \tag{8}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{Etoh} + \frac{1}{Y_{Ac/Etoh}} * v_{Ac} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{9}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{La} + \frac{1}{Y_{Aco/La}} * v_{Aco} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{10}$$
$$-1000 * \frac{K_{inh}}{Etoh + K_{inh}} \leq v_{Etoh} + \frac{Y_{Etoh/X}}{Y_{La/X}} * v_{La} \leq 1000 * \frac{K_{inh}}{Etoh + K_{inh}} \tag{11}$$
$$0 \leq v_{ATP} - 900 * v_{biomass} \leq 0 \tag{12}$$

Population dynamics

$$\frac{dX}{dt} = v_{biomass} * X \tag{13}$$

Metabolite dynamics

$$\frac{dEtoh}{dt} = -v_{Etoh} * X * \frac{1g}{1000mg} \tag{14}$$
$$\frac{dLa}{dt} = -v_{La} * X * \frac{1g}{1000mg} \tag{15}$$
$$\tag{16}$$
$$\frac{dAc}{dt} = -v_{Ac} * X * \frac{1g}{1000mg} \tag{17}$$
$$\frac{dAco}{dt} = -v_{Aco} * X * \frac{1g}{1000mg} \tag{18}$$

Remarks
All biomass units assume cell dry weight.
All metabolite consumption fluxes are positive and production fluxes are negative, except for L-asparagine and L-proline, for which the consumption fluxes are negative.

Annex Chapter 7

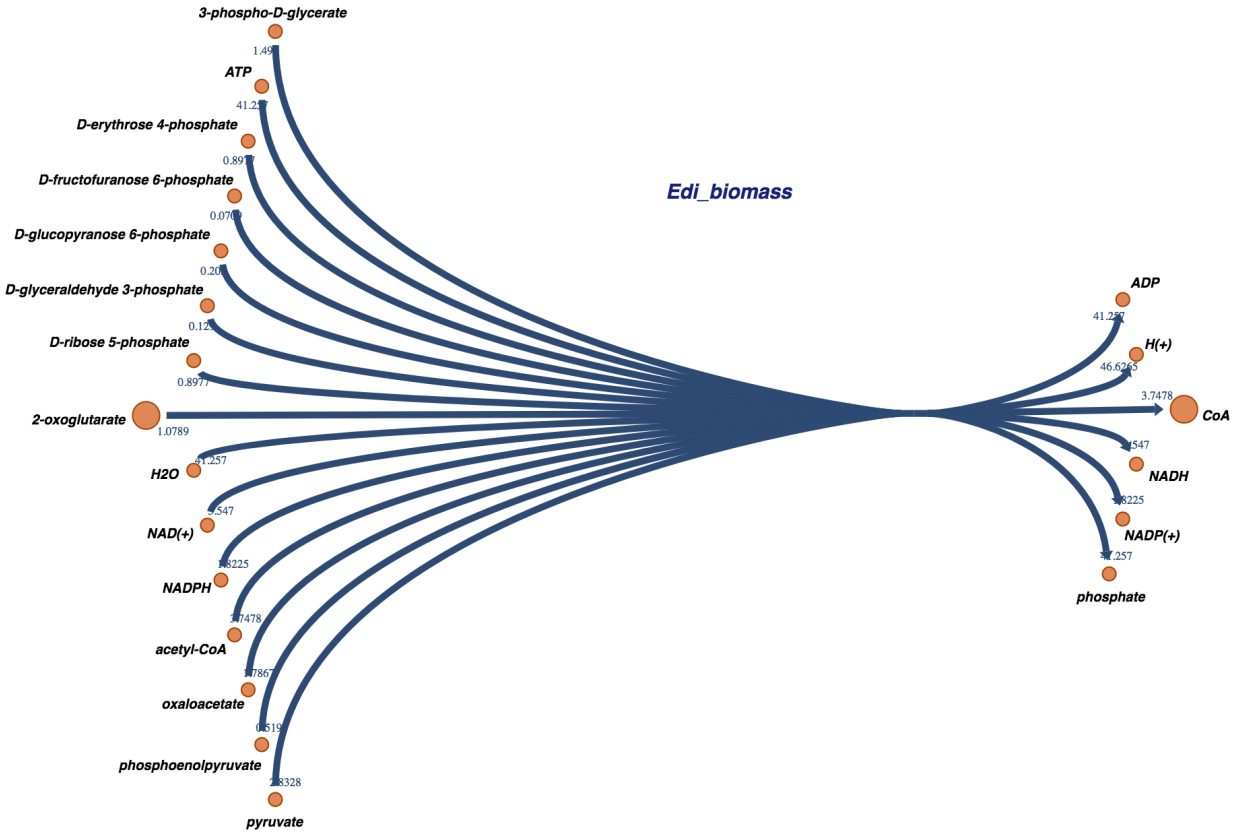
Supplementary File 1

1. Supplementary Tables

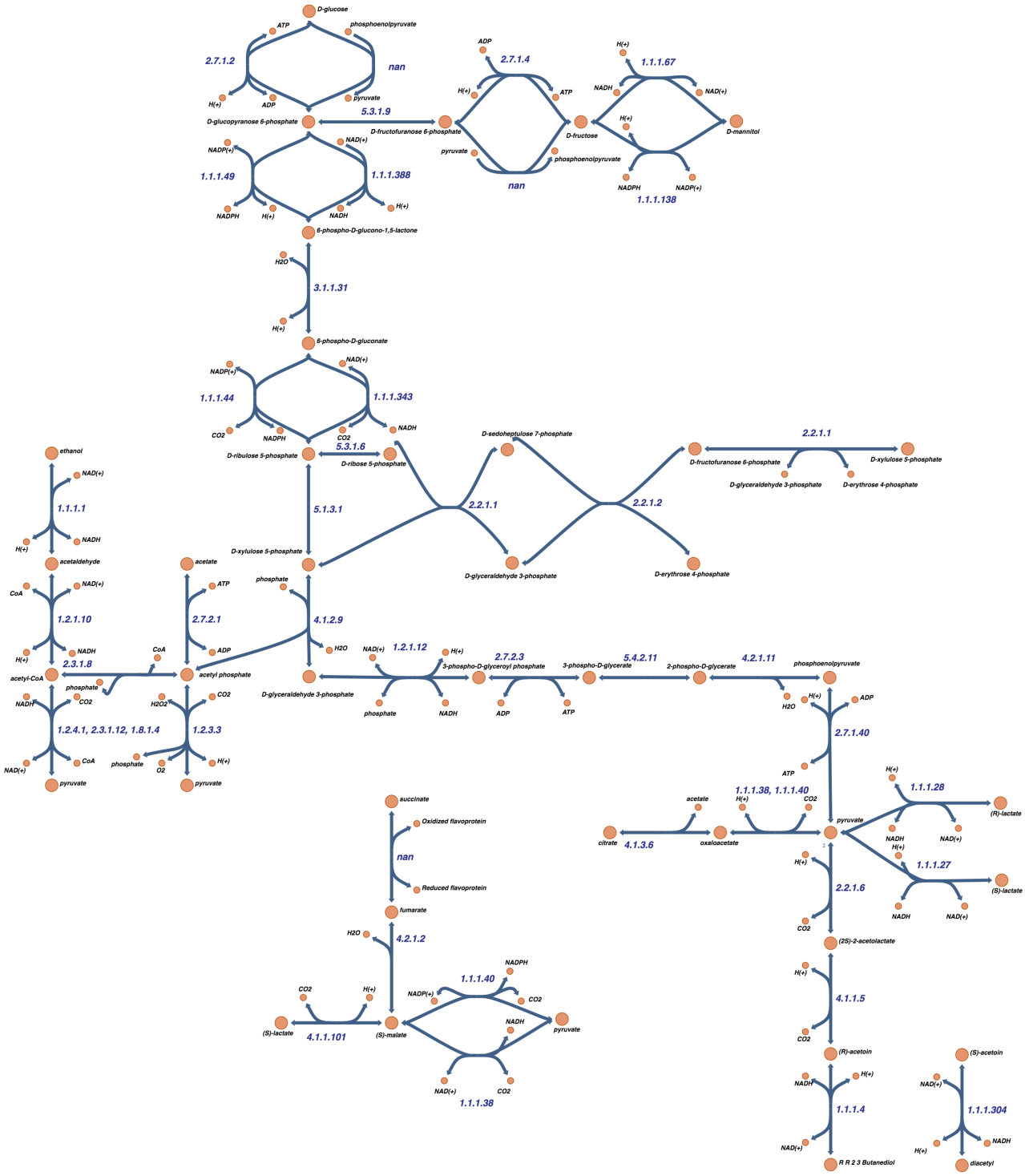
Table S1. Best blast hits of protein sequences of characterized enzymes to the *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 genomes

				Lactobacillus fermentum 222				*Lactobacillus plantarum* 80			
UniProt ID	**Annotation**	**Organism**	**Reference**	**Gene ID**	**Similarity (%)**	**Identity (%)**	**Coverage (%)**	**Gene ID**	**Similarity (%)**	**Identity (%)**	**Coverage (%)**
Q04GZ9	glucose PTS system IIA component	*Oenococcus oeni* PSU-1	Jamal *et al*., 2013	LFER_636	66.39	43.70	70	LP80_3084	61.83	38.17	77
Q04GZ8	glucose PTS system IIBC component	*Oenococcus oeni* PSU-1	Jamal *et al*., 2013	LFER_938	64.52	41.94	90	LP80_3084	65.40	47.26	100
A0NKD4	fructose PTS system IIBC component	*Oenococcus oeni* ATCC BAA-1163	Jamal *et al*., 2013	LFER_1552	67.35	47.96	20	LP80_904	54.57	35.41	88
A0NKD5	fructose PTS system IIA component	*Oenococcus oeni* ATCC BAA-1163	Jamal *et al*., 2013	LFER_1551	48.87	31.58	72	LP80_1854	59.17	34.17	66
A2Q6K5	FruA protein	*Bifidobacterium breve* UCC2003	Mazé *et al*., 2007	LFER_636	48.10	30.94	99	LP80_3084	49.00	30.09	97
A5HNZ4	annotated as ribokinase, possible phospfructokinase (PFK)	*Lactobacillus reuteri* ATCC 55730	Arskold *et al*., 2008	LFER_1545	81.58	68.75	99	LP80_732	69.23	49.83	98
A5HNZ6	annotated as ribokinase, possible phospfructokinase (PFK)	*Lactobacillus reuteri* ATCC 55730	Arskold *et al*., 2008	LFER_1545	74.10	56.39	99	LP80_732	72.37	54.93	99
Q1KM01	Fructose-1,6-bisphosphate aldolase, class II	*Lactobacillus reuteri* ATCC 55730	Arskold *et al*., 2008	LFER_RS10160	58.82	32.35	12	LP80_362	76.57	62.94	99
Q0GL75	Triose phosphate isomerase	*Lactobacillus reuteri* ATCC 55730	Arskold *et al*., 2008	LFER_648	77.60	55.60	96	LP80_1725	65.86	44.18	95
Q6ECH5	NADP-dependent mannitol 2-dehydrogenase	*Lactobacillus reuteri* ATCC 53608	Sasaki *et al*., 2005	LFER_58	94.35	88.39	100	LP80_2898	47.26	31.12	98
Q83XM4	UnkR repressor	*Lactobacillus fermentum* NRRL-B-1932	Helanto *et al*., 2006	LFER_1133	92.10	86.93	100	LP80_1985	60.80	39.51	98
Q83XM3	phosphoglucose isomerase (fruI)	*Lactobacillus fermentum* NRRL-B-1932	Helanto *et al*., 2006	LFER_RS04265	100.00	98.18	12	LP80_2834	83.11	73.11	100
Q8KMQ5	fructokinase (fruK)	*Lactobacillus fermentum* NRRL-B-1932	Helanto *et al*., 2006	LFER_856	100.00	98.97	99	LP80_1988	63.45	48.62	99
Q8KRG4	NADH oxidase NOX2	*Lactobacillus brevis* DSM20054	Geueke *et al*., 2003	LFER_1005	48.79	26.97	70	LP80_1537	78.57	64.06	99
Q6JI89	Pyruvate oxidase	*Lactobacillus plantarum* Lp80	Lorquet *et al*., 2004	LFER_1816	58.73	37.16	96	LP80_2753	99.83	99.67	100
Unknown1	N-terminal region NOX1 S. mutans	*Streptococcus mutans* NCIB 11723	Higuchi *et al*., 1993	LFER_RS05855	61.54	53.85	76	LP80_RS13055	88.89	77.78	53
Unknown2	N-terminal region NOX2 S. mutans	*Streptococcus mutans* NCIB 11723	Higuchi *et al*., 1993					LP80_2721	76.92	76.92	93
P37062	NADH peroxidase	*Enterococcus faecalis* ATCC 700802	Yeh *et al*., 1996	LFER_1005	47.97	29.94	71	LP80_2680	74.44	60.99	99

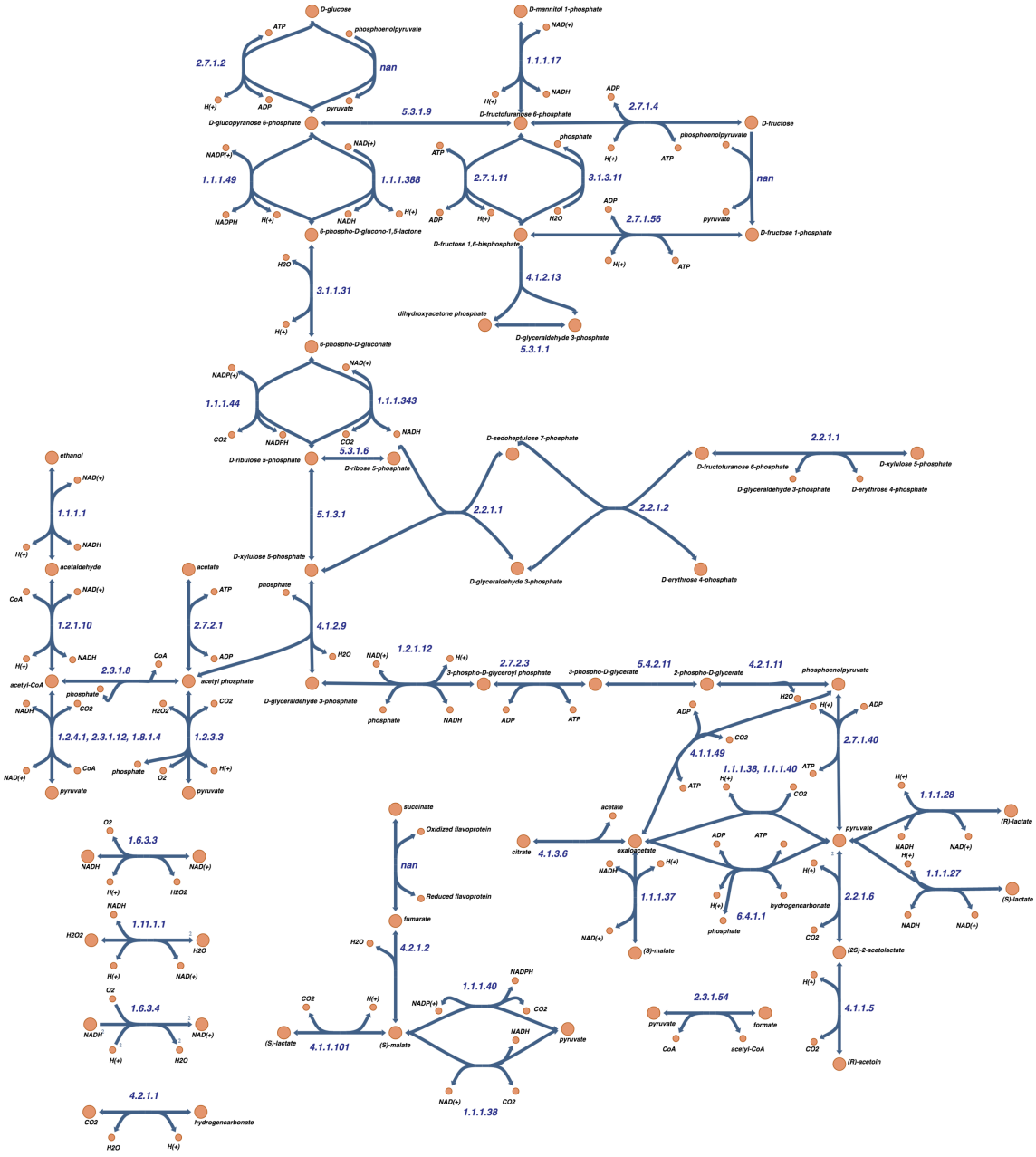
2. Supplementary Figures



Supplementary Figure 1. Universal biomass reaction to simulate microbial growth (Edirisinghe *et al*., 2016).



Supplementary Figure 2. Reconstructed metabolic network of *Lactobacillus fermentum* 222. Reactions are identified with EC numbers.



Supplementary Figure 3. Reconstructed metabolic network of *Lactobacillus plantarum* 80. Reactions are identified with EC numbers.

Supplementary File 2

Monoculture ODE model: *Lactobacillus fermentum* 222

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Glc	Glucose concentration	mM
Frc	Fructose concentration	mM
Cit	Citrate concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM
Man	Mannitol concentration	mM

Parameters

symbol	name	unit
μ_{Glc}	Maximum specific growth rate on glucose	h$^-$1
μ_{Frc}	Maximum specific growth rate on fructose	h$^-$1
μ_{Cit}	Maximum specific growth rate on citrate	h$^-$1
$Y_{Glc/X}$	Glucose-to-biomass yield	mmol/mg
$Y_{Frc/X}$	Fructose-to-biomass yield	mmol/mg
$Y_{Cit/X}$	Citrate-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-from-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-from-biomass yield	mmol/mg
$Y_{Man/Frc}$	Mannitol-from-fructose yield	mmol/mmol
K^M_{Glc}	Monod constant of glucose	mM
K^M_{Frc}	Monod constant of fructose	mM
K^M_{Cit}	Monod constant of citrate	mM
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h
$v_{max\ La}$	Maximum lactate production flux	mmol/g*h
$v_{max\ Ac}$	Maximum acetate production flux	mmol/g*h
$v_{max\ Man}$	Maximum mannitol production flux	mmol/g*h
K_{Glc}	Michaelis-Menten constant of glucose	mM
K_{Frc}	Michaelis-Menten constant of fructose	mM
K_{Cit}	Michaelis-Menten constant of citrate	mM
K_{La}	Michaelis-Menten constant of lactate	mM
K_{Ac}	Michaelis-Menten constant of acetate	mM
K_{Man}	Michaelis-Menten constant of mannitol	mM
n	Hill coefficient fluxes	

Monod model

Growth equations

$$f_{Glc} = \mu_{Glc} * \frac{Glc}{Glc + K^M_{Glc}} \tag{1}$$

$$f_{Frc} = \mu_{Frc} * \frac{Frc}{Frc + K^M_{Frc}} \tag{2}$$

$$f_{Cit} = \mu_{Cit} * \frac{Cit}{Cit + K^M_{Cit}} \tag{3}$$

$$\tag{4}$$

Population dynamics

$$\frac{dX}{dt} = (f_{Glc} + f_{Frc} + f_{Cit}) * X \tag{5}$$

Metabolite dynamics: using yields

$$\frac{dGlc}{dt} = -Y_{Glc/X} * f_{Glc} * X \tag{6}$$

$$\frac{dFrc}{dt} = -Y_{Frc/X} * f_{Frc} * X \tag{7}$$

$$\frac{dCit}{dt} = -Y_{Cit/X} * f_{Cit} * X \tag{8}$$

$$\frac{dLa}{dt} = Y_{La/X} * \frac{dX}{dt} \tag{9}$$

$$\frac{dAc}{dt} = Y_{Ac/X} * \frac{dX}{dt} \tag{10}$$

$$\frac{dMan}{dt} = Y_{Man/Frc} * Y_{Frc/X} * f_{Frc} * X \tag{11}$$

Metabolite dynamics: using fluxes

$$\frac{dGlc}{dt} = -v_{max\ Glc} * \frac{Glc^n}{Glc^n + K^n_{Glc}} * X \tag{12}$$

$$\frac{dGlc}{dt} = -v_{max\ Frc} * \frac{Frc^n}{Frc^n + K^n_{Frc}} * X \tag{13}$$

$$\frac{dCit}{dt} = -v_{max\ Cit} * \frac{Cit^n}{Cit^n + K^n_{Cit}} * X \tag{14}$$

$$\frac{dLa}{dt} = v_{max\ La} * \frac{K^n_{La}}{La^n + K^n_{La}} * X \tag{15}$$

$$\frac{dAc}{dt} = v_{max\ Ac} * \frac{K^n_{Ac}}{Ac^n + K^n_{Ac}} * X \tag{16}$$

$$\frac{dMan}{dt} = v_{max\ Man} * \frac{K^n_{Man}}{Man^n + K^n_{Man}} * X \tag{17}$$

Remarks
All biomass units assume cell dry weight.
n = 6

Supplementary File 3

Monoculture ODE model: *Lactobacillus fermentum* NCC 575

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Glc	Glucose concentration	mM
Frc	Fructose concentration	mM
Cit	Citrate concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM
Man	Mannitol concentration	mM

Parameters

symbol	name	unit
μ_{Glc}	Maximum specific growth rate on glucose	h^-1
μ_{Frc}	Maximum specific growth rate on fructose	h^-1
μ_{Cit}	Maximum specific growth rate on citrate	h^-1
$Y_{Glc/X}$	Glucose-to-biomass yield	mmol/mg
$Y_{Frc/X}$	Fructose-to-biomass yield	mmol/mg
$Y_{Cit/X}$	Citrate-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-from-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-from-biomass yield	mmol/mg
$Y_{Man/Frc}$	Mannitol-from-fructose yield	mmol/mmol
K^M_{Glc}	Monod constant of glucose	mM
K^M_{Frc}	Monod constant of fructose	mM
K^M_{Cit}	Monod constant of citrate	mM
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h
$v_{max\ La}$	Maximum lactate production flux	mmol/g*h
$v_{max\ Ac}$	Maximum acetate production flux	mmol/g*h
$v_{max\ Man}$	Maximum mannitol production flux	mmol/g*h
K_{Glc}	Michaelis-Menten constant of glucose	mM
K_{Frc}	Michaelis-Menten constant of fructose	mM
K_{Cit}	Michaelis-Menten constant of citrate	mM
K_{La}	Michaelis-Menten constant of lactate	mM
K_{Ac}	Michaelis-Menten constant of acetate	mM
K_{Man}	Michaelis-Menten constant of mannitol	mM
n	Hill coefficient fluxes	
p	Hill coefficient growth	

Monod model

Growth equations

$$f_{Glc} = \mu_{Glc} * \frac{Glc^p}{Glc^p + K_{Glc}^{M\ p}} \tag{1}$$

$$f_{Frc} = \mu_{Frc} * \frac{Frc^p}{Frc^p + K_{Frc}^{M\ p}} \tag{2}$$

$$f_{Cit} = \mu_{Cit} * \frac{Cit^p}{Cit^p + K_{Cit}^{M\ p}} \tag{3}$$

$$\tag{4}$$

Population dynamics

$$\frac{dX}{dt} = (f_{Glc} + f_{Frc} + f_{Cit}) * X \tag{5}$$

Metabolite dynamics: using yields

$$\frac{dGlc}{dt} = -Y_{Glc/X} * f_{Glc} * X \tag{6}$$

$$\frac{dFrc}{dt} = -Y_{Frc/X} * f_{Frc} * X \tag{7}$$

$$\frac{dCit}{dt} = -Y_{Cit/X} * f_{Cit} * X \tag{8}$$

$$\frac{dLa}{dt} = Y_{La/X} * (f_{Glc} + f_{Cit}) * X \tag{9}$$

$$\frac{dAc}{dt} = Y_{Ac/X} * (f_{Glc} + f_{Cit}) * X \tag{10}$$

$$\frac{dMan}{dt} = Y_{Man/Frc} * Y_{Frc/X} * f_{Frc} * X \tag{11}$$

Metabolite dynamics: using fluxes

$$\frac{dGlc}{dt} = -v_{max\ Glc} * \frac{Glc^n}{Glc^n + K_{Glc}^n} * X \tag{12}$$

$$\frac{dGlc}{dt} = -v_{max\ Frc} * \frac{Frc^n}{Frc^n + K_{Frc}^n} * X \tag{13}$$

$$\frac{dCit}{dt} = -v_{max\ Cit} * \frac{Cit^n}{Cit^n + K_{Cit}^n} * X \tag{14}$$

$$\frac{dLa}{dt} = v_{max\ La} * \frac{K_{La}^n}{La^n + K_{La}^n} * X \tag{15}$$

$$\frac{dAc}{dt} = v_{max\ Ac} * \frac{K_{Ac}^n}{Ac^n + K_{Ac}^n} * X \tag{16}$$

$$\frac{dMan}{dt} = v_{max\ Man} * \frac{K_{Man}^n}{Man^n + K_{Man}^n} * X \tag{17}$$

Remarks
All biomass units assume cell dry weight.
Metabolite dynamics using yields: p = 5
Metabolite dynamics using fluxes: n = 6; p = 1

Supplementary File 4

Monoculture ODE model: *Lactobacillus plantarum* 80

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Glc	Glucose concentration	mM
Frc	Fructose concentration	mM
Cit	Citrate concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM

Parameters

symbol	name	unit
μ_{Glc}	Maximum specific growth rate on glucose	h$^-$1
μ_{Frc}	Maximum specific growth rate on fructose	h$^-$1
μ_{Cit}	Maximum specific growth rate on citrate	h$^-$1
$Y_{Glc/X}$	Glucose-to-biomass yield	mmol/mg
$Y_{Frc/X}$	Fructose-to-biomass yield	mmol/mg
$Y_{Cit/X}$	Citrate-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-from-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-from-biomass yield	mmol/mg
K^M_{Glc}	Monod constant of glucose	mM
K^M_{Frc}	Monod constant of fructose	mM
K^M_{Cit}	Monod constant of citrate	mM
K^M_{La}	Monod inhibition constant of lactate	mM
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h
$v_{max\ La}$	Maximum lactate production flux	mmol/g*h
$v_{max\ Ac}$	Maximum acetate production flux	mmol/g*h
K_{Glc}	Michaelis-Menten constant of glucose	mM
K_{Frc}	Michaelis-Menten constant of fructose	mM
K_{Cit}	Michaelis-Menten constant of citrate	mM
K_{La}	Michaelis-Menten constant of lactate	mM
K_{Ac}	Michaelis-Menten constant of acetate	mM
n	Hill coefficient fluxes	
p	Hill coefficient lactate inhibition	

Monod model

Growth equations

$$f_{Glc} = \mu_{Glc} * \frac{Glc}{Glc + K_{Glc}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{1}$$

$$f_{Frc} = \mu_{Frc} * \frac{Frc}{Frc + K_{Frc}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{2}$$

$$f_{Cit} = \mu_{Cit} * \frac{Cit}{Cit + K_{Cit}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{3}$$

$$\tag{4}$$

Population dynamics

$$\frac{dX}{dt} = (f_{Glc} + f_{Frc} + f_{Cit}) * X \tag{5}$$

Metabolite dynamics: using yields

$$\frac{dGlc}{dt} = -Y_{Glc/X} * f_{Glc} * X \tag{6}$$

$$\frac{dFrc}{dt} = -Y_{Frc/X} * f_{Frc} * X \tag{7}$$

$$\frac{dCit}{dt} = -Y_{Cit/X} * f_{Cit} * X \tag{8}$$

$$\frac{dLa}{dt} = Y_{La/X} * \frac{dX}{dt} \tag{9}$$

$$\frac{dAc}{dt} = Y_{Ac/X} * \frac{dX}{dt} \tag{10}$$

Metabolite dynamics: using fluxes

$$\frac{dGlc}{dt} = -v_{max\ Glc} * \frac{Glc^{n}}{Glc^{n} + K_{Glc}^{n}} * X \tag{11}$$

$$\frac{dGlc}{dt} = -v_{max\ Frc} * \frac{Frc^{n}}{Frc^{n} + K_{Frc}^{n}} * X \tag{12}$$

$$\frac{dCit}{dt} = -v_{max\ Cit} * \frac{Cit^{n}}{Cit^{n} + K_{Cit}^{n}} * X \tag{13}$$

$$\frac{dLa}{dt} = v_{max\ La} * \frac{K_{La}^{n}}{La^{n} + K_{La}^{n}} * X \tag{14}$$

$$\frac{dAc}{dt} = v_{max\ Ac} * \frac{K_{Ac}^{n}}{Ac^{n} + K_{Ac}^{n}} * X \tag{15}$$

Remarks
All biomass units assume cell dry weight.
Metabolite dynamics using yields: p = 8; $K_{La}^{M} = 85$.
Metabolite dynamics using fluxes: n = 6; p = 12; $K_{La}^{M} = 85$.

Supplementary File 5

Monoculture ODE model: *Lactobacillus plantarum* NCC 2829

Variables

symbol	name	unit
X	Biomass concentration	mg/l
Glc	Glucose concentration	mM
Frc	Fructose concentration	mM
Cit	Citrate concentration	mM
La	Lactate concentration	mM
Ac	Acetate concentration	mM

Parameters

symbol	name	unit
μ_{Glc}	Maximum specific growth rate on glucose	h$^-$1
μ_{Frc}	Maximum specific growth rate on fructose	h$^-$1
μ_{Cit}	Maximum specific growth rate on citrate	h$^-$1
$Y_{Glc/X}$	Glucose-to-biomass yield	mmol/mg
$Y_{Frc/X}$	Fructose-to-biomass yield	mmol/mg
$Y_{Cit/X}$	Citrate-to-biomass yield	mmol/mg
$Y_{La/X}$	Lactate-from-biomass yield	mmol/mg
$Y_{Ac/X}$	Acetate-from-biomass yield	mmol/mg
K^M_{Glc}	Monod constant of glucose	mM
K^M_{Frc}	Monod constant of fructose	mM
K^M_{Cit}	Monod constant of citrate	mM
K^M_{La}	Monod inhibition constant of lactate	mM
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h
$v_{max\ La}$	Maximum lactate production flux	mmol/g*h
$v_{max\ Ac}$	Maximum acetate production flux	mmol/g*h
K_{Glc}	Michaelis-Menten constant of glucose	mM
K_{Frc}	Michaelis-Menten constant of fructose	mM
K_{Cit}	Michaelis-Menten constant of citrate	mM
K_{La}	Michaelis-Menten constant of lactate	mM
K_{Ac}	Michaelis-Menten constant of acetate	mM
n	Hill coefficient fluxes	
p	Hill coefficient lactate inhibition	

Monod model

Growth equations

$$f_{Glc} = \mu_{Glc} * \frac{Glc}{Glc + K_{Glc}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{1}$$

$$f_{Frc} = \mu_{Frc} * \frac{Frc}{Frc + K_{Frc}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{2}$$

$$f_{Cit} = \mu_{Cit} * \frac{Cit}{Cit + K_{Cit}^{M}} * \frac{K_{La}^{Mp}}{La^{p} + K_{La}^{Mp}} \tag{3}$$

$$\tag{4}$$

Population dynamics

$$\frac{dX}{dt} = (f_{Glc} + f_{Frc} + f_{Cit}) * X \tag{5}$$

Metabolite dynamics: using yields

$$\frac{dGlc}{dt} = -Y_{Glc/X} * f_{Glc} * X \tag{6}$$

$$\frac{dFrc}{dt} = -Y_{Frc/X} * f_{Frc} * X \tag{7}$$

$$\frac{dCit}{dt} = -Y_{Cit/X} * f_{Cit} * X \tag{8}$$

$$\frac{dLa}{dt} = Y_{La/X} * \frac{dX}{dt} \tag{9}$$

$$\frac{dAc}{dt} = Y_{Ac/X} * \frac{dX}{dt} \tag{10}$$

Metabolite dynamics: using fluxes

$$\frac{dGlc}{dt} = -v_{max\ Glc} * \frac{Glc^{n}}{Glc^{n} + K_{Glc}^{n}} * X \tag{11}$$

$$\frac{dGlc}{dt} = -v_{max\ Frc} * \frac{Frc^{n}}{Frc^{n} + K_{Frc}^{n}} * X \tag{12}$$

$$\frac{dCit}{dt} = -v_{max\ Cit} * \frac{Cit^{n}}{Cit^{n} + K_{Cit}^{n}} * X \tag{13}$$

$$\frac{dLa}{dt} = v_{max\ La} * \frac{K_{La}^{n}}{La^{n} + K_{La}^{n}} * X \tag{14}$$

$$\frac{dAc}{dt} = v_{max\ Ac} * \frac{K_{Ac}^{n}}{Ac^{n} + K_{Ac}^{n}} * X \tag{15}$$

Remarks
All biomass units assume cell dry weight.
Metabolite dynamics using yields: p = 3; $K_{La}^{M} = 47$.
Metabolite dynamics using fluxes: n = 6; p = 3; $K_{La}^{M} = 47$.

Supplementary File 6

Monoculture dFBA model: *Lactobacillus fermentum* NCC 575

Model variables

symbol	name	unit	initial value
X	Biomass concentration	mg/l	10.
Glc	Glucose concentration	mM	130.
Frc	Fructose concentration	mM	130.
Cit	Citrate concentration	mM	44.
La	Lactate concentration	mM	0.
Ac	Acetate concentration	mM	0.
Man	Mannitol concentration	mM	0.

FBA variables

symbol	name	unit
$v_{biomass}$	Biomass production flux	mmol/g*h
v_{Glc}	Glucose consumption flux	mmol/g*h
v_{Frc}	Fructose consumption flux	mmol/g*h
v_{Cit}	Citrate consumption flux	mmol/g*h
v_{La}	Lactate production flux	mmol/g*h
v_{Ac}	Acetate production flux	mmol/g*h
v_{Man}	Mannitol production flux	mmol/g*h
v_{But}	2,3 Butanediol production flux	mmol/g*h
v_{OA}	Oxaloacetate consumption flux	mmol/g*h
$v_{Acetolac}$	(2S)-2-acetolactate consumption flux	mmol/g*h
v_{OAD}	Oxaloacetate decarboxylase flux	mmol/g*h
v_{TAL}	Transaldolase flux	mmol/g*h
v_{ATP}	ATP hydrolysis flux	mmol/g*h

Parameters

symbol	name	unit	value
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h	20.
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h	20.
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h	10.
K_{Glc}	Michaelis-Menten constant of glucose	mM	95.
K_{Frc}	Michaelis-Menten constant of fructose	mM	50.
K_{Cit}	Michaelis-Menten constant of citrate	mM	1.5
lowerbound	Lower bound of constraint	mmol/g*h	-30
upperbound	Upper bound of constraint	mmol/g*h	30
n	Hill coefficient		6

Dynamic FBA

Growth predicted by FBA

$$max\ v_{biomass}$$
$$subject\ to:$$

$$S * v = 0 \quad (1)$$
$$-v_{max\ Glc} * \frac{Glc^n}{Glc^n + K_{Glc}^n} \leq v_{Glc} \leq 0 \quad (2)$$
$$-v_{max\ Frc} * \frac{Frc^n}{Frc^n + K_{Frc}^n} \leq v_{Frc} \leq 0 \quad (3)$$
$$-v_{max\ Cit} * \frac{Cit^n}{Cit^n + K_{Cit}^n} \leq v_{Cit} \leq 0 \quad (4)$$
$$0 \leq v_{La} \leq 1000 \quad (5)$$
$$0 \leq v_{Ac} \leq 1000 \quad (6)$$
$$0 \leq v_{Man} \leq 1000 \quad (7)$$
$$0 \leq v_{But} \leq 1000 \quad (8)$$
$$-.8 * \frac{K_{Cit}^n}{Cit^n + K_{Cit}^n} \leq v_{OA} \leq 0 \quad (9)$$
$$-.100 * \frac{K_{Cit}^n}{Cit^n + K_{Cit}^n} \leq v_{Acetolac} \leq 0 \quad (10)$$
$$-.100 * \frac{Cit^n}{Cit^n + K_{Cit}^n} \leq v_{OAD} \leq 0 \quad (11)$$
$$0 \leq v_{TAL} \leq 0 \quad (12)$$
$$-.1 \leq v_{Man} + .9 * v_{Frc} \leq .1 \quad (13)$$
$$lowerbound * \frac{Cit^n}{Cit^n + K_{Cit}^n} \leq v_{ATP} - 100 * v_{biomass} \leq upperbound * \frac{Cit^n}{Cit^n + K_{Cit}^n} \quad (14)$$

Population dynamics

$$\frac{dX}{dt} = v_{biomass} * X \quad (15)$$

Metabolite dynamics

$$\frac{dGlc}{dt} = v_{Glc} * X * \frac{1g}{1000mg} \quad (16)$$
$$\frac{dFrc}{dt} = v_{Frc} * X * \frac{1g}{1000mg} \quad (17)$$
$$\frac{dCit}{dt} = v_{Cit} * X * \frac{1g}{1000mg} \quad (18)$$
$$\frac{dLa}{dt} = v_{La} * X * \frac{1g}{1000mg} \quad (19)$$
$$\frac{dAc}{dt} = v_{Ac} * X * \frac{1g}{1000mg} \quad (20)$$
$$\frac{dMan}{dt} = v_{Man} * X * \frac{1g}{1000mg} \quad (21)$$

Remarks
All biomass units assume cell dry weight.
All metabolite consumption fluxes are negative and production fluxes are positive.

Supplementary File 7

Monoculture dFBA model: *Lactobacillus plantarum* NCC 2829

Model variables:

symbol	name	unit	initial value
X	Biomass concentration	mg/l	10.
Glc	Glucose concentration	mM	130.
Frc	Fructose concentration	mM	130.
Cit	Citrate concentration	mM	42.
La	Lactate concentration	mM	0.
Ac	Acetate concentration	mM	0.

FBA variables:

symbol	name	unit
$v_{biomass}$	Biomass production flux	mmol/g*h
v_{Glc}	Glucose consumption flux	mmol/g*h
v_{Frc}	Fructose consumption flux	mmol/g*h
v_{Cit}	Citrate consumption flux	mmol/g*h
v_{La}	Lactate production flux	mmol/g*h
v_{Ac}	Acetate production flux	mmol/g*h
v_{TAL}	Transaldolase flux	mmol/g*h
v_{ATP}	ATP hydrolysis flux	mmol/g*h

Parameters:

symbol	name	unit	value
$v_{max\ Glc}$	Maximum glucose consumption flux	mmol/g*h	3.
$v_{max\ Frc}$	Maximum fructose consumption flux	mmol/g*h	3.
$v_{max\ Cit}$	Maximum citrate consumption flux	mmol/g*h	1.
K_{Glc}	Michaelis-Menten constant of glucose	mM	10.
K_{Frc}	Michaelis-Menten constant of fructose	mM	100.
K_{Cit}	Michaelis-Menten constant of citrate	mM	30.
K_{La}	Michaelis-Menten constant of lactate	mM	60.
lowerbound	Lower bound of constraint	mmol/g*h	-500
upperbound	Upper bound of constraint	mmol/g*h	500
n	Hill coefficient		6

Dynamic FBA

Growth predicted by FBA

$$max\ v_{biomass}$$
$$subject\ to:$$

$$S * v = 0 \quad (1)$$

$$-v_{max\ Glc} * \frac{Glc^n}{Glc^n + K_{Glc}^n} \leq v_{Glc} \leq 0 \quad (2)$$

$$-v_{max\ Frc} * \frac{Frc^n}{Frc^n + K_{Frc}^n} \leq v_{Frc} \leq 0 \quad (3)$$

$$-v_{max\ Cit} * \frac{Cit^n}{Cit^n + K_{Cit}^n} \leq v_{Cit} \leq 0 \quad (4)$$

$$0 \leq v_{La} \leq 50 * \frac{K_{La}^n}{La^n + K_{La}^n} \quad (5)$$

$$0 \leq v_{Ac} \leq 1000 \quad (6)$$

$$0 \leq v_{TAL} \leq 0 \quad (7)$$

$$lowerbound * \frac{K_{La}^n}{La^n + K_{La}^n} \leq v_{ATP} - 1000 * v_{biomass} \leq upperbound * \frac{K_{La}^n}{La^n + K_{La}^n} \quad (8)$$

Population dynamics

$$\frac{dX}{dt} = v_{biomass} * X \quad (9)$$

Metabolite dynamics

$$\frac{dGlc}{dt} = v_{Glc} * X * \frac{1g}{1000mg} \quad (10)$$

$$\frac{dFrc}{dt} = v_{Frc} * X * \frac{1g}{1000mg} \quad (11)$$

$$\frac{dCit}{dt} = v_{Cit} * X * \frac{1g}{1000mg} \quad (12)$$

$$\frac{dLa}{dt} = v_{La} * X * \frac{1g}{1000mg} \quad (13)$$

$$\frac{dAc}{dt} = v_{Ac} * X * \frac{1g}{1000mg} \quad (14)$$

$$\frac{dMan}{dt} = v_{Man} * X * \frac{1g}{1000mg} \quad (15)$$

Remarks
All biomass units assume cell dry weight.
All metabolite consumption fluxes are negative and production fluxes are positive.