THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

In silico analysis of microbial communities through constraint-based metabolic modelling

- Investigation of the human gut microbiota and bacterial consortia in food fermentation

Parizad Babaei

Systems & Synthetic Biology

Department of Biology and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2019

***In silico* analysis of microbial communities through constraint-based metabolic modelling**

- Investigation of the human gut microbiota and bacterial consortia in food fermentation

Parizad Babaei

ISBN: 978-91-7597-876-5

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4557
ISSN 0346-718X

Department of Biology and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden

Telephone + 46 (0)31-772 1000

Cover:
Studying microbial communities using -omics and modelling approaches, Parizad Babaei 2019

Printed by Chalmers Reproservice
Gothenburg, Sweden 2019

In silico analysis of microbial communities through constraint-based metabolic modelling

PARIZAD BABAEI

Systems and Synthetic Biology
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

Microbial communities are involved in many vital biological processes from elemental cycles to sustaining human health. The bacterial assemblages are remarkably under-studied as they are reluctant to grow in the laboratory conditions. Therefore, alternative omics-based approaches and computational modelling methods have been an active area of research to investigate microbial communities physiologically, ecologically and biochemically. In this thesis different microbial consortia involved in food production and also the human gut microbiota have been modelled and investigated. In the case of the human gut microbiota, the effects of malnutrition on the overall health of children from three different countries, namely, Malawi and Bangladesh, and Sweden have been studied. In each of the first two countries, a group of malnourished children going through food therapy as well as a healthy cohort were monitored to investigate the effect of food intervention on malnutrition, with their gut microbiota being the focal point. In this project, using metagenomics data we identified the dominant strains in each cohort, reconstructed genome-scale metabolic models (GEMs) for the most abundant ones and used our models to predict diet-microbe, microbe-microbe, and microbe-host interactions. Based on our results in this project, in addition to being less diverse, the gut microbiota of malnourished children showed a lower potency regarding the production of valuable metabolites. The second investigated microbial consortia were the ones used in fermented milk products. Based on the genome sequence and also experimental data for five selected strains, we reconstructed GEMs, curated the models and performed community modelling to predict their metabolic interactions. Using the simulation outcomes, we could predict a ratio for bacterial strains used in yogurt starter culture to maximise the production of acetaldehyde which is a key contributor to yogurt's unique taste and aroma.

GEMs are powerful tools to model an organism's metabolic capabilities, and although numerous GEMs have been reconstructed, their quality control has not gained enough attention. Evaluation of a repository of semi-automatically reconstructed GEMs related to the human gut microbiota and another repository of manually curated ones was performed comparatively. Assessing these models from topological and functional aspects, it was shown that semi-automatically reconstructed models required extensive manual curation before they could be used for target-specific simulations.

In constraint-based modelling, an objective function is usually optimised under particular environmental conditions, however, in case of the microbial communities, there is no distinct and relevant objective function. Therefore, an unbiased uniform randomised sampling algorithm was implemented for microbial communities. The samples acquired from the solution space were analysed statistically to see clustering patterns of the reactions and commensalistic relationships between the community members were identified. Overall, computational modelling paves the way towards gaining a mechanistic understanding of microbial communities and provides us with testable hypotheses and insight.

Keywords: Microbial community, gut microbiota, systems biology, metagenomics, malnourishment, genome-scale metabolic model, lactic acid bacteria, fermented food production, community modelling, uniform randomised sampling

List of publications

This thesis is based on the following publications:

Paper I: Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling
Manish Kumar*, Boyang Ji*, Parizad Babaei*, Promi Das, Dimitra Lappa, Girija Ramakrishnan, Todd E. Fox, Rashidul Haque, William A. Petri, Fredrik Bäckhed, and Jens Nielsen (2018) *Metabolic Engineering*

Paper II: Challenges in modeling the human gut microbiome
Parizad Babaei, Saeed Shoaie, Boyang Ji, and Jens Nielsen (2018), *Nature Biotechnology*

Paper III: Computational design of optimum bacterial consortia for milk fermentation using genome-scale metabolic models
Parizad Babaei, Christian Chervaux, Chloë Beal, Jean-Michel Faurie, and Jens Nielsen (2019), *submitted for publication*

Paper IV: Uniform randomised sampling of microbial communities
Parizad Babaei, Promi Das, Adil Mardinoglu, and Jens Nielsen (2019), *submitted for publication*

Paper V: Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome
Promi Das, Parizad Babaei, and Jens Nielsen (2019), *submitted for publication*

Paper VI: Human gut microbiota and healthy aging: Recent developments and future prospective
Manish Kumar, Parizad Babaei, Boyang Ji and Jens Nielsen (2016), *Nutrition and Healthy Aging*

Additional publications not included in this thesis:

Paper VII: Heterogeneity of amino acids metabolism affects lung cancer prognosis
Pouyan Ghaffari, Dijana Djureinovic, Pinja Kurppa, Johanna Mattsson, Parizad Babaei, Hans Brunström, Adil Mardinoglu, Mathias Uhlen, Patrick Micke, and Jens Nielsen (2019), *submitted for publication*

Paper VIII: Memote: A community driven effort towards a standardized genome-scale metabolic model test suite
Christian Lieven , Moritz E. Beber, Brett G. Olivier , Frank T. Bergmann ,Meric Ataman , Parizad Babaei, , Jennifer A. Bartell , Lars M. Blank , Siddharth Chauhan,…, Nikolaus Sonnenschein (2019), *submitted for publication*

*Authors contributed equally to this work.

Contribution summary

Paper I: Reconstructed, curated and evaluated genome-scale metabolic models, performed pairwise growth simulations and metabolic productions, assisted in the preparation of the manuscript.

Paper II: Performed quality control assessment, analysed the results, prepared and submitted the paper.

Paper III: Reconstructed, curated and evaluated the models, analysed fermentation data, prepared and submitted the manuscript.

Paper IV: Curated the models, analysed the experimental data, formulated and implemented the sampling algorithm, analysed the results, prepared and submitted the manuscript.

Paper V: Mapped and analysed the vitamin-related genes to the reactions of the gut bacterial models, analysed the results, assisted in the preparation of the manuscript.

Paper VI: Assisted in writing and preparation of the manuscript.

Paper VII: Reconstructed the tissue-specific models, performed anti-metabolite identification using tINIT algorithm, analysed the results, assisted in the preparation of the manuscript.

Paper VIII: Tested the software, suggested the test for direct biomass precursors and proofread the manuscript.

Preface

This dissertation is submitted for the partial fulfilment of the degree of Doctor of Philosophy at the Department of Biology and Biological Engineering at Chalmers University of Technology. It is based on the work carried out between April 2015 and March 2019 in the Systems and Synthetic Biology division under the supervision of Prof. Jens Nielsen. The research was funded by the Bill & Melinda Gates Foundation, Danone Research, Novo Nordisk Foundation through the Center for Biosustainability and the Knut and Alice Wallenberg Foundation.

Parizad Babaei
March 2019

Table of Contents

Abbreviations

AA	amino acid
ATP	adenosine triphosphate
CFU	colony-forming unit
CO_2	carbon dioxide
CTP	cytidine triphosphate
EC	enzyme commission
EDR	energy dissipating reaction
EGC	energy generating cycle
EGR	energy generating reaction
$FADH_2$	flavin adenine dinucleotide (reduced)
FBA	flux balance analysis
FCA	flux coupling analysis
$FMNH_2$	flavin mononucleotide (reduced)
FVA	flux variability analysis
g_{DW}	gram dry weight
GEM	genome-scale metabolic model
GTP	guanosine triphosphate
HBM	human breast milk
ITP	inosine triphosphate
KO	KEGG orthology
LAB	lactic acid bacteria
LP	linear programming
MILP	mixed integer linear programming
mRNA	messenger RNA
NADH	nicotinamide adenine dinucleotide (reduced)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
OD	optical density
pFBA	parsimonious flux balance analysis
rRNA	ribosomal RNA
RUTF	ready-to-use therapeutic food
SCFA	short-chain fatty acid
UTP	uridine triphosphate
WHO	world health organization

"The wholeness is just as necessary for the understanding of its parts as the parts are necessary for the understanding of the wholeness."

- David Bohm, Professor of Theoretical Physics

1. Introduction

Living organisms consist of complex sub-microscopic molecular systems which govern various biological processes. Therefore, organisms could be regarded as a whole in a systemic view. System-level understanding of biology started with biological cybernetics as early as the days of Norbert Weiner more than 40 years ago (Kitano 2001, 2002). The goal of systems biology is to gain insight into biological systems at functional and regulatory levels. The scope of the systems approach could vary from an omics-based scale at genomics or transcriptomics level to smaller sub-circuits of the system consisted of a few proteins working as a logical gate (Ferrell 2009). In this way, systems biology has been likened to Newtonian planetary astronomy in which simplification and abstraction of the whole system elucidate system mechanics. Similarly, reductionist biological approaches in which biological elements are isolated and characterised individually resembles pre-Keplerian astronomical era during which the astronomical objects were named and observed separately. Similar to Kepler's efforts to catalogue astronomical objects comprehensively, advances in molecular biology methods have resulted in an avalanche of biological data which subsequently have led to the emergence of systems biology.

A biological system is more than an arithmetic sum of its components, and new capabilities emerge as the number of elements increase. For instance, a negative feedback loop consisted of three proteins can show an oscillatory behaviour which is not achievable by a bi-member feedback system. The two-member negative feedback loop can generate pluses from constant inputs which obviously is not doable by a single protein. However, it is not possible to only understand the building blocks of a system to predict its behaviour. For example, one can catalogue every part of an aeroplane, but this information cannot explain how a plane works. Similarly, systems biology seeks to elucidate phenotypes from genotypes and decipher the inner workings of life (Voit 2016).

Microbial communities comprising of bacteria, fungi, archaea, microeukaryotes, and viruses are omnipresent, complex and adaptive biological systems populating the Earth and its multicellular inhabitants. People had used selective procedures to produce their desired food long before bacteria were discovered. Starter cultures were used for milk fermentation and pickles, even processes that favour the growth of nitrogen-fixing cyanobacteria were unknowingly used for rice production (Barton and Northup 2011). Louis Pasteur, in 1864, highlighted the role of microorganisms in fermentative processes. Similar to systems biology, microbial ecology is also a relatively young discipline in biology. Microbiology was traditionally based on experimenting and characterising microbial physiology (Schaechter 2009). Founders of microbial ecology are considered by many to be Sergei Winogradsky (1856 –1953) and Martinus Beijerinck (1851 –1931) who described microbial roles in nitrogen and carbon fixation and microbial interactions in soil (Barton and Northup 2011). However, it was Ernst Haeckel who coined the term *ecology* in 1866 in his book *Generell Meorpholog* where he describes it as "physiology of relationships" and "the science of economy of external relations of organisms to each other" (Stauffer 1957). Today, the discipline of microbial ecology encompasses many fields with a focus on functionality, robustness, and evolution of various microbial communities. Research in microbial communities has been revolutionized by the availability of omics data and gains more attention day by day.

In this thesis, I have focused on computational modelling of microbial communities mainly involved in fermentative food production and the human gut microbiota. By this means,

genome-scale metabolic models for microbes were constructed and combined to simulate different microbial communities in their respective microenvironment. In the case of the human gut microbiota, the metabolic capabilities of healthy and disturbed gut bacterial communities as the result of malnutrition, have been compared. For microbial communities in starter cultures, simulations have led to the proposal of an optimum ratio between the consortia members in order to maximise the production potential of an important metabolite for the final product's taste and aroma. Furthermore, I have focused on genome-scale metabolic models from a technical point of view, comparatively assessing their quality using two model repositories. Finally, I have proposed a uniform random sampling approach for microbial communities to circumvent the need for a relevant biological objective for modelling microbial consortia.

2. Background

2.1 Microbial communities

Microbial communities are widespread in nature and are ubiquitously present in various habitats participating in biogeochemical processes via their biochemical capabilities. Populations of interacting microorganisms have influenced Earth's environmental conditions, more specifically its surface chemistry, in a way that it could support the evolution of more complicated organisms and they continue to sustain life; being key players in the maintenance of stable environmental conditions (Konopka 2009; Falkowski, Fenchel, and Delong 2008).

Definition of a microbial community may vary depending on the ecological perspective. It is generally regarded as a biological assemblage comprising of multiple interacting species coexisting in a contiguous habitat. However, there might be alternative interpretations of the abovementioned definition based on how one defines interaction or contiguous habitat. There are methanogenic microbes in anaerobic environments who could transport electrons over the range of centimetres (Lovley 2017) or even microbes in stratified saturated marine water columns generating chemical gradients via metabolite production and consumption as long as several meters (Wakeham et al. 2007), whereas soil microorganisms usually influence their adjoining surroundings on a scale of microns (Young et al. 2008).

From a functional perspective, microbial communities act as functional units that provide various ecosystem services (ES), ranging from nitrogen-fixing soil bacteria functioning as fertilizers to host-associated human gut microbiota providing energy sources and valuable metabolites for epithelial cells (Pommier et al. 2018; Wang et al. 2018). As depicted in Figure 1A, ecosystem functions increase as taxonomic breadth expands to the point that it reaches functional redundancy which, subsequently, leads to community resilience and resistance to disturbances (Konopka 2009). Functional redundancy, resistance, and resilience are all emergent properties of microbial communities which appear when a community is regarded as a system. As shown in Figure 1B, community resistance is related to the stability of the composition of a community against stress, whereas community resilience is defined as how fast a microbial community is able to return to its pristine composition. Microbial communities could be highly resilient because of relatively fast growth rates, high physiological flexibility and the ability of horizontal gene transfer and mutation of microbes. However, the time for a particular microbial community to recover its original composition might vary greatly, ranging from hours to years depending on the type, strength, and frequency of the environmental perturbation (Allison and Martiny 2008). Even in the case of an altered composition, microbial communities may still be able to perform the original functionality due to their functional redundancy. Functional redundancy occurs when phylogenetically distinct members of a community perform the same biochemical transformations. In other words, microbial communities are able to retain their functionality with different taxonomic makeups which is the case in the observed taxonomic differences in human gut microbiota across individuals while having the same genetic content (Turnbaugh et al. 2009). The same decoupling between taxonomic architecture and genetic profile have been reported for terrestrial and aquatic microbial communities (Tringe et al. 2005), green microalgae *Ulva australis* surface-associated bacterial flora (Burke et al. 2011) and plant-related archaeal and bacterial communities (Louca et al. 2017). Several other studies in bioreactors also suggest that compositionally different microbial communities can conduct similar metabolic functions (Fernández et al. 1999; Vanwonterghem et al. 2014; Wittebolle et al. 2008; Wang et al. 2011).

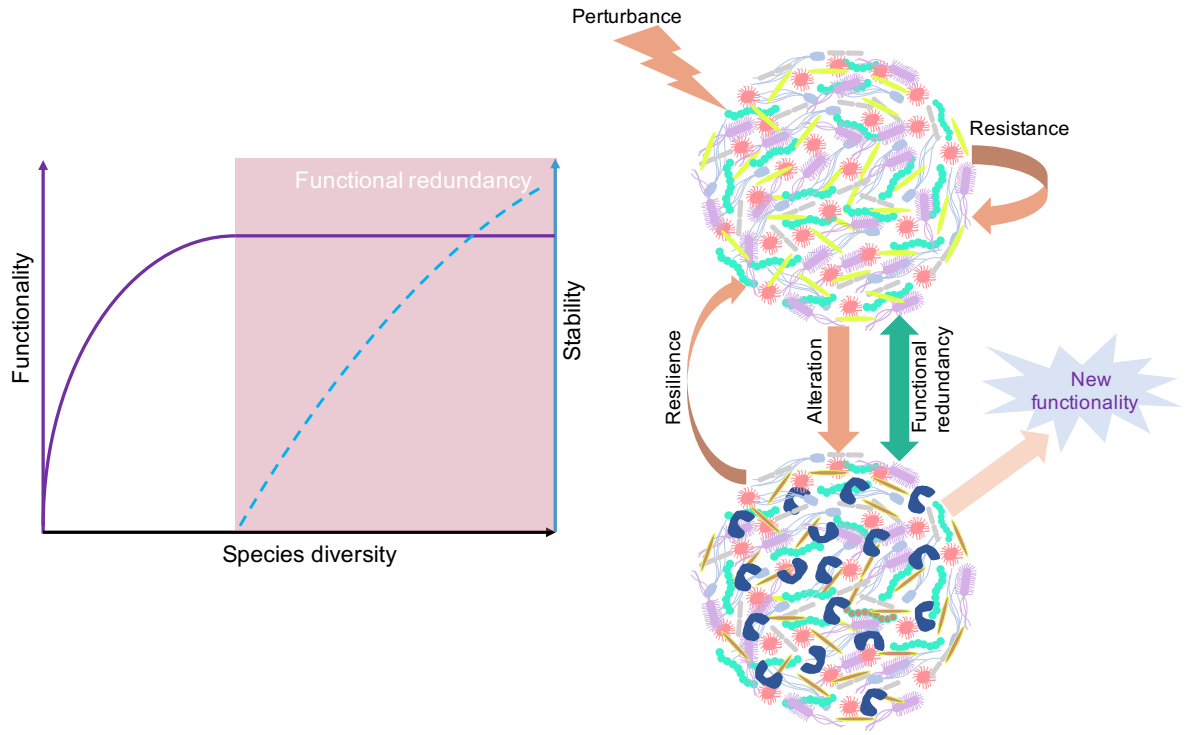


Figure 1. Resistance, resilience and functional diversity in microbial communities. A) The functionality of microbial communities increases with the number of species until it reaches functional redundancy and from that point, the stability of the community starts. B) Schematic representation of microbial communities' resistance against environmental disturbances resulting in compositional stability, its resilience which leads to compositional reconfiguration to the original state and functional redundancy where the original and the altered microbial communities have the same operational state. The altered community may also shift at the functional level by gaining new functionality.

2.2 Studying microbial communities

There are various approaches in characterising microbial communities ranging from analysis of functional and taxonomic diversity to the use of predictive models and ecological theories to provide a context in order to gain mechanistic insight. Advances in sequencing technologies and computational algorithms and power have led to the emergence of a myriad of molecular data which correspondingly have revolutionized the field of microbial ecology (Prosser et al. 2007). However, ecological theories and predictive models that are required for delineating ecosystem processes are still in their infancy. Below, omics data-based analysis and modelling strategies to elucidate microbial ecologies are discussed.

2.2.1 Metagenomics and meta-transcriptomics

Traditionally, the field of microbiology was limited to culture-dependent techniques focused on single species which subsequently restricted scientists to solely study the microbes which were culturable in a known particular medium. However, as it appeared with the advances of molecular techniques, this approach can only cover a narrow taxonomic breadth, as small as 1% in some environments, of microorganisms in both free-living and host-associated microbial communities that are recalcitrant to grow in experimental conditions (Ji and Nielsen 2015; Riesenfeld, Schloss, and Handelsman 2004). It was first mentioned by Norman Pace that our

understanding of taxonomic diversity could be increased by deploying molecular tools (Rappé and Giovannoni 2003). In their papers in the eighties, they proposed a combination of sequencing and recombinant DNA techniques to investigate microbial communities both phylogenetically and quantitatively based on ribosomal RNA macromolecules or gene sequences (Olsen et al. 1986; Lane et al. 1985). This approach transformed the field of microbial ecology and the idea continues to be employed by many scientists of the field.

Meta means "transcendent" in Greek, and metagenomics approaches have circumvented the uncluturability of a vast diversity of microorganisms. During the past four decades, the number of identified bacterial phyla has increased from 12 to 92. In the case of archaea, since their discovery as a separate domain in 1977, 26 phyla were characterised (Woese and Fox 1977; Gutleben et al. 2018). The aforementioned identified microorganisms comprise a total of 400,000 bacterial and archaeal species (Hug et al. 2016; Youssef et al. 2015). Metagenomics methodology based on the 16S rRNA marker gene as shown in Figure 2, includes amplification and subsequent detection of taxonomic diversity in DNA molecules extracted from environmental samples (Lane et al. 1985). This approach characterises the taxonomy of members of a microbial community, however, does not provide further information regarding community functionality. In order to identify other genes, functions and biochemical pathways present in an environmental sample, other methods such as shotgun sequencing are required (Eisen 2007; Venter et al. 2004). In addition to providing a robust way of sampling for genes other than 16S rRNA, this procedure also has higher randomness and breadth and is applicable for viruses which do not possess 16S rRNA gene (Eisen 2007).

Even though shotgun sequencing provides partial insight into the functionality of microbial communities, complementary approaches such as metatranscriptomics, metaproteomics, and metabolomics are employed to infer the functionality of microbial communities in different environmental conditions (Aguiar-Pulido et al. 2016). As depicted in Figure 2, metatranscriptomics is the sequencing and annotation of the collection of all mRNA molecules present in an environmental sample, shedding light on the expressed genes of the community (Moran 2009). Metaproteomics aims to catalogue the proteins in microbial consortia and link the microbial composition to its functionality, however, it is still in its infancy and current methods either have a low taxonomic resolution or are low-throughput (Wilmes and Bond 2006; Kleiner et al. 2018). Metabolomics techniques are also deployed in microbial communities to identify metabolic by-products being secreted to the microenvironment (Aguiar-Pulido et al. 2016). While all these omics data approaches are of great value and provide insight into various microbial communities, a shared downside of them is that genes, transcripts, and proteins are dissociated from the organisms and are normally not regarded in the genetic and physiological context of the microorganism that they originate, therefore, in order to put all these data in a mechanistic structure complementary modelling approaches are required.

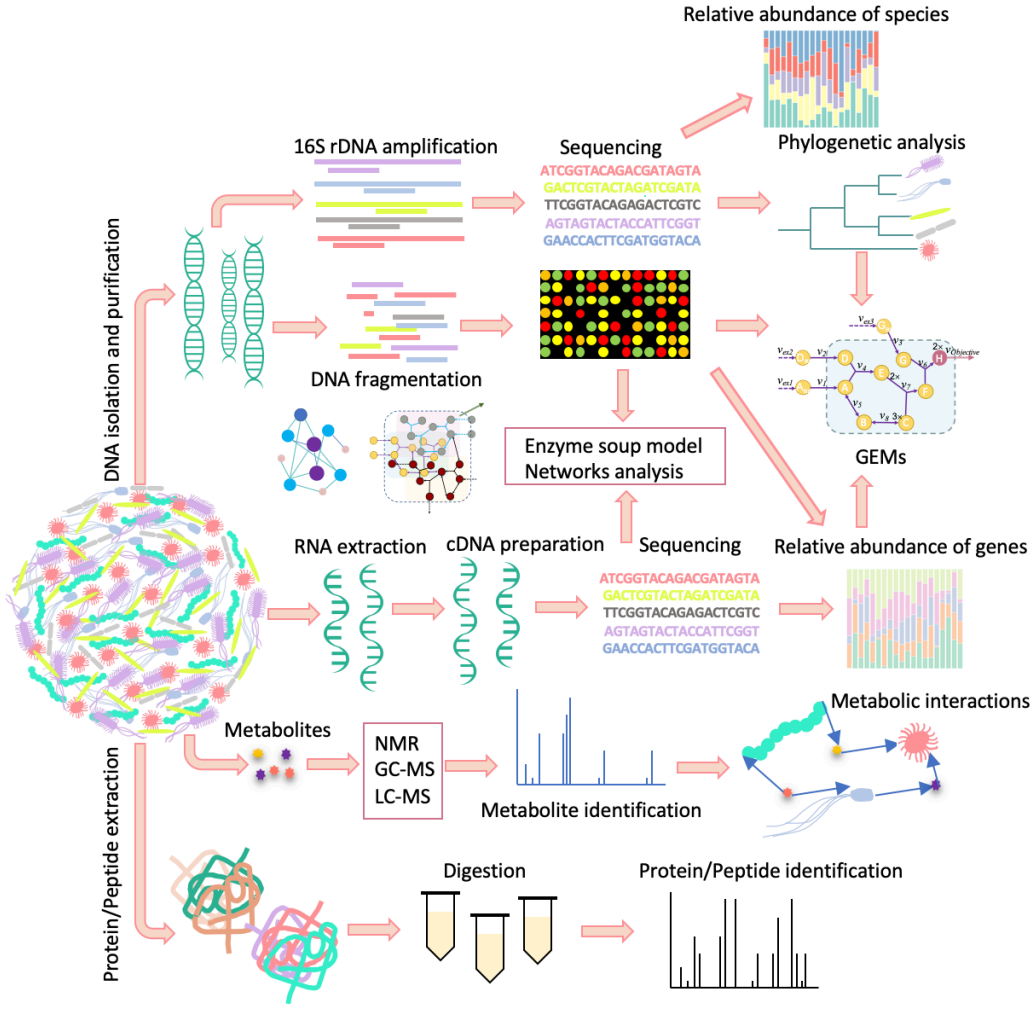


Figure 2. Omics approaches to analyse microbial communities. 16S rRNA-based studies characterise microbial communities taxonomically while whole-genome sequencing provides more insight into the operational capabilities of the microbial system. Metatranscriptomics elucidates the expressed genes in the environmental sample and metaproteomics reveals the protein contents of a sample. Metabolomics analyses could be employed to infer metabolic interactions between the community members.

2.2.2 Modelling approaches

Mechanistic understanding of the principles that govern microbial consortia depends on predictive tools to contextualize the plethora of data generated by omics-based descriptive approaches. These tools might pave the way for the generation of hypothesis regarding cause-consequence of different observations related to microbial ecosystems and their interplay with environmental states, dynamics of consortia assemblage, evolutionary path or fate of microbial communities (Zomorrodi and Segre 2016). There are various types of mathematical modelling methods that have been used to investigate microbial ecosystems which range from the classic ecological theory-based procedures (Lafferty et al. 2015; Bellucci et al. 2015) and Lotka–Volterra population dynamics models (Wangersky 1978; Stein et al. 2013) to spatial modelling (Cantrell and Cosner 2004; Cosner 2008; D'Acunto et al. 2015), game theory-based approaches (Mao, Blanchard, and Lu 2014; Lambert, Vyawahare, and Austin 2014) and agent-based

modelling to study species interactions (Hellweger et al. 2016; Kreft et al. 2013; Schluter and Foster 2012).

All the above-mentioned methods are based on relative abundances of community members and their simplified interactions in order to predict community dynamics; meanwhile, internal biochemical reactions of members are usually not considered. Internal modelling of biochemical capabilities of living organisms can be achieved through genome-scale metabolic models. These models combined with constrained-based methods could be employed to link genotype to phenotype and have become a fundamental and powerful tool in systems biology.

2.3 Genome-scale metabolic models

A genome-scale metabolic model (GEM) is a collection of all of the biochemical transformations performed by an organism of interest. Each biotransformation is carried out by a metabolic reaction which might be catalysed by one or a set of enzymes encoded by a single or a collection of genes. From the first prokaryotic genome-scale metabolic models to the more recent and complicated ones, the field of metabolic modelling has been expanding to increase gene coverage and also include additional elements such as metabolite formulas, charges, and different identifiers. These models could be integrated with various omics data to both contextualize data and also increase the predictive power of the model.

2.3.1 Definition

A genome-scale metabolic model is a mathematical representation of metabolic capabilities encoded by the genome of a target organism. Starting from an annotated genome, all available biochemical information is compiled into a matrix, called stoichiometric or S matrix. Other reactions are also added which include:

1. biomass reaction; a fictitious reaction which consumes cellular components with their molar ratio as their stoichiometric coefficient.
2. Exchange reactions; a set of reactions that define a particular medium on which the modelled organism grows.
3. Transport reactions; a set of reactions which transport metabolites from extracellular space to cytoplasm.

All of the reactions mentioned above are added to the S matrix as shown in Figure 3.

In the stoichiometric matrix, each column and row represent a reaction and a metabolite, respectively. Changes in metabolites' concentrations could be calculated using the S matrix and vector of reaction fluxes. Since metabolism is a fast process compared to the other cellular processes (such as regulation or cell division), then it is acceptable to presume that it has reached steady-state (Terzer et al. 2009). Therefore, the S matrix is used to imply a set of constraints known as the mass balance constraints. Another set of thermodynamic constraints are also used when the flux capacity of the involved reactions is known. These constraints could be categorized as follows:

1. Directionality constraints; if directionality of a reaction is known, then it could be implied by a zero lower or upper bound in case of forward or reverse reactions, respectively.
2. Uptake or secretion fluxes; changes in concentrations of extracellular metabolites could be measured and consumption or production fluxes could be calculated, and subsequently imposed as constraints on exchange reactions.
3. Internal fluxes; in case of availability of ^{13}C data, similar to the previous section, bounds on internal reactions could also be calculated and applied.

After implying all the constraints mentioned above, a feasible flux space is defined which satisfies all the mass-balanced and the thermodynamic constraints. Finally, flux through a particular reaction, usually the biomass reaction in case of prokaryotes, is optimised based on an evolutionary assumption which postulates that bacteria tend to maximise growth (Feist and Palsson 2010).This method, known as Flux Balance Analysis or FBA, calculates a flux vector which supports the optimal state (Orth, Thiele, and Palsson 2010).

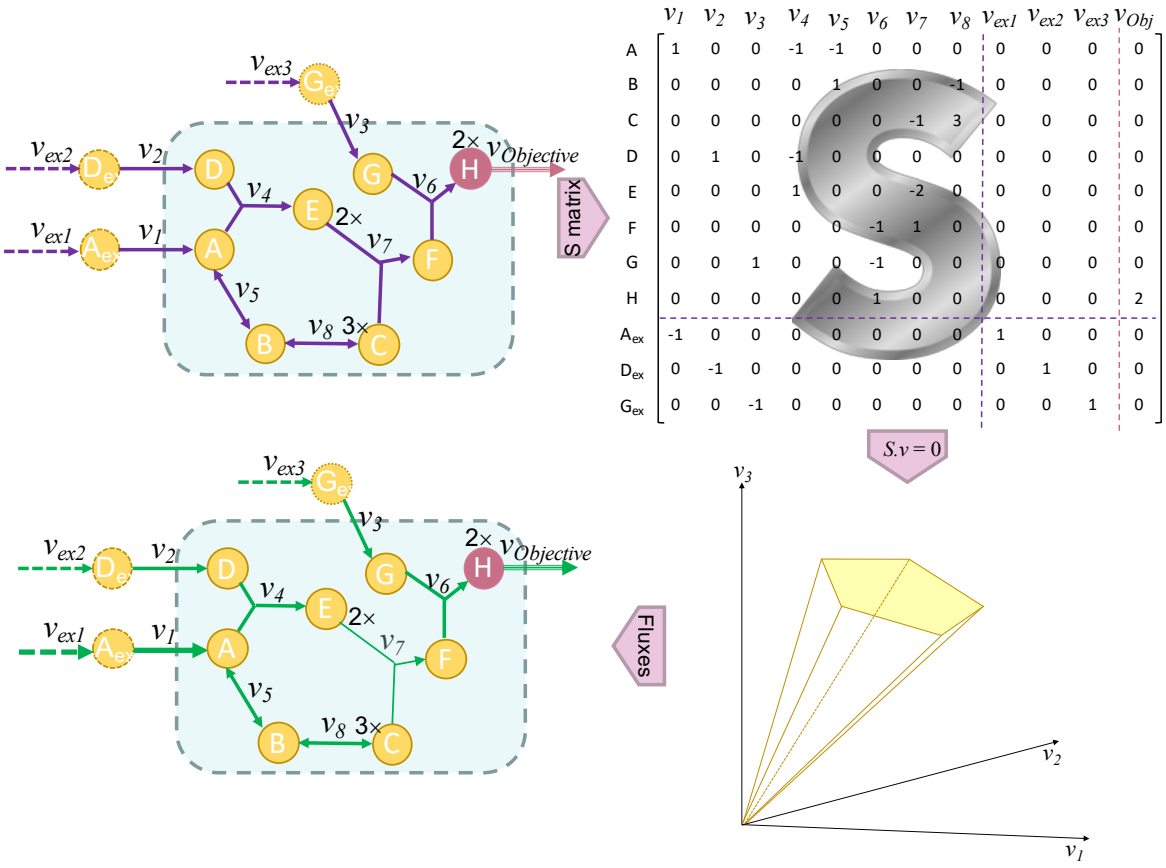


Figure 3. Schematic representation of metabolic network reconstruction and flux prediction. A simple metabolic network with exchange reactions (v_{ex1}, v_{ex2}, and v_{ex3}), transport reactions (v_1, v_2, and v_3), internal reactions, and an objective function is shown. Stoichiometric matrix related to the network is formulated and used to impose mass balance constraints, resulting in a feasible solution space. Finally, flux values are calculated by optimising the flux through $v_{objective}$.

2.3.2 Reconstruction

The first step in genome-scale metabolic model reconstruction process is genome annotation. Based on this step, a list of reactions is gathered and assembled into the stoichiometric matrix. At this stage, the GEM is called a draft network which is most probably not able to produce several or even any of the biomass precursors. A draft model, therefore, contains "gaps" or missing information due to our incomplete understanding of metabolism (Orth and Palsson 2010). A follow-up procedure, called gap-filling, is necessary to find network gaps and eliminate them by adding appropriate metabolic reactions in order to render the draft model functional. There are many different methods for reconstruction and gap-filling which could be crudely divided into manual and semi-automatic methods which are discussed below.

2.3.2.1 *Bottom-up approach*

The reconstruction procedure of early genome-scale metabolic models has been tremendously laborious and time-consuming. This bottom-up approach is based on genomic and bibliomic data, organism-specific information, protein , and physiological data if available. Afterwards, each reaction and gene annotation are checked manually to see whether it fits in the scope of reconstruction. Enzyme substrates and cofactors, reaction directionality and localization, metabolite formulas and charges, determination of biomass composition, the addition of exchange and transport reactions and many more detailed steps follows data retrieval as described in a protocol published in 2010 (Thiele and Palsson 2010). Subsequently, in an iterative process, the model is assessed for a set of quality-control and quality-assurance tests. The most critical criteria include control of mass and charge balance of the network, gap-filling analysis, identification of blocked reactions and dead-end metabolites, type-III extreme pathway analysis and their curation, and the model's ability to grow in a biologically feasible rate. It is evident that while this approach would most probably result in a high-quality model, it is not comparable, in terms of the required time and manual effort, to other more recent and advanced methods in which at least parts of the reconstruction process is performed automatically.

2.3.2.2 Semi-automatic approaches

Model reconstruction have been drastically accelerated with the advent of semi-automatic or fully automatic methods such as GEM Systems (Arakawa et al. 2006), Model Seed (Henry et al. 2010), SuBliMinaL (Swainston et al. 2011), FAME (Boele, Olivier, and Teusink 2012), RAVEN toolbox (Agren et al. 2013), CoReCo (Pitkänen et al. 2014), merlin (Dias et al. 2015), Pathway Tools (Karp et al. 2015), pyFBA (Cuevas et al. 2016),Kbase (Arkin et al. 2018), CarveMe (Machado et al. 2018), and AutoKEGGRec (Karlsen, Schulz, and Almaas 2018) that have minimised the amount of required time and manual work to produce a functional metabolic model. These methodologies work based on either previously published template models of phylogenetically close organisms or the genome sequence of an organism of interest. Thorough comparisons between the aforementioned methods have been discussed elsewhere (Hamilton and Reed 2014; Faria et al. 2018). Development of automated model reconstruction algorithms have been an area of active research with an obvious benefit of less required time, however, these tools do not completely eliminate the need for manual curation in GEMs (Karlsen, Schulz, and Almaas 2018). The reason for the need for manual curation is twofold; firstly, some information such as reaction directionality, ATP maintenance, biomass composition, intracellular pH and organism specific data remain difficult to obtain automatically and secondly, during the automatic reconstruction procedures some mathematical artefacts and misleading information, such as poor annotation or a misplaced reaction, might be added automatically to a model which would subsequently affect the model's prediction accuracy. Therefore, a trade-off between automation and manual curation is essential to reconstruct a high-quality genome-scale metabolic model which can produce biologically relevant predictions.

2.3.3 Gap-filling processes

The draft version of metabolic networks usually is not functional, meaning that at least one biomass precursor can be synthesized under steady-state by the set of reactions included in the model; in this stage, the model contains missing information. The missing information is generally divided into two categories, namely gaps and orphan reactions (Orth and Palsson 2010). Gaps are reactions occurring in the modelled organism that are absent from the network. When there is a hole in a pathway in the network, adjacent metabolites to it cannot be either consumed or produced, called dead-end metabolites. Reactions associated with a dead-end

metabolite cannot carry flux under steady-state conditions; therefore, they are *blocked*. Orphan reactions, on the other hand, are reactions that are known to occur; however, their corresponding gene or enzyme is not identified. It is estimated that up to 40% of all recognised enzymatic activities are not associated with sequence data (Pouliot and Karp 2007; Lespinet and Labedan 2005; Shearer, Altman, and Rhee 2014). GEMs can be useful in identifying the genes related to orphan reactions. For example, in the case of *P. aeruginosa* and *E. coli*, a manually curated metabolic network led to a series of annotation refinements (Reed et al. 2006; Oberhardt et al. 2008).

There are various methods for gap identification and closure in metabolic networks. Generally, these methods first find holes in the network and then suggest a set of reactions that should be added to the draft model to restore network connectivity. Some algorithms also provide genes that could be associated with the candidate reactions (Pan and Reed 2018). These methods often solve a mixed- integer linear programming problem in which a minimal set of reactions required to either render the model functional or decrease model-data inconsistencies, is found from either a universal database, list of reactions or a template functional model (Herrgård, Fong, and Palsson 2006; Reed et al. 2006; Kumar, Dasika, and Maranas 2007; Kumar and Maranas 2009; Brooks et al. 2012; Hosseini and Marashi 2017). These methods assume that the most parsimonious pathway has a higher chance to occur which would result in a smaller GEM (Biggs et al. 2015).

The MILP-based approaches are, however, computationally demanding and may not be applicable to larger genome-scale models. Therefore, alternative methods have been developed to circumvent this problem. Algorithms such as FASTGAPFILL (Thiele, Vlassis, and Fleming 2014) and GLOBALFIT (Hartleb, Jarre, and Lercher 2016) re-formulate the aforementioned MILP problem into a series of linear programming or bi-level linear programming problems, respectively. In another LP-based approach, dead-end filling (DEF), gaps are eliminated by constructing a quasi-endosymbiosis model containing two compartments; the draft model as a eukaryotic cell and a list of external reactions as the mitochondrion. It is based on co-evolution of primitive eukaryotes and mitochondria postulating that during this process the eukaryotes evolved towards the most efficient usage of dead-end metabolites such as oxygen. Other methods have also been developed based on network topology (Prigent et al. 2017), likelihood-based gene annotation (Benedict et al. 2014) and metabolic pattern recognition (Ganter, Kaltenbach, and Stelling 2014) to propose additional reactions to a network which would ultimately increase the model's predictive power. Although all these methods aim at reconciling computational predictions with experimental measurement, it has been shown in a recent study that they might add a significant number of incorrect reactions to a network, meaning that, a follow-up manual examination of the proposed reactions by any of the mentioned methods is crucial (Karp, Weaver, and Latendresse 2018).

2.3.4 Flux analysis methods

Once a draft metabolic network has gone through the gap-filling process, flux balance methods can be performed to gain functional insight into its metabolic capabilities. Constraints are necessary to define what metabolic and physiological states are "achievable" in the genetic and environmental context of the modelled organism, and flux balance methods are applied to explore the aforementioned feasible space defined by the stoichiometric and flux capacity constraints. There are various variants of flux analysis techniques that are frequently used in constraint-based modelling of metabolism.

2.3.4.1 *Flux Balance Analysis (FBA)*

Flux Balance Analysis seeks to optimise a linear objective function under a set of linear constraints. Therefore, in FBA a Linear Programming (LP) problem is solved, and the optimal value of the objective function, as well as the value for all variables that support the optimal state, are found as depicted in Figure 4. In constraint-based modelling, the objective function is usually assumed to be the flux through biomass reaction; however, it can be any physiologically meaningful objective such as the production of any metabolite of interest. The variables in the LP problem are all of the metabolic reactions included in the model (Varma and Palsson 1993; Orth, Thiele, and Palsson 2010). The linear constraints usually concern mass-conservation (quasi-steady-state assumption), thermodynamic (reversibility of the reactions), and enzyme capacity (uptake and secretion rates of transporters, known internal reaction rates) constraints. The outcome of FBA is a flux distribution state that would at the same time satisfy all the imposed constraints and support the optimal state. Even though FBA is the most frequently used method in constrained-based metabolic modelling, the choice of a relevant objective function that would be consistent with *in vivo* metabolic fluxes in many cases remains a challenge (Schuetz, Kuepfer, and Sauer 2007). Furthermore, the flux distribution found by FBA is not unique and it only provides one of the many possible flux distributions that would support optimality. Alternative solutions normally exist due to the presence of parallel biochemical pathways, and therefore, FBA alone might not be sufficient to predict physiological characteristics.

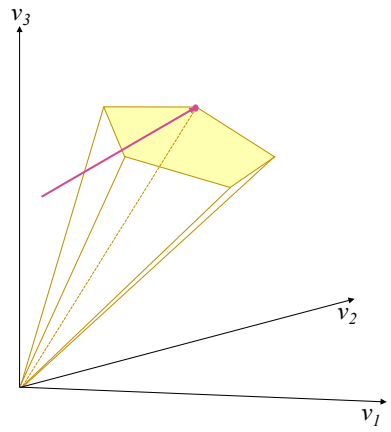


Figure 4. Feasible solution space and a linear objective function. A linear objective function is optimised within the boundaries of the solution space and its maximum value, as well as, the values of all the variables in the respective optimum state are defined.

2.3.4.2 *Parsimonious Flux Balance Analysis (pFBA)*

Parsimonious Flux Balance Analysis or pFBA is a variant of FBA that takes enzyme efficiency into account. In this method, the underlying assumption is that under the growth pressure, the most efficient pathways are the ones that use the minimum number of enzymes (Lewis et al. 2010). pFBA is a two-step linear programming problem in which growth optimisation is followed by minimisation of the sum of the net fluxes of all gene-associated reactions in the network. Therefore, pFBA leads to characterisation of the most efficient subset of genes and enzymes under particular environmental conditions. Using pFBA flux through thermodynamically infeasible loops, a set of reactions forming a closed circle, can also be circumvented since it seeks to minimise the sum of fluxes. This approach has been successfully employed to investigate gene regulation changes in *E. coli* during adaptive laboratory evolution

(Lewis et al. 2010). Additionally, in a systematic comparison with other predictive methods for flux distributions using *E. coli* model and ^{13}C measurements, pFBA showed the lowest median of error and the smallest variation in error distribution compared to other methods (Machado and Herrgård 2014).

2.3.4.3 Flux Variability Analysis (FVA)

In a metabolic model with m metabolites and n reactions, usually, the number of variables (reactions) exceeds the number of compounds ($n > m$), resulting in the network being under-constrained . It means that for an optimum state, regardless of the objective function, there exists a countless alternative number of flux vectors which could be attained as shown in Figure 5. Another reason for the existence of alternative flux distributions in a single optimal state is the presence of silent biological phenotypes that would result in achieving maximum growth rate (or optimum of any other objective function) via different metabolic pathways (Raamsdonk et al. 2001; Reed and Palsson 2004).In order to fully explore the range of allowable fluxes, it is possible to see the minimum and maximum value of flux of the reactions in which the objective function remains at its optimum point. For this means, a method called Flux Variability Analysis (FVA) is employed to reveal the range of alternative flux distributions. In FVA, objective the function is fixed at its optimum value and flux through each reaction in the network is minimised and maximised under the same mass-balance and flux capacity constraints (Mahadevan and Schilling 2003).

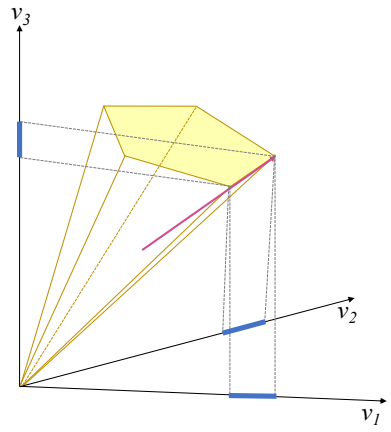


Figure 5. Flux variability in metabolic networks. There are alternative solutions for achieving an optimum. Flux through v_1, v_2, and v_3 can have any value in the highlighted ranges without affecting the value of the objective function.

2.3.4.4 Flux Coupling Analysis (FCA)

GEMs can be employed to find dependencies and relationships, known as flux coupling, between metabolic reactions based on their flux values. This feature reflects a network's topology and the level of connectivity between its reactions. In Flux Coupling Analysis or FCA, first all of the blocked reactions, i.e. the reactions unable to carry a non-zero flux under a specific uptake scenario, are identified in the network and then, all possible types of coupling between the remaining active reactions are investigated. There are mainly three types of flux coupling between non-blocked reactions:

1. Directional coupling: Reaction v_i and v_j are directionally coupled if a non-zero flux through one of them implies a non-zero flux for the other. Notice that this type of

coupling is unidirectional. $v_i \rightarrow v_j$ means that reaction j is directionally coupled with reaction i, therefore, only a non-zero flux through i would activate j, and not vice versa.

2. Partial coupling: Reaction v_i and v_j are partially coupled if a non-zero flux through either of them implies a non-zero flux through the other. Partial coupling is a bidirectional relationship; $v_i \leftrightarrow v_j$ means that v_i would have a non-zero flux *if and only if* v_j is active.
3. Full coupling: Reaction v_i and v_j are fully coupled if they are partially coupled *and* the ratio between their flux values are constant. $v_i \Leftrightarrow v_j$ means that a non-zero flux through v_i would result in a fixed flux through v_j and vice versa.

If reactions do not fall into any of the abovementioned categories, they are *uncoupled*. In FCA, first every reversible reaction is divided into two irreversible ones; therefore, flux space is restricted to positive values. Then, the reactions whose maximum is zero are identified as blocked reactions. Subsequently, a ratio between the flux values of any two reactions is defined and its minimum and maximum are calculated. Finally, reactions are grouped as shown in Figure 6. Coupling analysis helps to identify enzyme subsets which work together and provide detailed information of the topology of the network that could be used to pinpoint reactions involved in a thermodynamically infeasible loop.

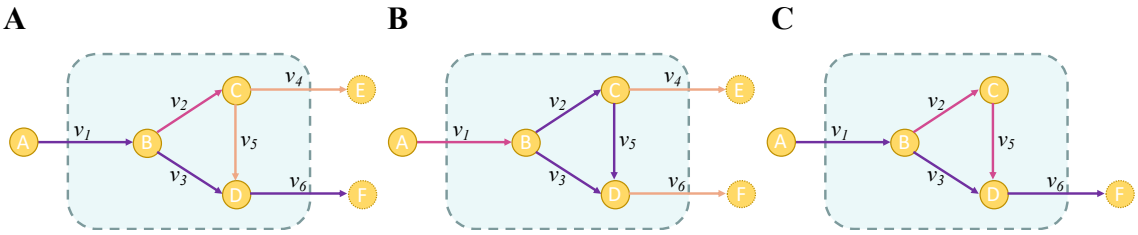


Figure 6. Different types of flux coupling in a simple network. A) v_2 is directionally coupled to v_4 and v_5 because a non-zero flux through either of these two would imply the activity of v_2, but not vice versa. B) v_4 and v_6 are partially coupled with v_1 since a non-zero flux through the first two would imply a non-zero flux through v_1 and vice versa. C) v_2 and v_5 are fully coupled as any non-zero flux though v_2 continues via v_5.

2.3.4.5 Uniform random sampling of the flux space

Flux space contains all the feasible functional scenarios of a metabolic network under its physiochemical laws. Uniform random sampling of the flux space is an unbiased means of characterising its contents. The flux space is a polyhedron in n-dimensional space defined by mass-balance and flux capacity constraints as shown in Figure 3. First, a parallelepiped is found which encloses the feasible space as tight as possible. Then, random sampling is performed by assigning uniformly random weights on its spanning edges (Price, Schellenberger, and Palsson 2004). By uniformly sampling the solution space a set of random samples are generated as depicted in Figure 7 which can show probability distribution of possible flux values by the reactions and can also be used to investigate statistical correlations between the reactions. For example, by calculating pairwise correlation coefficients, correlated sets of reactions known as *co-sets* can be identified. While $|r_{ij}| = 1$ correlation coefficient means prefect co-dependence between reaction i and j, $r_{ij} = 0$ implies that the pair are independent. There are also imperfectly correlated reactions in a network ($0 < r_{ij} < 1$). This approach is helpful in informative experimental design; measurement of the flux of any of the reactions in a perfectly correlated set is sufficient to know the flux through all of them.

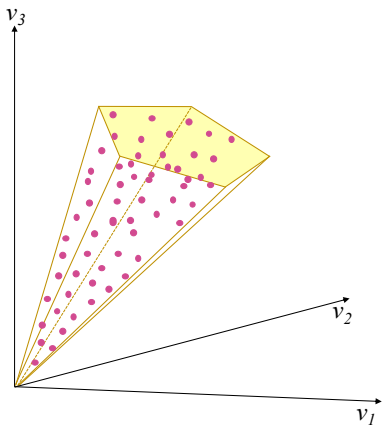


Figure 7. Uniform random sampling of the feasible solution space.

2.3.5 Evaluation of modelling predictions

Genome-scale metabolic networks are employed to compute testable predictions regarding physiological properties such as optimal growth state, gene essentiality and biochemical potentials and incapabilities of the modelled organism. For bacterial models, the growth rate is calculated by optimising flux through the biomass reaction of the network. If the predicted value is too high compared to the experimentally measured one, it might be due to the presence of falsely included reactions, wrong estimation of biomass precursors' molar ratios, incorrect growth or non-growth associated ATP values. On the other hand, if the predicted value is too slow compared to the actual rate, the reason could be one or a series of limiting biomass precursors; cellular components that cannot be synthesized efficiently. In this case, normally, the bottlenecks are identified and curated until the predicted growth rate reaches the experimentally observed one (Thiele and Palsson 2010).

Binary growth profiles of microorganisms of various carbon or nitrogen sources could also be predicted and evaluated using GEMs. This approach has been deployed for numerous models using BIOLOG substrate utilization data in order to evaluate and expand the models (Puchałka et al. 2008; Bochner, Gadzinski, and Panomitros 2001; Babaei, Marashi, and Asad 2015; Henry et al. 2009; Balagurunathan et al. 2012; Henry et al. 2017; Norsigian et al. 2018; Oberhardt et al. 2008). Results of predictions regarding substrate utilization could be categorized into 4 groups; true positives, false positives, true negatives, and false negatives. While true positives and true negatives show an agreement between computation and experiments, false positives and false negatives pinpoint to the pathways that require further refinements.

Another common test for accuracy of metabolic model predictions is the ability to predict gene essentiality. With the advent of molecular genetic tools and high-throughput phenotypical assays data regarding gene knockout phenotypes has expedited. It is also quite fast to acquire growth predictions for single or combinatory gene knockout predictions using a genome-scale metabolic model as these models include gene-protein-reaction rules. Similar to substrate utilization data, in case of discrepancy the model should be investigated to find the root of the disagreement. In case of false positives, an incomplete modelled reactome might be expanded to be able to reflect phenotypical observations and for false negatives, a falsely incorporated reaction might be the culprit (O'Brien, Monk, and Palsson 2015).

Flux through exchange reactions calculated by exo-metabolomics data or internal reactions based on ^{13}C labelling experiments data could be both used as a constraint or as a means of evaluation. For example, in case of exo-metabolomics data availability, consumption fluxes

could be used to fix the corresponding exchange reactions' bounds to the experimentally measured values while predicted production fluxes are used to evaluate the model's accuracy or vice versa. Additionally, the models should also be able to predict an organism's biochemical inability, meaning that if an organism is known not to consume or produce one or a set of certain metabolites, the model should be able to reflect it.

Inconsistent predictions cannot be completely removed since some of them might be due to other cellular processes or constraints which are not in the scope of metabolic models. For example, cellular spatial constraints, protein costs, post-translational modifications, and regulatory processes influence metabolism without being considered in GEMs.

2.3.6 Quality control of GEMs

In addition to a model's predictive ability, there are other features that one should test to increase the quality of a GEM iteratively. These characteristics include:

1. Network connectivity in terms of permanently blocked reactions and dead-end metabolites. In a well-connected and flux consistent network, all of the reactions should be able to carry non-zero flux when all of the exchange reactions are unbounded.
2. Percentage of gene-associated reactions; It is important for a GEM to have relatively high gene-associated reactions, otherwise, if a reaction is in the network without any GPR rules, there should be biochemical data supporting its existence in the modelled organism.
3. The absence of thermodynamically infeasible cycles. Reactions in a metabolic network might form a loop which violates thermodynamics second law since a non-zero flux through a closed loop would require flux against the chemical potential of metabolites.
4. Percentage of biomass precursors that are directly consumed from the media. There is a certain degree of auxotrophy in many modelled organisms, however, when the majority of biomass precursors are directly taken from the media and consumed by the biomass reactions, one might suspect that such a shortcut might be a side-effect of addition of reactions proposed by the gap-filling method which tries to minimise the number of added reactions.
5. The absence of energy-generating cycles; it has been shown in a recent study that wrong directionality of energy-generating reactions coupled with an internal loop might lead to excess production of energy, consequently leading to up to 25% increase in the growth rate production (Fritzemeier et al. 2017).

2.4 Metabolic modelling of microbial communities using GEMs

GEMs can serve as structured scaffolds for omics data contextualization and are powerful predictive tools for single cell's physiological characteristics. With the increasing availability of high-throughput omics data and ever increasing number of reconstructed GEMs, it is now possible to extend them to multispecies systems which would offer insight into metabolic machinery and web of interactions in microbial communities. Several frameworks have been developed for this means which vary considering their scope, resolution, *a priori* assumption, and potential usages. Some of the most frequently used methods to simulate microbial communities are the gene-soup approach, compartmentalization, separation of species-level and community-level objective functions, and dynamic modelling which are discussed below.

Gene-soup or enzyme-soup (also known as mixed-bag) approach is the simplest of all of the methods mentioned above to interrogate a community's overall metabolic capabilities. As shown in Figure 8A in this method, a model is built upon the annotation of meta-omics data obtained from an environmental sample. Briefly, a list of present enzymatic reactions is gathered and the whole community is regarded as a supra-organism without accounting for the

existing boundaries between the consortia members (Abubucker et al. 2012; Henry et al. 2016; Faria et al. 2016). Subsequently, reactions are compiled into a single stoichiometric matrix and finally, a generalized biomass reaction is formed and optimised (Tobalina et al. 2015). In this procedure, any two reactions could potentially be connected and it is more suited for under-studied microbial communities as there is a limited requirement for *a priori* assumptions (Taffs et al. 2009). The focus in this approach is on environment-consortia interactions and it does not provide species-resolution information such as metabolic or biological interactions. There is also a lack of accuracy associated with neglecting the cellular boundaries.

To predict cross-species metabolic exchanges and biological interactions, a set of GEMs can be linked together in a compartmentalized model, similar to a eukaryotic cell in which each organelle is represented as a separate compartment as depicted in Figure 8B. In this approach, S matrices are combined into a "meta-stoichiometric matrix". In the compartmentalized model each compartment is one member of the modelled microbial consortium and these units interact through transmembrane reactions via a shared environment. Relative abundance of microbes in the community is used to formulate a weighted linear combination of the biomass reactions of the community members which is used as the community objective function (Biggs et al. 2015). This procedure is the most frequently used method to simulate microbial communities because of its simplicity, species-level resolution and potential to scale-up for larger communities and has resulted in reasonable agreements with experimental observations (Stolyar et al. 2007; Shoaie and Nielsen 2014; Bordbar et al. 2010; Klitgord and Segrè 2010; Freilich et al. 2011; Heinken et al. 2013; Khandelwal et al. 2013; Nagarajan et al. 2013; Shoaie et al. 2013; Ye et al. 2014). However, the assumption of a linearly combined biomass function in this approach is difficult to justify biologically and might not apply to every microbial community. To circumvent this, one way is to separate species-level and community level objective functions. For example, in OptCom method, a community-level objective function is optimised in a nested bi-level optimisation as shown in Figure 8C. The outer or the community-level objective function is subject to the inner problem's objective function, or individual biomass reactions (Zomorrodi and Maranas 2012). Another study concerning two separate objective functions is the CASINO toolbox, in which the community-level and species-level objective functions are optimised iteratively until the local and global optimums converge (Shoaie et al. 2015). In addition to a simple objective function, the compartmentalization approach assumes microbial communities to be in steady-state which would not allow metabolite accumulations and abundance changes over time. These community characteristics can be captured using dynamic FBA (dFBA) which performs time-dependent growth optimisation of each community member independent of the others according to the nutrient availability (Mahadevan, Edwards, and Doyle III 2002). In this framework, kinetic expressions for nutrient utilization are required, and concentration of metabolites and cell densities can change, rather than being in steady-state. Therefore, dFBA enables us to monitor a microbial community over time rather than providing a snapshot of community-level metabolism. It can also be combined with other approaches, such as OptCom, to upgrade them into dynamic-level resolution (Zomorrodi, Islam, and Maranas 2014). dFBA has been implemented to study spatio-temporal and metabolic dynamics of small microbial consortia comprised of two or three microbes (Harcombe et al. 2014; Zhuang et al. 2011; Chiu, Levy, and Borenstein 2014; Louca and Doebeli 2015; Tzamali et al. 2011). Even though dFBA circumvents the need for optimisation of a single objective function, paucity of kinetic parameters, on the other hand, hampers its use for more complicated and little-understood microbial communities. Besides, it is more time-intensive and computationally expensive to implement dFBA compared to its simpler version.

There are other approaches to simulate microbial communities with the help of genome-scale metabolic models. For instance, GEMs have also been combined with Agent-based modelling to simulate biofilm formation by *Pseudomonas aeruginosa* (Biggs and Papin 2013) and also polymicrobial microbial communities (Bauer et al. 2017) to explore spatio-temporal dynamics of the consortium. All of these computational advances pave the way towards a deeper understanding of dynamics and metabolic interactions of microbial communities.

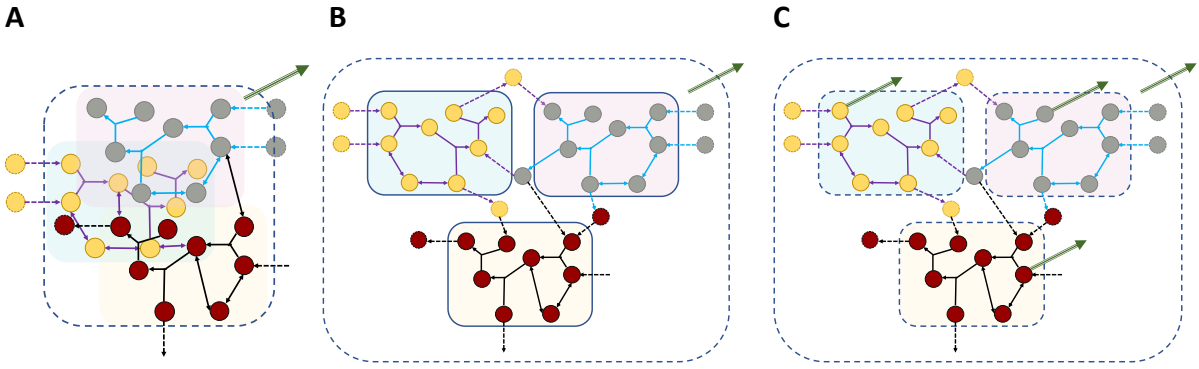


Figure 8. Microbial community simulation using genome-scale metabolic models. A) enzyme-soup. B) Compartmentalisation and C) bi-level optimisation approach. In each case the green arrows represent the objective function. In enzyme-soup and compartmentalisation methods, one community objective function is optimised while in a bi-level method there are inner and outer objective functions.

2.5 Gut microbiota

Microbial communities colonize almost every nook and cranny of the human body. Their diversity and abundance differ greatly depending on the site of the body they thrive on, but the gut dwellers form the most densely populated microbial community amongst all. This dense collection of microorganisms comprising of bacteria, archaea, viruses, and eukarya, is called the human gut microbiota and their collective genomic content is referred to as the gut microbiome (Thursby and Juge 2017). The gut microbiota is immensely diverse; including more than 1,000 species, however, it is dominated by two bacterial phylotypes: Bacteroidetes and the Firmicutes (Kovatcheva-Datchary, Tremaroli, and Bäckhed 2013). It has been estimated that the human gut microbiota outnumbers human somatic cells by a factor of 10 (Bäckhed et al. 2005; Savage 1977); however, a more recent study has suggested the ratio is actually 1:1 (Sender, Fuchs, and Milo 2016). Regardless of the ratio, their physiological role in human health is undeniable and being increasingly characterised. The gut colonisers are involved in gut homeostasis (Jandhyala et al. 2015; Lin and Zhang 2017), defence against pathogens (Pultz et al. 2005; Kommineni et al. 2015; Ubeda, Djukovic, and Isaac 2017), energy harvest (Cani et al. 2019; Rosenbaum, Knight, and Leibel 2015), immune system maturation and response (Kim et al. 2018; Hooper, Littman, and Macpherson 2012), and synthesis of neurotransmitters (Luan, Wang, and Cai 2019). These functions may be disrupted as a result of an altered microbial composition (dysbiosis). Perturbed gut microbiota is associated with many diseases, ranging from metabolic disorders such as obesity to neurobehavioral conditions like autism, that are being steadily recognized (Chang and Lin 2016; Hughes, Rose, and Ashwood 2018).

2.5.1 Metabolic functionality of the human gut microbiota

The gut microbiota is an essential metabolic repertoire which is complementary to the human digestive system. One of the most critical functions of the human gut microbiota is digestion of food components that are otherwise undigestible (Rowland et al. 2018). The gut commensals contribute to the human digestive systems by providing enzymes which are not encoded by the human genome for the breakdown of complex polysaccharides and polyphenols as well as the synthesis of vitamins. The gut bacteria mainly rely on carbohydrates which have escaped the digestive enzymes and reached the colon. In the absence of such substrates though, bacteria shift to fermentation of other energy sources such as proteins, resulting in metabolites that might be detrimental for human health. Colonic fermentation of proteins is associated with higher risk for colorectal cancer since it results in the production of branched-chain fatty acids and other potentially toxic metabolites such as thiols and amines (Kovatcheva-Datchary, Tremaroli, and Bäckhed 2013).

2.5.1.1 Short-chain fatty acids

Short-chain fatty acids are the primary saccharolytic fermentation end products of the gut microbiota using complex carbohydrates as substrates. Acetate, propionate, and butyrate are the most abundantly produced SCFAs with the molar ratio in the range of 3:1:1 to 10:2:1 (Rowland et al. 2018). These metabolites play vital roles in the host metabolism regulation (Koh et al. 2016). Acetate, the most abundant SCFA, is transported via the bloodstream to other organs where it is used in cholesterol metabolism and lipogenesis (Nicholson et al. 2012). Propionate is also carried in the blood and transferred to the liver where it is involved in gluconeogenesis (Tremaroli and Bäckhed 2012). Bacterially produced butyrate is the preferred energy source for the colonic epithelial cells. Colonocytes of germ-free mice are energy deprived and show a decreased expression of enzymes in TCA cycle (Donohoe et al. 2011). Even though butyrate has been reported to have anti-cancer potentials by inducing apoptosis in tumour cells in some studies, the opposite effect has been revealed in some others (known as the butyrate paradox) (Tilg et al. 2018). It has been shown in a study that tumour load was decreased by antibiotic treatment and less intake of fiber (Belcheva et al. 2014). The butyrate paradox, still an open question, highlights the need for the mechanistic understanding of the multi-factor host-microbe interactions.

2.5.1.2 Amino acids

Among many metabolic capabilities characterised in the gut microorganisms' biochemical repertoire is the ability to produce amino acids from nitrogen sources using the energy extracted from dietary carbohydrates (Metges 2000; Matteuzzi, Crociani, and Emaldi 1978; Neis, Dejong, and Rensen 2015). Lysine and threonine of microbial origin were identified in rat models that were fed a protein-free diet containing fermentable carbohydrates and labelled ^{15}N in $^{15}NH_4Cl$ (Torrallardona et al. 1996). Even though amino acid absorption occurs mainly in the small intestine, the importance of amino acid production by the gut microflora should not be neglected. There is some evidence for this phenomenon; for instance, peptide transporters have been identified in rabbit colon (Döring et al. 1998) and utilization of microbially produced essential amino acids has been reported in pigs and rats (Torrallardona et al. 1996; Torrallardona et al. 1994). Additionally, enrichment of amino acid biosynthesis genes in the human gut microbiome has been reported (Gill et al. 2006), however, further studies are required to gain a clearer understanding of the contribution the gut flora to amino acid production in humans.

2.5.1.3 Vitamins

Biosynthesis genes for most of the vitamins are not encoded by the human genome and these nutrients need to be provided exogenously; either through diet or by the human gut commensals which can synthesize these micronutrients *de novo*. These microorganisms are capable of synthesizing vitamin K and the majority of B vitamins which are absorbed in the colon, unlike the diet-originated vitamins that are absorbed in the small intestine (Said and Mohammed 2006). Folate, for instance, categorized in B vitamins group, can be synthesized by certain species of bifidobacteria. It has been shown that hepatic and faecal folate concentration increases after administration of bifidobacteria in rats and humans, respectively (Pompei et al. 2007; Strozzi and Mogna 2008). Cobalamin, another vitamin in B group, can be produced by *L . reuteri* and propionibacteria which are residents of the human colon and can, therefore, provide this vitamin for the host. Vitamin K, also known as menaquinone, has been reported to be bacterially produced. Although, in this case, the majority of daily requirement is provided from the diet since germ-free model animals can still get sufficient menaquinone (Suttie 1995; Davidson et al. 1998).

2.5.2 Composition

The taxonomic composition of the complex gut microflora is diverse and host-specific. In general, Bacteroidetes and Firmicutes prevail in the gut flora, followed by Actinobacteria and Proteobacteria. These four phyla together account for 97% of the gut microbial population (Donaldson, Lee, and Mazmanian 2016; Rosenbaum, Knight, and Leibel 2015). The two predominant phyla are seen in an inverse association meaning that whenever one is dominant the other is present in smaller diversity (Huttenhower et al. 2012). The ratio between these two phyla (Firmicutes/Bacteroidetes) changes during an individual's lifetime from ranging from 0.4 in infancy to 10.9 in adulthood and decreasing as one ages to 0.6 (Mariat et al. 2009). Bacteroidetes are specialized fiber degraders since as high as 10% of the genome of some of the species in this phylum is dedicated to polysaccharide degradation resulting in the production of fermentation end products such as acetate (Xu et al. 2007). Firmicutes, on the other hand, can produce butyrate, the colonocytes' preferred energy source, using acetate produced by Bacteroidetes (Louis and Flint 2017). Many factors affect the composition of the human gut microbiota; amongst which, diet, host genetics, environmental factors, and early microbial exposure play critical roles. Despite its dynamic nature, the adult gut microbiota is relatively stable, as a dominant core microbiota consisting of 40% of its species is persistent for at least one year in people (Martínez, Muller, and Walter 2013). However, this composition is prone to change as a result of antibiotic treatment, bacterial infection, long-term shift in lifestyle and GI surgery (Rodríguez et al. 2015).

2.5.3 Development

The first microbial exposure depends on the mode of delivery. It has been reported that the microbiota of infants that are delivered naturally is similar to that of their mothers' vaginal communities dominated by *Lactobacillus* and *Prevotella* spp., whereas the microbiota of babies delivered through C-section resembles adult skin microbiota dominated by *Staphylococcus* and *Corynebacterium* spp. The neonate's microbiota has been shown to be analogous across different body parts (Dominguez-Bello et al. 2010). C-section delivery has been linked to the development of gastrointestinal symptoms, sensitivity to nutritional allergens, allergic rhinoconjunctivitis, and also higher risk of childhood-onset type I diabetes during the first year after birth in neonates (Laubereau et al. 2004; Renz-Polster et al. 2005; Cardwell et al. 2008).

The maternal microbiota, regardless of the mode of delivery, forms the microbial "seed" for the infant. This microbial inoculum subsequently increases in diversity until it reaches an adult-like composition after 2-5 years (Rodríguez et al. 2015). Many postnatal factors influence the infants' microbiota development process; namely, sanitation, diet (breast milk or formula), gestational age and antibiotic treatment. Colonization of the infant's gut starts with facultative anaerobes and is followed by strict anaerobic genera such as *Bacteroides*, *Clostridium*, and *Bifidobacterium*. A neonate's gut is initially colonized by Proteobacteria and Actinobacteria and has a low diversity, then progress towards higher diversity and dominance of Bacteroidetes and Firmicutes occurs (Bäckhed 2011). By the age of one, infants acquire a unique microbial composition which continues to mature towards an adult gut flora (Yatsunenko et al. 2012). This is one of the reasons behind the importance of food interventions in early infancy since it plays a key role in the maturation of the intestinal microbiota which in turn, influences the child's growth and neurodevelopment (Borre et al. 2014).

The majority of the studies concerning the composition of the human gut microbiota rely on faecal samples and sequencing 16S rRNA gene using next-generation sequencing techniques. Faecal samples, however, may not reflect what resides in the human gut accurately. Besides, the 16S rRNA identification methodologies have a resolution level which allows characterisation of microorganisms in the level of genus or species. Therefore, it remains a challenge to accurately describe the real diversity of healthy adults' gut microflora. However, the consensus is that this microbial community is dominated by a coalition of Bacteroidetes and Firmicutes and once matured, it is relatively stable unless it is exposed to host-related changes (Faith et al. 2013).

2.5.4 Diet and malnourishment

Diet is a primary modulator of the gut microbiota. The increasing trend of microbiome-associated diseases in the past century may pinpoint the hypothesis that changes in lifestyle, and more specifically daily nutritional intakes affect the microbial composition and function which have shown to have detrimental consequences. For instance, the western diet which is deficient in fiber and high in fat and simple carbohydrates may lead to long-term loss of beneficial bugs in the gut community (Sonnenburg et al. 2016).Therefore, it is not surprising to see a perturbed gut microbial composition in case of malnourished children. According to the World Health Organization's (WHO) definition malnutrition occurs when the diet is deficient, excess, inadequate or imbalanced in energy or nutrients (World Health Organization 2016).

Infant gut bacterial maturation is characterised by a rise in bifidobacteria; facultative anaerobes which pave the way towards blooming of anaerobes by their protective effects against pathogens. Depletion of *Bifidobacterium longum* leads to gut microbial compositional shift which is associated with severe acute malnutrition. Subsequently, an altered community of microbes would not be able to harvest energy, provide vitamins, and have a functional barrier effect which would lead to infections and other harmful consequences (Million, Diallo, and Raoult 2017). For instance, it has been shown that an immature gut microbiota combined with diet plays a crucial role in under-development of children diagnosed with kwashiorkor (a form of severe malnutrition). In this study, the gut microbiota of homozygotic twin pairs discordant for kwashiorkor was transferred to germ-free mice, and it was observed that mice lost weight severely on a poor diet. Additionally, the weight loss was only partially regained when the modelled mice for kwashiorkor underwent food therapy (Smith et al. 2013). Another study identified a group of 12 gut symbionts which were exclusively present in healthy children and absent from their stunted counterparts in a cohort living in a close area (Tidjani Alou et al.

2017). These findings highlight the pivotal role of the gut microbiota in postnatal growth and potential for microbiotherapy for the treatment of malnutrition (Schwarzer 2018).

2.6 Microbial consortia in food production

The spectrum of benefits gained by microbial metabolic engines extends to include fermented food production. Microbes, either fermenting in single form or a consortium, have been traditionally used to manufacture a variety of products; wine by grape juice fermenting yeast, dairy products by milk souring lactic acid bacteria, and fish and milk products are a few examples in which microbial fermentation results in both flavour and aroma development and creating an environment which would prevent spoilage or growth of pathogenic strains (Bokulich et al. 2016; Cocolin and Ercolini 2015). Food microbiology has traditionally followed a reductionist approach; however, nowadays it frequently benefits from high-throughput sequencing methodologies and computational modelling procedures. Metagenomic and metatranscriptomics have been employed to explore metabolic interaction within microbial consortia and even optimise them towards the desired direction.

2.6.1 Lactic acid bacteria

In spontaneous food fermentation, lactic acid bacteria are among the abundant microorganisms. Lactic acid bacteria (LAB) are gram-positive, non-spore forming, microaerophilic bacteria that typically convert carbohydrates to lactic acid through the non-respiratory metabolism (Bachmann et al. 2017). There are two main categories of hexose fermentation in these bacteria, namely, homo-fermentative and heterofermentative pathways and each lead to a different set of end-products. Glycolysis in homofermentative strains results in lactate formation, while in heterofermentative strains CO_2, ethanol, and acetate are also formed in addition to lactate (Kandler 1983). Lactic acid bacteria are classified as obligate homofermentative, obligate heterofermentative and facultative heterofermentative. Homofermentative LABs ferment glucose via Emden–Meyerhoff pathway in which 2 moles of lactate and ATP are produced from 1 mole of glucose, and pyruvate is the main branching point. In heterofermentative LABs, phosphoketolase pathway is used for glucose catabolism which results in the production of 1 mole of ATP, lactate, and ethanol, while acetyl-phosphate is the main branching point of metabolism. LAB act food-specific; homofermentative LABs are used as the sole microbial strains in dairy and meat fermentation, and work together with heterofermentative LABs on other raw material such as vegetables. (Gaenzle 2015)

2.6.2 Fermented milk products

It is estimated that humans have been consuming fermented milk products for as far as 12,000 years ago (Hill et al. 2017). Lactic acid bacteria and bifidobacteria produce a variety of compounds during milk fermentation which contributes to the unique taste and aroma of the final product. For example, in yogurt, there are two LAB strains as the bacterial starter culture, namely *Lactobacillus delbrueckii subsp. bulgaricus* and *Streptococcus thermophilus*, the former acidifies the environment by producing lactate and the latter contributes to yogurt's unique flavour by producing acetaldehyde. There have been attempts to enhance *S. thermophilus*'s ability to produce acetaldehyde by metabolic engineering strategies (Chaves et al. 2002). Other volatiles such as diacetyl, acetoin, acetone, and 2-butanone are found in yogurt which are microbially produced and contribute to the flavour (Cheng 2010), therefore, it is evident that optimisation for production of a range of compounds along with other features of yogurt (optimum ratio between the starter culture bacteria), would be demanding and challenging. Computational models have the potential to provide testable hypotheses for achieving the optimum product in food microbiology.

2.6.3 Probiotics

Lactic acid bacteria and bifidobacteria display probiotic potentials. Probiotic strains are non-pathogenic live microorganisms in food that are able to deliver certain health benefits when provided in adequate amount. Historically, consumption of live bacteria in food gained popularity over a century ago when Élie Metchnikoff hypothesized that the longevity of Bulgarian farmers is because of their daily intake of yogurt since the bacteria found in yogurt have anti-ageing capabilities (Brown and Valiere 2004; Metchnikoff 2004).

Proposed health benefits for probiotics include alleviation of lactose intolerance, prevention against infections, positive influence on immune system and treatment of antibiotic-resistant diarrhoea (Ouwehand, Salminen, and Isolauri 2002; Schrezenmeir and de Vrese 2001). However, the scientific evidence for the claimed health benefits remains to be investigated as the current information is relatively sparse and conflicting as their survival through the human gastrointestinal tract and mechanism of action and dynamics of colonization in the gut are still debated. (Zmora et al. 2018).

Some studies have shown that the presence of the probiotics in the gut and their subsequent delivery of the proposed health benefits, is restricted to the consumption of the bacteria supplemented food (Wang et al. 2015; McNulty et al. 2011). Whether probiotics have an effect on the host's gut microflora is also not clear; as contradictory reports have shown both possible scenarios of modulating the microbial composition (Ferrario et al. 2014; Wang et al. 2015) or ineffectiveness on the gut dwellers make-up (Kristensen et al. 2016; Laursen et al. 2017). A recent study, using an invasive approach of colonoscopy on humans, have shown that the ability of probiotic strains to colonize the mucus layer depends on the host, probiotic strain, and the region of the gut, therefore, multiple factors play key roles in probiotic dynamics in the gut (Zmora et al. 2018).

3. Results & discussion

3.1 PAPER I & V: Genome-scale metabolic models and pairwise simulations in health and disease

Diet, especially in early infancy when the fragility of the gut microbiota is high, can influence the gut microbial make-up in terms of richness, evenness, and diversity (Brüssow 2016). Correspondingly, a diet deficient in certain nutrients or calories in case of malnourished children would alter their gut microbiota composition drastically as shown by metagenomics studies on the faecal samples of the children from Bangladesh and Malawi (Smith et al. 2013; Subramanian et al. 2014). These studies investigated the differences in taxonomic configurations of the gut microbiota between health and malnutrition based on either 16S rRNA analysis (Subramanian et al. 2014) or metagenome shotgun sequencing (Smith et al. 2013; Blanton et al. 2016). The malnourished children underwent food therapy which transiently influenced their health and maturation of the gut microbiota; however, underlying mechanisms for governing the difference between a healthy and diseased gut bacterial community requires mechanistic models. Therefore, in paper I, we have introduced a constrained-based metabolic modelling approach to investigate the biochemical capabilities of the gut microbiota in different states of health and dietary intakes. By this means, we have analysed the relative taxonomic abundance of the gut bacteria in three groups of children from Bangladesh (Subramanian et al. 2014; Blanton et al. 2016) and Malawi (Smith et al. 2013) each consisted of healthy and malnourished cohorts and a group of healthy Swedish infants (Bäckhed et al. 2015). For each community, we selected the 20 most abundant species to simplify the modelling platform by restricting the species to the dominant ones. An overall of 68 species was chosen, and their phylogenetic relationships and genetic properties are shown in Figure 9A.

Since low diversity has been associated with a dysfunctional gut (Menni et al. 2017), we calculated the Shannon-Wiener index (Hurlbert 1971) for four datasets of the aforementioned communities. In the Shannon-Wiener index as shown below, P_i is the relative abundance of species i, and S is the number of species in the microbial community. As depicted in Figure 9B, the Swedish community has a significantly higher diversity index compared to the other groups ($P < 0.01$). Firmicutes to Bacteroidetes ratio (F/B) was also calculated and was found to be significantly lower in Swedish cohort compared to the data from other the two other countries as shown in Figure 9C ($P < 0.01$). F/B ratio is associated with disorders such as Inflammatory Bowel Disease (IBD) and obesity (Sokol et al. 2009; Turnbaugh et al. 2006).

$$Shannon - Wiener\ index = -\sum_{i=1}^{S} P_i log_2(P_i)$$

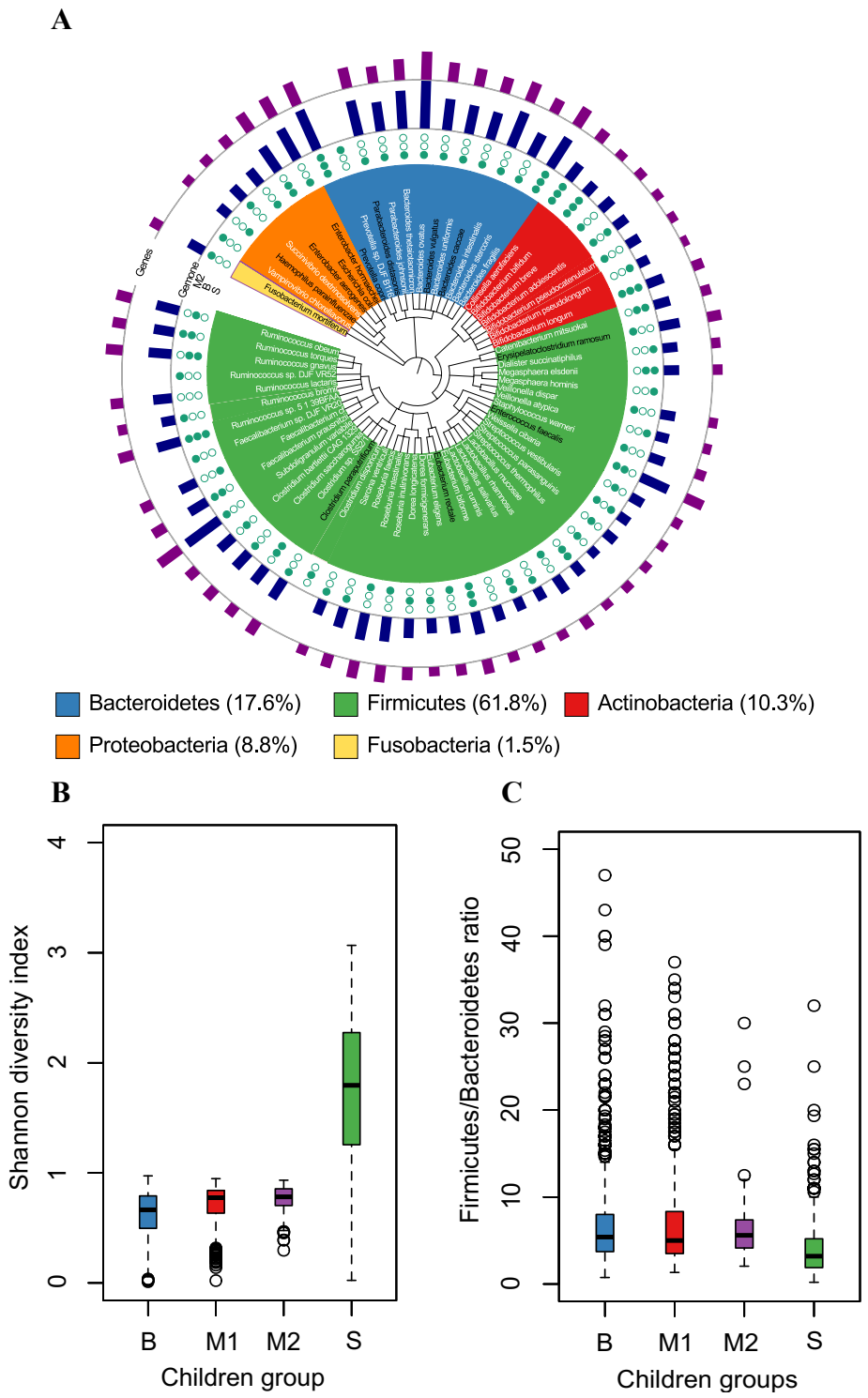


Figure 9. Gut microbes in healthy and malnourished children. A) Phylogenetic relationship between the 20 most abundant species identified in Swedish (S), Bangladeshi (B), and Malawian (M2) populations. Species written in black are pathogenic or conditionally pathogenic bacteria. The bars on the outermost circle shows the number of the genes, and the circle with blue bars represents genome size. Solid and empty green circles mean presence and absence in each of the communities respectively. B) Shannon- Wiener diversity index and C) F/B ratio of the gut microbiota in all groups.

The species whose genomes were not fully sequenced were excluded (10 species) and genome-scale metabolic models were reconstructed for the remaining 58 species. As explained in section 2.3, a GEM is a collection of an organism's biochemical capabilities in the context of its physiochemical, thermodynamic, genetic and environmental constraints. Draft models were reconstructed using Kbase platform (Arkin et al. 2018) which is based on the ModelSEED pipeline (Henry et al. 2010), followed by a gap-filling step in the same platform to render the draft models functional. Since reactions with low confidence score might be added to the draft models during the gap-filling process, the GEMs that had at least 75% gene-associated reactions were retained for further curations, analysis, and validation. Curations include anaerobic growth, reactions directionality, metabolic tasks, and improvement of annotations.

Oxygen pressure is relatively low in the gut environment; therefore, the majority of the gut commensals either grow anaerobically or micro-aerobically. To enable anaerobic growth, all oxygen-dependent essential and synthetic lethal reactions were identified. Subsequently, these reactions were replaced by alternative reactions in which the main substrates and products were the same as the original reactions, but the chemical conversion was performed with molecules other than oxygen. In the case of facultative anaerobic species, both oxygen-dependent and oxygen-independent reactions were included so that micro-aerobic and anaerobic growth could be simulated. Reaction directionalities with a focus on energy-generating reactions were curated in a way that no excess energy would be generated due to the wrong directionality.

Metabolic capabilities, in terms of the ability to produce Short-Chain Fatty Acids (SCFAs) and Amino Acids (AAs), of the models were investigated by adding a temporary exchange reaction for each of them one at a time and optimising for their production while biomass reactions was fixed at its optimum. Based on data availability in the scientific literature, case-specific manual curations were performed to close the gap between predictions and *in vitro* observations in case of inconsistency.

Models were also improved in regards to supplementary information such as chemical formulas for metabolites, EC numbers for enzymatic reactions, and nomenclature identifiers for both metabolites and reactions from external databases. We examined the metabolic distance between each pair of models based on their reactions. The metabolic distance was calculated as subtracting Jaccard similarity index from 1 (metabolic distance = 1 – Jaccard index). Jaccard index between two sets, A and B, is defined as below:

$$\text{Jaccard index} = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A} \cup \mathbf{B}|}$$

Features of the models and their metabolic distances are depicted in Figure 10.

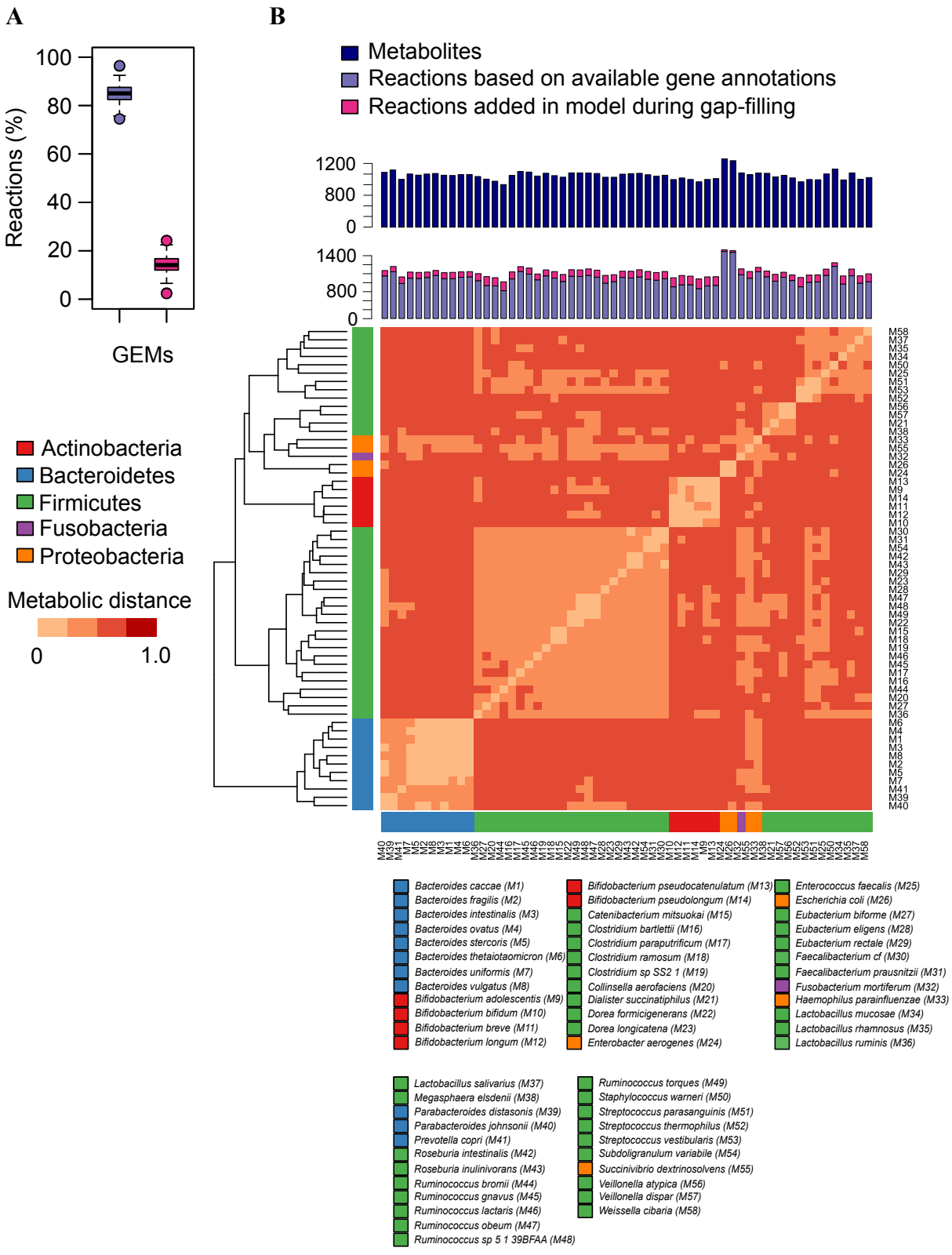


Figure 10. Metabolic features and distance between the models. A) Percentage of the gene-associated reactions and reactions added in gap-filling. B) The number of metabolites and reactions in each GEM shown by orange and blue/pink bars. Clustergram shows the metabolic distance between each pair of the models and how species in the same phylum cluster together.

Once GEMs were gap-filled and curated as explained in the previous section, we simulated their growth rate on complex media and validated the predicted growth rate using experimental data for six of them, namely *Bacteroides thetaiotaomicron* (M6), *Bifidobacterium adolescentis* (M9), *Eubacterium rectale* (M29), *Faecalibacterium prausnitzii* (M31), *Prevotella copri* (M41), and *Roseburia inulinivorans* (M43) as shown in Figure 11A. These strains were grown anaerobically in YCFA medium, and the tested models were able to predict anaerobic growth rates with sufficient consistency.

Since the children were either breast-fed or under food treatment using Ready-to-Use-Therapeutic Food (RUTF), we simulated the growth of the modelled organisms on each of these diets as media. In each case, the exchange reactions were adjusted according to the nutrients present in each diet and flux through the biomass reaction was optimised. As depicted in Figure 11B, predicted growth rates were either higher on Human Breast Milk (HBM) or similar in both cases. These results pinpoint the importance of breastfeeding in supporting the growth of gut dwellers in line with previous studies (Subramanian et al. 2015). Additionally, the composition of RUTF could be ameliorated so that it would provide essential nutrients for the gut microbiota.

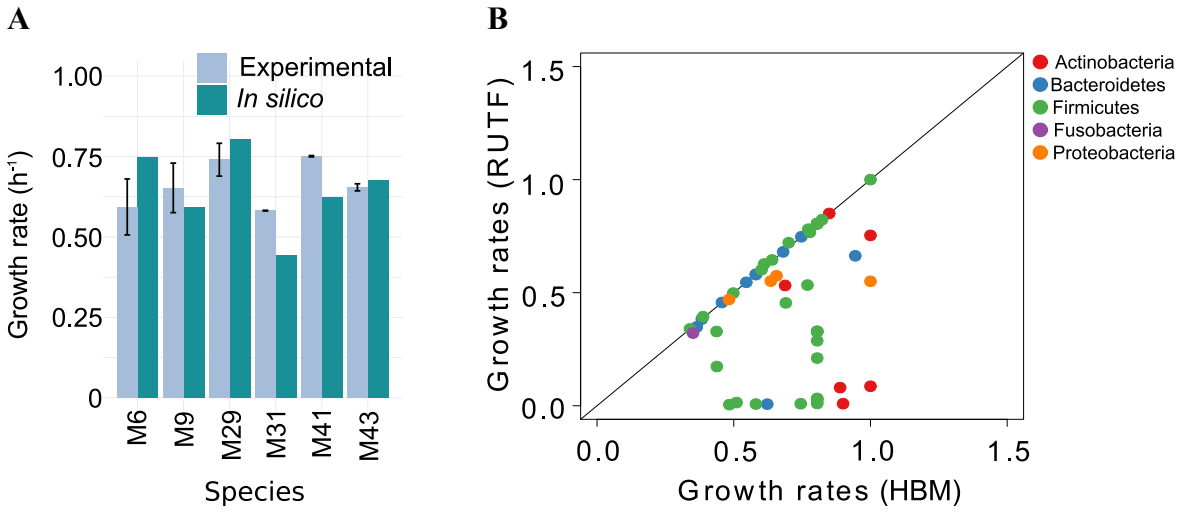
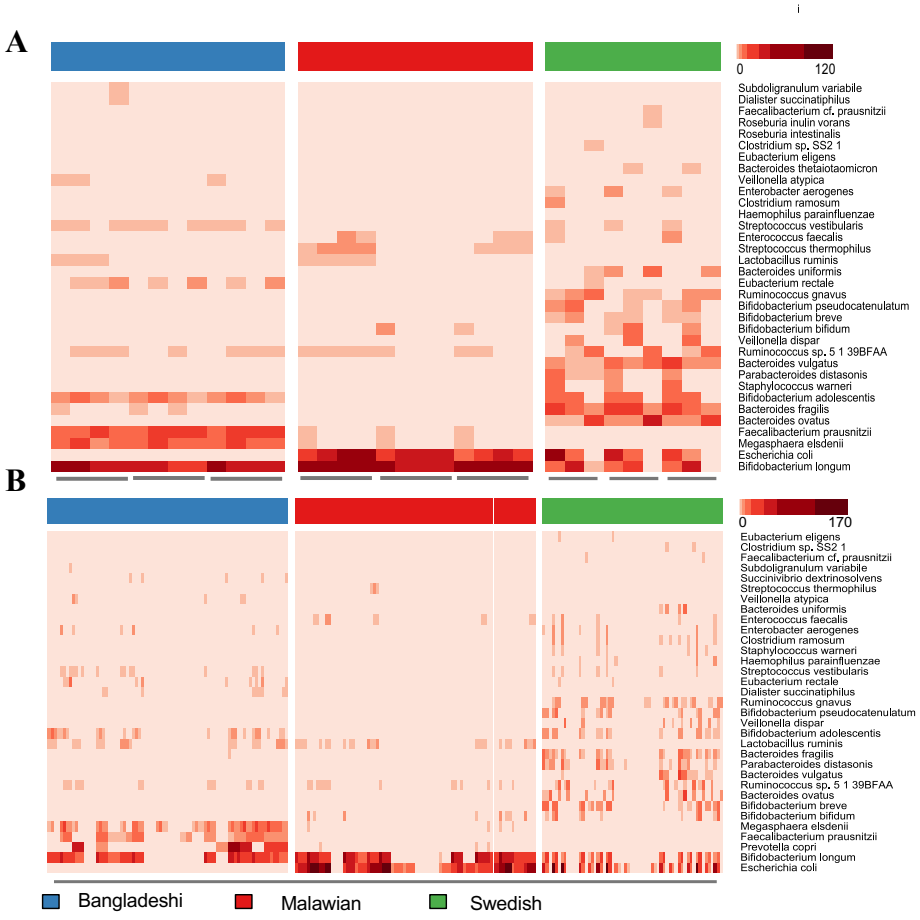


Figure 11. Prediction of growth rates using GEMs on three different media. A) Evaluation of the predicted growth rates for six gut bacterial strains; *Bacteroides thetaiotaomicron* (M6), *Bifidobacterium adolescentis* (M9), *Eubacterium rectale* (M29), *Faecalibacterium prausnitzii* (M31), *Prevotella copri* (M41), and *Roseburia inulinivorans* (M43). B) Prediction of the growth rate on human breast milk and RUTF.

As discussed in section 2.5.1, the gut microbiota provides the host with valuable metabolites such as SCFAs and AAs. Therefore, we performed a community- and diet-specific analysis of the potential of each microbial community for the production of the mentioned metabolites. For three SCFAs, namely acetate, propionate and butyrate, and all of the 20 amino acids, we calculated the Estimated Maximal Metabolic Potential (EMPP) to produce each of these metabolites by multiplying the production fluxes into the relative abundance of the relevant species:

$$EMPP = A_i \times v_j$$

A_i denotes the relative abundance of species i and v_j is the production flux of metabolite j. Figure 12 shows the EMPPs. According to the results of this section, the Swedish microbial community has a higher potential of production for the tested metabolites. These results applied for both healthy and diseased populations of the other two countries. In regards to the difference between healthy and malnourished children, there was no significant difference between the two groups in Malawian cohort; however, Bangladeshi children had a different profile in each state. Figure 12C shows the significant difference in terms of EMPP between the healthy children and the malnourished ones before, during and even after food therapy. Further analysis of plasma metabolomics related to these children confirmed these findings since several essential, and semi-essential amino acids were significantly lower in malnourished Bangladeshi children (Figure 12D). These results show how geographical, nutritional, and health status could influence the metabolic potential of gut microbiota.



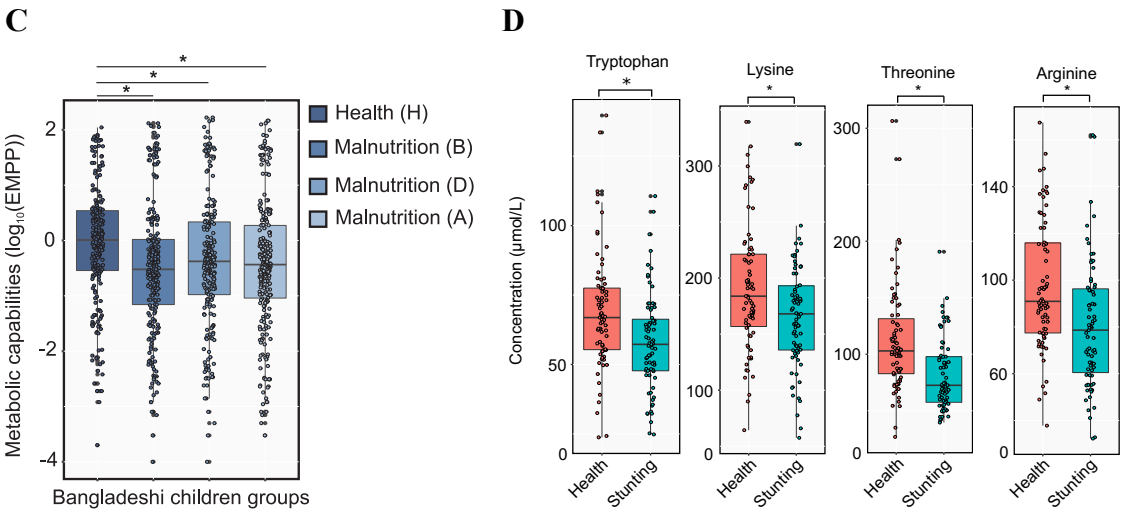


Figure 12. Estimated maximal metabolic potential. A) Short-chain fatty acids. B) Amino acids. C) Differences in metabolic potential in Bangladeshi children between health and malnourishment and before (B), during (D), and after (A) food intervention. D) The concentration of three essential amino acids (tryptophan, lysine, and threonine) and one conditionally essential (arginine) in the blood plasma of healthy and stunted children of Bangladesh.

We also investigated the biological interactions and metabolic potential of the aforementioned microbial communities in pairwise growth simulations. For each pair of the models, a linear programming problem was solved first:

$$Maximize \quad v_{biomass1} + v_{biomass1}$$

$$Subject\ to:$$

$$S.v = 0$$

$$lb \leq v \leq ub$$

$$v_{sharedMet_i1} + v_{sharedMet_i2} \leq [medium_i]$$

$$v_{consumption_i} \leq v_{production_i} + [medium_i]$$

In this LP, the sum of biomass reactions is optimised while each model is under its mass-balance and thermodynamic constraints. There are two additional constraints regarding metabolites in the shared environment; first, each shared metabolite from the medium is constrained in a way that the sum of consumption fluxes would not exceed the available amount. If there are any metabolites that are produced by one member and consumed by the other one in the pair, the last constraint implies that for the mentioned linking metabolite the consumption flux should be less or equal to the production flux plus its original available amount in the shared environment. By solving this LP problem and comparing the growth rates to the ones in the single strain optimisations, one of the six scenarios summarized in Table 1 would occur (Heinken and Thiele 2015).

Table 1. Biological interactions in pairwise growth simulations.

	Member 1	Member 2
Mutualism	Gain	Gain
Commensalism	Gain	No effect
Parasitism	Gain	Loss
Amensalism	Loss	No effect
Competition	Loss	Loss
Neutralism	No effect	No effect

Results of pairwise growth simulations show that competition, parasitism, and amensalism are the most frequent outcomes. In all of these scenarios at least one of the members has a decreased growth rate compared to the single strain optimisation state, which highlights that these microbes are more competitive than cooperative. Mutualism was only detected in RUTF which is a richer medium in terms of the number of the metabolites, so lesser degree of competition would occur as microbes can use different carbon and nitrogen sources. These results are in line with ecological theories stating that competition is a stabilizing factor for a community compared to mutualistic relationships (Coyte, Schluter, and Foster 2015).

In paper V, we analysed shotgun metagenomics data of individuals from four different countries, namely China, USA, Denmark, and Spain, in varying health conditions with a focus on the vitamin biosynthesis and transporter genes abundance. As explained in section 2.5.1.3, the human gut commensals are capable of producing vitamin K and B group. Comparing the normalized gene abundances across the healthy populations, we observed that Chinese individuals had the highest gene abundances for all vitamin biosynthesis and transport categories, followed by the American and European samples. Regarding the vitamin-related gene abundances in various health status, our analysis revealed a significant difference in Chinese cohort between diabetic and healthy population for all vitamin-related genes except for biotin and riboflavin as shown in Figure 13A, however, no significant difference was detected for other countries between healthy and IBD or obesity in American or Spanish populations, respectively, which is consistent with a previous study (Li et al. 2014).

Vitamin biosynthesis pathways have been observed to be typically incomplete in microorganisms, suggesting the need for complementary cooperation for *de novo* biosynthesis of these macromolecules (LeBlanc et al. 2013; Magnúsdóttir et al. 2015). Therefore, it is reasonable to see the vitamin-related genes spread across all phyla in the gut microbiota. Retrieving the lineage information from UniProt database (UniProt Consortium 2018) for the mapped genes to vitamin metabolism confirmed the previous observations; as shown in Figure 13B.

A **B**

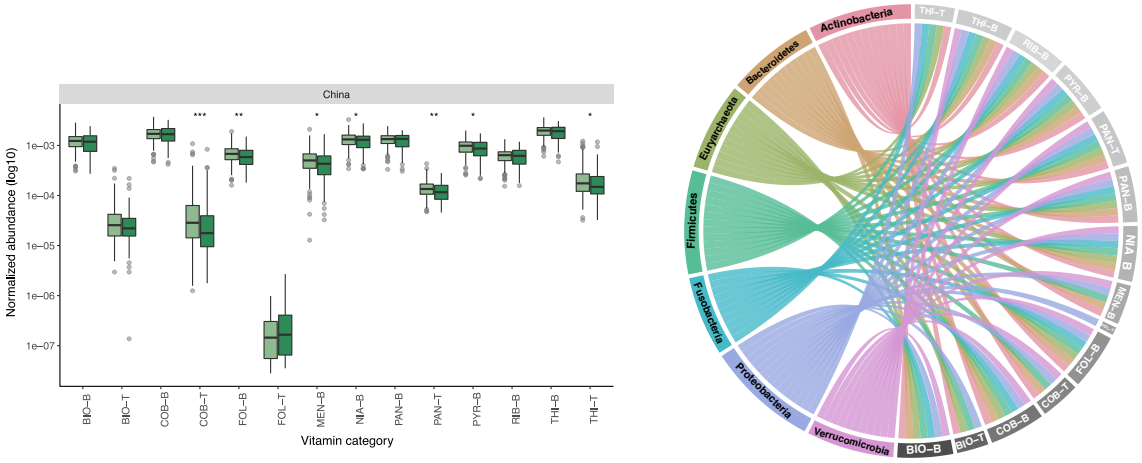


Figure 13. Abundance and phylogeny of vitamin-related genes of B and K group. A) Abundance of the genes related to vitamin metabolism across healthy and diabetic Chinese population**.** Light and dark green box plots represent healthy and diabetic individuals, respectively. B) Phylogenetic information of vitamin-associated genes. For each vitamin, there are two categories of biosynthesis (B) and transport (T). The analysed vitamins include biotin (BIO), cobalamin (COB), folate (FOL), menaquinone (MEN), niacin (NIA), pantothenate (PAN), pyridoxine (PYR), riboflavin (RIB), thiamine (THI).

Next, using GEMs from paper I, we investigated the presence and absence of the vitamin-related reactions in each of the cohorts. By this means, reactions in the vitamin-associated KEGG Orthology groups (KOs) were fetched from the KEGG database (Ogata et al. 1999) and compiled into a reaction catalogue. Then, as depicted in Figure 14, the presence and absence of each of the reactions were tested to elucidate the completeness and conserveness of the pathways of each vitamin group. As shown in Figure 14A, riboflavin and folate synthesis

pathways have the highest coverage which is consistent with the predicted capability of many gut symbionts to synthesize these two vitamins (Magnúsdóttir et al. 2015). Menaquinone, on the other hand, lies at the other end of the spectrum with the lowest coverage. This might be due to the fact that vitamin K requirements are provided by dietary intake and do not rely on bacterially produced ones in the gut since germ-free animals can still get enough menaquinone (LeBlanc et al. 2013). Other vitamins show an incomplete pathway with varying degrees which might suggest a co-dependence between two or multiple bacteria for their biosynthesis.

Furthermore, we analysed the American and Chinese healthy and diseased populations in terms of the abundance of reactions related to vitamin metabolism. Figure 14B and 14C shows American and Chinese populations, respectively. As shown here, the individuals from the two countries have a different gut bacterial make-up, and there are subtle differences between healthy and diseased compositions in both communities, however, it is not clear whether these structural differences would lead to functional varieties.

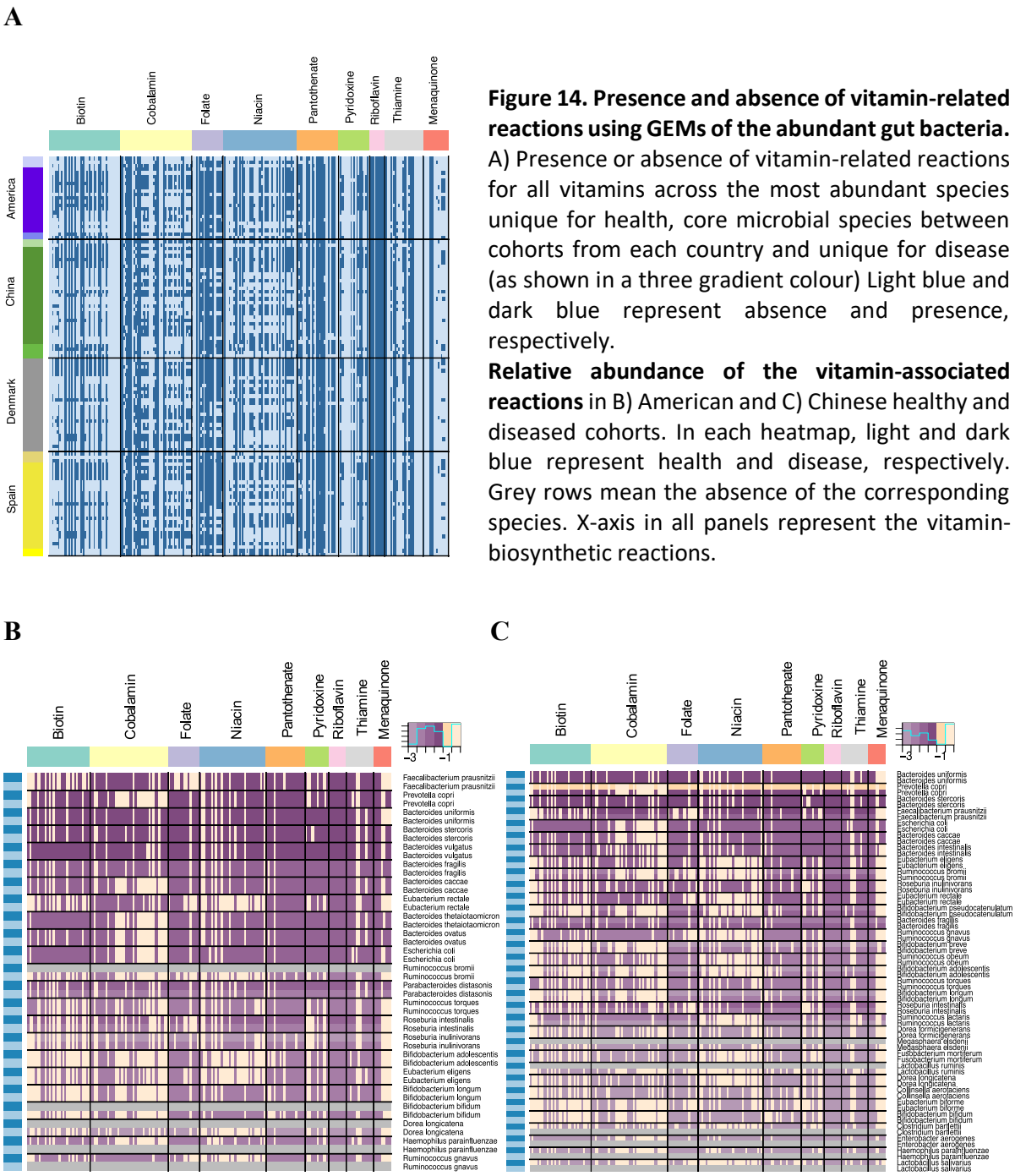


Figure 14. Presence and absence of vitamin-related reactions using GEMs of the abundant gut bacteria. A) Presence or absence of vitamin-related reactions for all vitamins across the most abundant species unique for health, core microbial species between cohorts from each country and unique for disease (as shown in a three gradient colour) Light blue and dark blue represent absence and presence, respectively.

Relative abundance of the vitamin-associated reactions in B) American and C) Chinese healthy and diseased cohorts. In each heatmap, light and dark blue represent health and disease, respectively. Grey rows mean the absence of the corresponding species. X-axis in all panels represent the vitamin-biosynthetic reactions.

3.2 PAPER II: Quality control of genome-scale metabolic models

GEMs are potent tools to simulate various metabolic scenarios and have the potential to be combined with -omics data to improve the accuracy of their predictions. It is essential for a GEM, therefore, to have a high quality in terms of specific criteria. In its simplest form, a bacterial GEM should be able to predict the growth rate as close as possible to the experimentally measured ones. High connectivity of metabolites and flux consistency also reflect a GEM's high quality. Furthermore, GEMs should be curated to eliminate any probable mathematical artefacts, such as the formation of thermodynamically infeasible cycles due to incorrect directionality, which would consequently influence their predictive power.

As mentioned in section 2.3.2 there are numerous methods for GEM reconstruction; here, we have analysed the quality of GEMs reconstructed following a semi-automated pipeline consisted of gut bacterial models (AGORA) (Magnúsdóttir et al. 2017) and compared them to a repository of manually-curated models (BiGG) (King et al. 2015) from several aspects and concluded that manual curation is a crucial step in GEM reconstruction to ensure that the models are capable of biologically meaningful predictions.

Firstly, we optimised the models in each group for growth with the original bounds on their exchange reactions without imposing further constraints. The predicted growth rates by the AGORA models, depicted in Figure 15A, have a wide range; from 0.004 h^{-1} to 255 h^{-1} which is clearly out of the biologically feasible spectrum. We noticed that meaningful growth rates are related to the models that have very tight constraints on their exchange reactions and the other group are not constrained on their respective exchange bounds. Although limiting the exchange reactions using experimental data is a conventional approach in constrained-based modelling, it seems that these bounds in AGORA models are instead controlling the value of the predicted growth rate which leaves the internal reactions not influential in the determination of the growth rate. The controlling role of the exchange reactions and the corresponding insignificance of the internal reactions are reflected in their respective reduced costs. The value of reduced costs shows the sensitivity of the objective function, growth rate in this case, to each variable which are the reactions in the model. For instance, in a random AGORA model, the active internal reactions had a median of 4.2877×10^{-19} for their associated reduced costs which is not comparable to the same variable for the active exchange reactions which was -1.7227×10^{-4}. The same values in the BiGG models are -9.1570×10^{-6} and –0.0062 for internal and exchange reactions, respectively, which shows an influential role both for exchange and internal reactions.

The high number of active exchange reactions encouraged us to take a closer look at the overall input flux of the AGORA models. Considering growth yields, defined as the growth rate divided by the total carbon source influx, we noticed that these models have very low growth yields, much lower than that of experimentally observed ones for bacteria (Figure 15B). This observation indicates that many carbon sources are taken from the environment and not used for biomass optimisation. To further investigate the fate of these compounds, we performed Flux Coupling Analysis (FCA) using F2C2 method (Larhlimi et al. 2012) in 100 randomly selected AGORA models to see which reactions are coupled with these carbon source uptake reactions. As mentioned in section 2.3.4.4., flux coupling analysis identifies reaction pairs in which non-zero flux through one of them would imply non-zero flux through the other. We observed that these reactions are coupled with numerous internal reactions activating, therefore, many of them. Among these carbon sources, we identified the essential ones and checked the gene association of their respective transporters and noticed that the majority of them (70% on average) were not gene-associated. Therefore, it highlights the fact that these non-gene associated carbon source transporters which in turn activate many internal reactions,

might have been added to the model during the gap-filling process without biological support. Such couplings and the low number of gene-associated transporters were not detected in BiGG models.

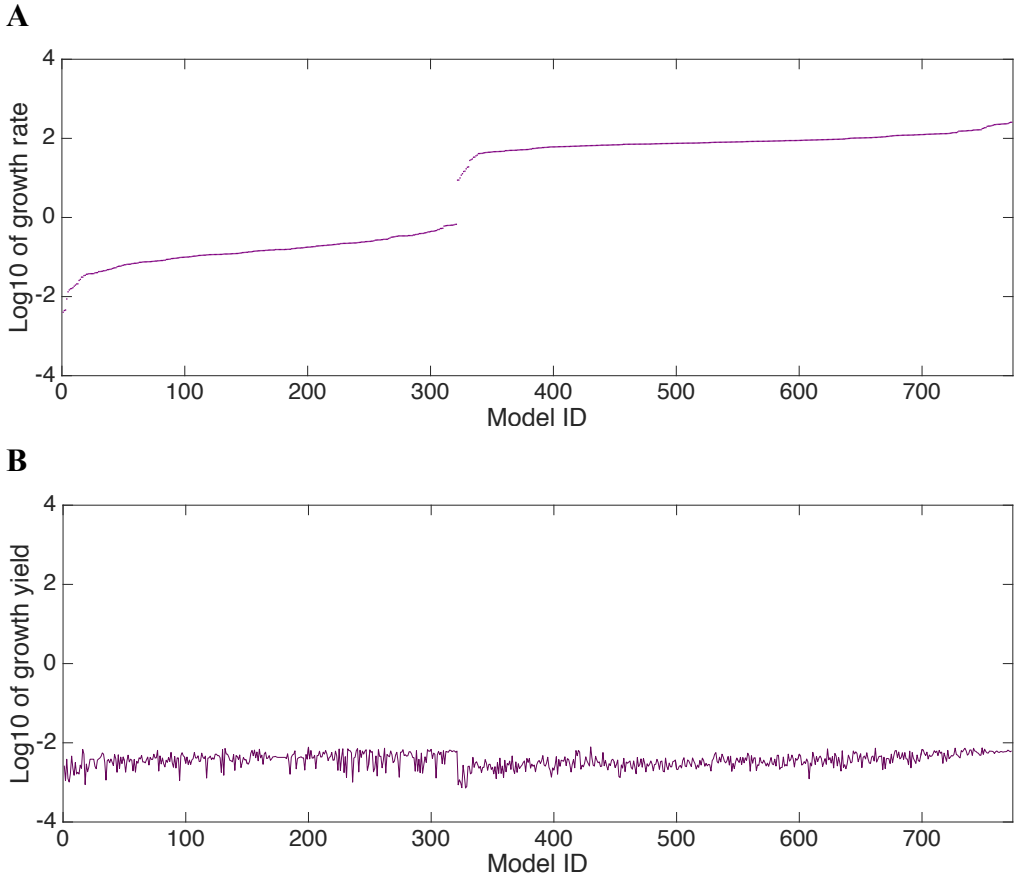
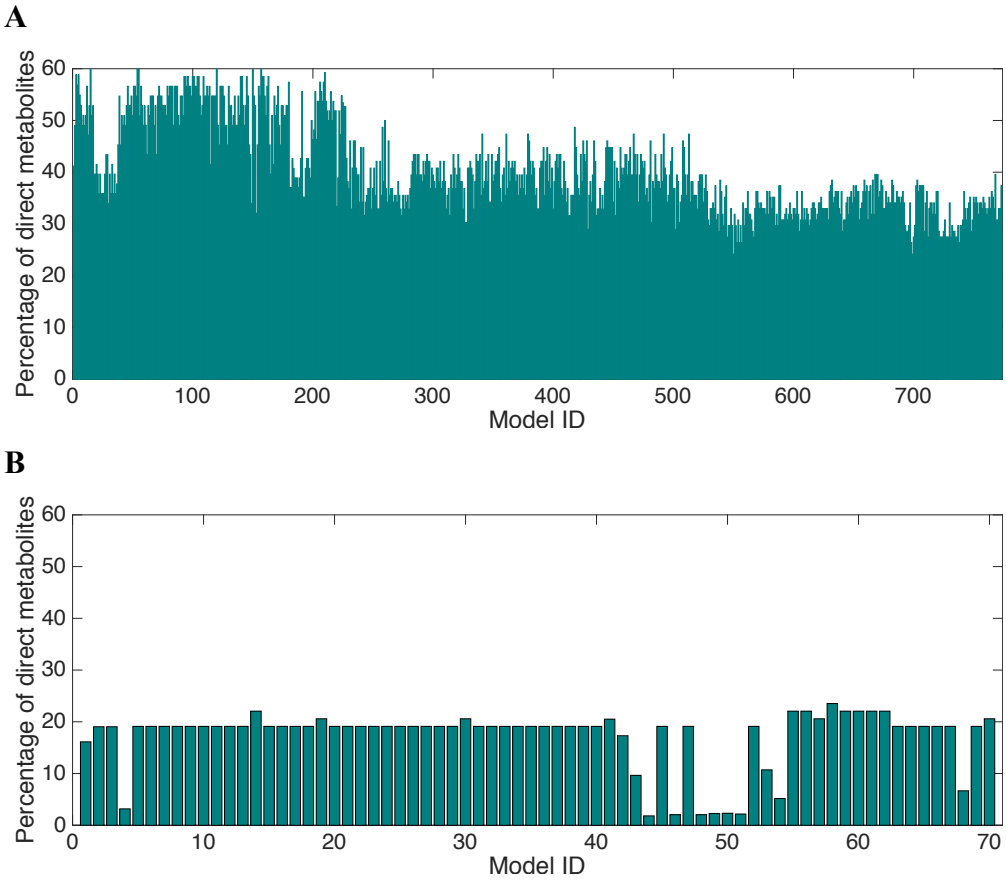


Figure 15. Growth rates and growth yields of the AGORA models. A) Anaerobic growth rates and B) Anaerobic growth yields of of the AGORA models using their original bounds in base 10 logarithmic scale.

Another characteristic of automatically generated models is the high percentage of direct metabolites in their biomass formulations. Direct metabolites are biomass precursors which are directly taken from the environment and consumed by the biomass reaction without any further biotransformations. It is another reason behind the high sensitivity of the objective function to the flux of exchange reactions since a majority part of the biomass precursors are provided directly by the exchange fluxes, and fixed constraints are therefore the only way to keep the value of the objective function in a biologically feasible range. Analysis of the biomass metabolites in AGORA and their associated reactions showed that the biomass reaction gets at least 30% of its precursors directly from the medium, which might be a consequence of an automatic gap-filling approach since as explained in section 2.3.3, MILP-based gap-filling algorithms seek to minimise the number of required reactions to render a draft model functional. Therefore, as the shortest pathway for any metabolite would be its direct consumption by a single exchange reaction and its associated transport reaction, these reactions are preferred over alternative solutions. Even though in MILP-based gap-filling methods it is possible to penalize exchange reactions in order not to end up with many of them, still these methods add exchange and transport reactions more than they should.

As shown in Figure 16A and 16B, the automatically reconstructed models have considerably higher direct metabolites compared to the manually-curated ones. Additionally, since all of the semi-automatically reconstructed models have gone through the same gap-filling approach, they are quite similar in their direct metabolites as depicted in Figure 16C showing Jaccard similarity index between model pairs regarding their direct biomass precursors.



C

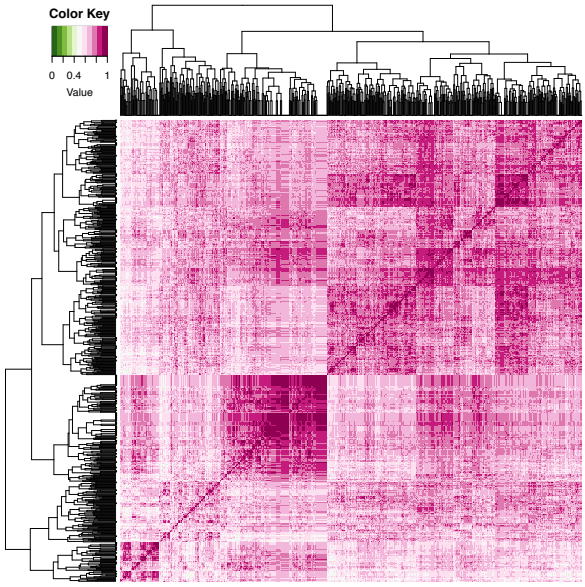


Figure 16. Direct metabolites on AGORA and BiGG models. A) Percentage of the direct biomass precursors in AGORA models. B) Percentage of the direct biomass precursors in BiGG models. C) Heatmap showing the Jaccard similarity index regarding the direct metabolites in each pair of AGORA models.

A high-quality GEM has high connectivity and is flux consistent. Therefore, it is expected for any reaction in a metabolic network to be able to carry a non-zero flux if all of the exchange reactions are unbounded; otherwise, they are blocked. Blocked reactions should be further curated to see whether they rise from annotation errors, therefore wrongly added to the network, by looking for further evidence in the scientific literature and experimental observations. If a reaction is unable to carry a non-zero flux, then it might be associated with a dead-end metabolite (metabolites that can only be consumed or produced). Any reaction associated with the dead-end metabolites would be blocked due to the mass balance constraints. Some blocked reactions, though, are not linked to any dead-end metabolites and still cannot carry flux if, for example, they are related to a metabolite which is involved in a thermodynamically infeasible loop . To sum up, dead-end metabolites and thermodynamically infeasible loops might inhibit metabolic reactions from carrying a non-zero flux and therefore, manual curation for them is necessary.

In order to identify blocked reactions in two model categories, we removed any bounds on the exchange reactions and performed Flux Variability Analysis (FVA) to detect the reactions whose minimum and maximum is zero. Figure 17A and 17B shows the percentage of the blocked reactions in AGORA and BiGG models, respectively. As shown in Figure 17A and 17B, except for a few BiGG models, the manually curated ones have a lower percentage compared to the automatically generated models. The same observation is valid for the percentage of the dead-end metabolites as the BiGG models have fewer of such metabolites compared to the AGORA models. Therefore, further manual curation is required for the AGORA models to increase network connectivity and flux consistency of these models.

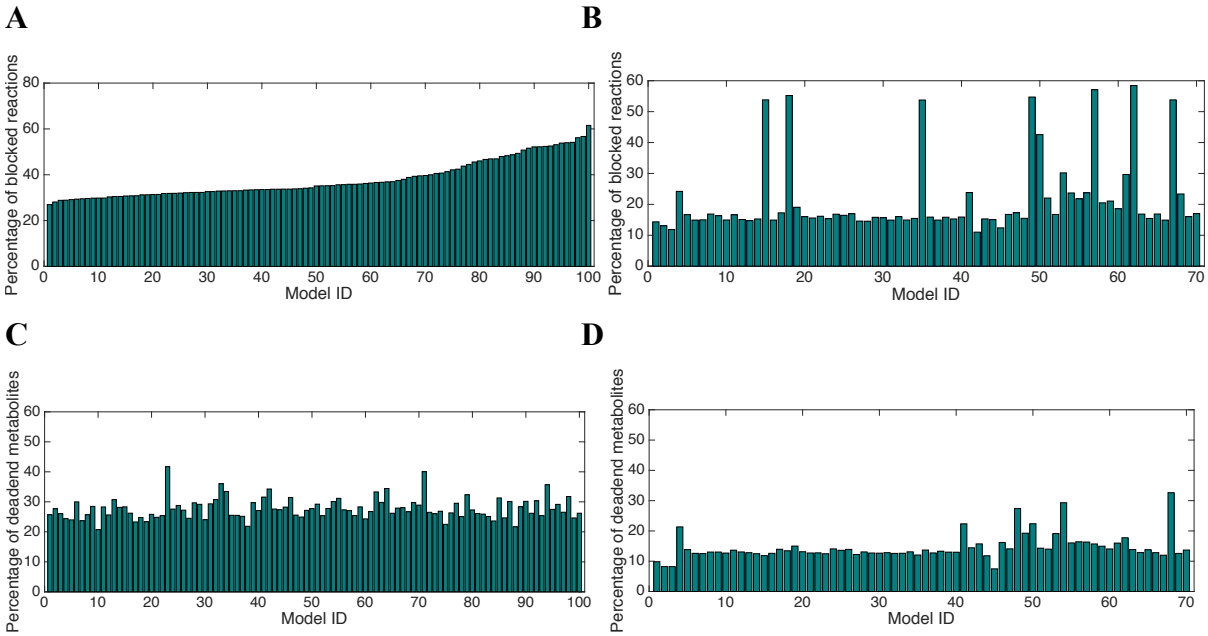


Figure 17. Percentage of blocked reactions and dead-end metabolites in AGORA and BiGG models. A) Percentage of the blocked reactions in AGORA models. B) Percentage of the blocked reactions in BiGG models. C) Percentage of the dead-end metabolites in AGORA models. D) Percentage of the dead-end metabolites in BiGG models.

As explained in the previous section, thermodynamically infeasible loops might disable some reactions from carrying a non-zero flux. Additionally, they violate the second law of thermodynamics. Similar to Kirchhoff's second law stating that the sum of potential differences around a closed circle is zero, the second law of thermodynamics implies that the sum of the chemical potential difference in a loop consisted of chemical reactions is zero. Therefore, since there are no net thermodynamic forces, the net flux in a loop must be zero (Price, Thiele, and Palsson 2006). We checked to see firstly if the models in the two aforementioned repositories contained any thermodynamically infeasible loops, and secondly if any of the loop-involved reactions were essential for growth. By this means, we blocked all of the exchange reactions and performed FVA to identify the reactions whose flux could reach the lower or the upper bound of the reaction. Figure 18 depicts the number of loop-related reactions in a 100 randomly chosen AGORA models, however, in case of the BiGG models we only detected one model containing loop reactions. Furthermore, all of the loop-related reactions in AGORA repository were essential for growth. These observations further pinpoint the need for manual curation for automatically generated models.

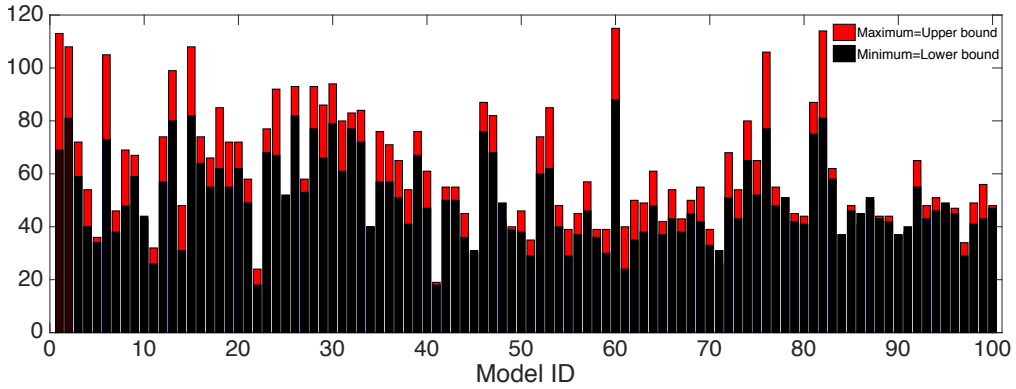
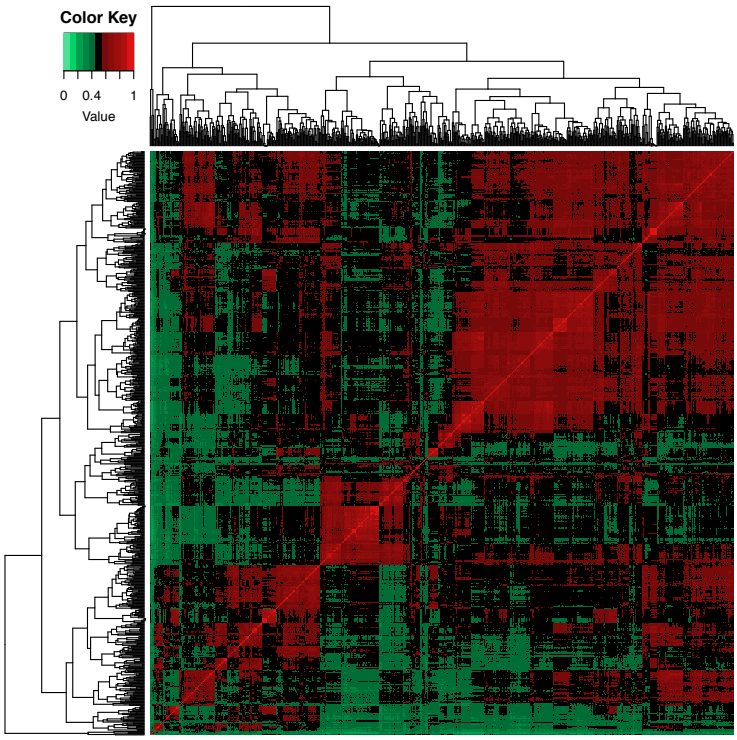


Figure 18. The number of reactions in 100 randomly chosen AGORA models which are associated with a thermodynamically infeasible loop.

Gut bacterial species are functionally redundant. We investigated the degree of similarity between the metabolic models of these microorganisms, to see how similar they are and how they cluster based on this comparison. To measure metabolic similarity, we calculated Jaccard similarity index based on all metabolic reactions, active metabolic reactions identified by FBA and pFBA, and also the biomass formulation between each possible model pair of the AGORA repository. Regarding all metabolic reactions, as depicted in Figure 19A, the majority of the AGORA models pairs show a high degree of similarity which highlights the functional redundancy in the human gut microbiota. However, the clustering pattern does not reflect their phylogenetic relationships as except for Bacteroidetes; all the other phyla are distributed across three groups. Considering the similarity between the internal active reactions, we used both FBA and pFBA to identify them. pFBA was necessary to ensure that loop-related reactions do not affect the results of this section. Based on the calculated Jaccard similarity indexes related to the active reactions, models cluster into three main groups (Figure 19C). All phyla were distributed between the groups as shown in Figure 19E. This pattern of clustering means that for each group, there is a highly conserved active core that drives the biotransformation of the model inputs and this core is not originating from biological closeness as not a single phylum is present in only one group. The main reason for this clustering is the similarity of the biomass formulation in each group (Figure 19D). The clustering pattern based on biomass similarity Jaccard indexes were exactly the same as that of active reactions. We concluded, therefore, that the biomass similarity as the objective function drives the similarity of the active reaction core. In other words, an automatically generated biomass formulation in a draft model, followed by the same gap-filling approach has led to an artificial similarity between the AGORA models. Even though culturing the gut commensals *in vitro* and measurement of their biomass composition still remains a challenge, there has been some efforts in that direction and only with more data the discussed shortcomings of the automatically generated models can be tackled.

A



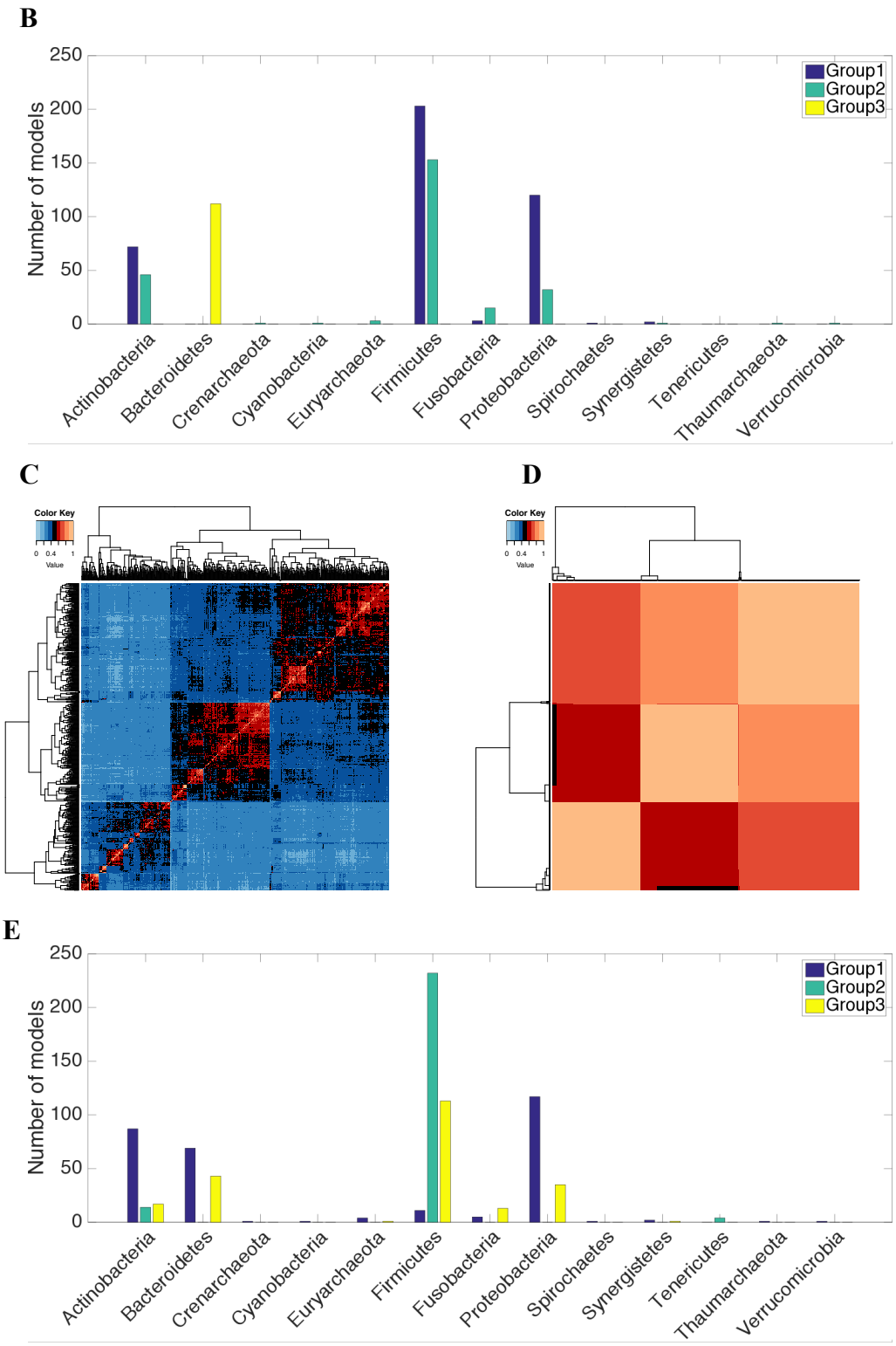


Figure 19. Similarity analysis for AGORA models. A) Clustergram based on Jaccard similarity index between each model pair calculated considering all reactions. B) Phyla distribution between the three main groups with the highest similarity regarding the metabolic reactions. C) Clustergram based on Jaccard similarity index between each model pair calculated considering the active reactions identified by FBA and pFBA. D) Clustergram based on Jaccard similarity index between each model pair calculated considering the biomass precursor metabolites. E) Phyla distribution between the three main groups with the highest similarity regarding the active reaction core and biomass precursors.

3.3 PAPER III: Lactic acid bacteria GEMs and simulations in fermented milk product

Fermentation has long been used in food production and preservation. Biotransformation of raw material is usually performed by multi-species microbial consortia with complex biological dynamics and metabolic interactions. It is not, therefore, straightforward to manipulate most of these food-related microbial communities towards a specific goal; such as optimisation of production of metabolites contributing to the fermented product's taste and aroma. Furthermore, some bacteria involved in food have probiotic capabilities. Studying these microbial from various aspects, ranging from the flavour of the final product to the potential of it to deliver health benefits, requires novel computational tools. GEMs of the relevant microbial species can be employed to both explore the biochemical machinery leading to food fermentation and also help to postulate metabolic strategies for the particular design of fermented food products.

In paper III, we have reconstructed genome-scale metabolic models for four lactic acid bacteria involved in milk fermentation and dairy production along with one Bifidobacterium strain with probiotic properties. We have curated and validated these five models both in single strain growth, bi-culture in yogurt and a community of five bacteria in a more complex fermented milk product. Finally, we have used the validated models to come up with an optimal ratio between them to achieve the maximum production of acetaldehyde which is the main volatile in yogurt's taste and aroma.

The mentioned four lactic acid bacteria consisted of *Lactobacillus delbrueckii* subsp. *bulgaricus* (*L. bulgaricus*) CNCM I-1519 and CNCM I-1632, *Lactococcus lactis* subsp. *lactis (L. lactis*) CNCM I-1631 and *Streptococcus salivarius* subsp. *thermophilus* (*S. thermophilus*) CNCM I-1630. *Bifidobacterium animalis* subsp. *lactis* (*B. lactis*) CNCM I-2494 with probiotic capabilities was also modelled as it is used in a fermented milk product with the four LABs above. For draft network reconstruction, two complementary approaches were applied. Firstly, we used RAVEN toolbox (Agren et al. 2013) which retrieves all the relevant biochemical reactions coded by the genome based on protein homology from the KEGG database (Ogata et al. 1999). Then, we used an online annotation server from the same database, namely BlastKOALA (Kanehisa, Sato, and Morishima 2016), to identify and include more metabolic reactions into the draft networks. We manually reviewed and curated the reactions that were exclusively identified by the second approach before adding them to the draft networks. For biomass formulation, we used phylogenetically close organisms as templates as shown in Table 2. Each of these previously-modelled organisms had a biomass formulation which was wholly or partially, experimentally measured. Since all these strains are involved with milk fermentation, milk composition was adopted as their environment; meaning that exchange reactions for milk molecular components were added to the model to represent the medium.

Table 2. Template models for biomass formulation and the measured macromolecules and parameters.

Strain	Template model	Measured macromolecules/parameters
***B. lactis* CNCM I-2494**	*Bifidobacterium adolescentis* (El-Semman et al. 2014)	Estimated from 2 other studies (Neidhardt, Ingraham, and Schaechter 1990; Teusink et al. 2006)
***L. bulgaricus* CNCM I-1519**	*Lactobacillus plantarum* (Teusink et al. 2006)	DNA, RNA, Protein, Lipids, Polysaccharides, Peptidoglycans, Wall teichoic acids, Lipoteichoic acids
***L. bulgaricus* CNCM I-1632**	*Lactobacillus plantarum* (Teusink et al. 2006)	DNA, RNA, Protein, Lipids, Polysaccharides, Peptidoglycans, Wall teichoic acids, Lipoteichoic acids
***L. lactis* CNCM I-1631**	*Lactococcus lactis* subsp. *cremoris* MG1363 (Flahaut et al. 2013)	Protein, GAM and NGAM calculated from FBA and compared to other LABs

Draft models are usually non-functional; meaning that one or several biomass precursors cannot be synthesized due to the incompleteness of the network. Therefore, a gap-filling process is required to detect metabolic holes and fill them with candidate reactions. In this paper, we used an LP-based gap-filling approach which works based on flux consistency of the draft network rather than a MILP-based one since the scalability issue of such methods hampers their usability for larger genome-scale models. The LP-based approach, fastGapFill, instead solves a series of L_1-norm regularized LPs (Thiele, Vlassis, and Fleming 2014). A universal database (KEGG in this case), is added to each compartment of the draft network. Then a compact flux consistent network is calculated and greedily extended to include the original draft network and a near-minimal set of newly added reactions. Exchange and transport reactions were penalized over the metabolic reactions by a factor of ten. All reaction suggested by fastGapFill were manually inspected before adding to the model. No exchange or transport reactions were added in this stage to the models since the metabolic reactions were sufficient to make the draft models functional.

Reversibility information of biochemical reactions are normally not included in databases like KEGG, and the majority of the reactions are assumed to be reversible which may not be the case in physiological conditions. Leaving all the reactions unconstrained would lead to the formation of internal loops such as futile cycles and energy-generating cycles. While the former is not a mathematical artefact and occurs in microorganisms living in energy-rich environments (Reidy and Weber 2002; Russell 2007), the latter combined with an energy-generating reaction (EGR) forms an energy generating cycle (EGC) which leads to erroneous energy generation without any inputs(Figure 20). It has been shown in a study that energy production via a combination of internal loops coupled with energy generating reactions could lead to a 25% increase in the value of the objective function (Fritzemeier et al. 2017). In order to detect any potential EGCs in the models, we added 15 energy dissipating reactions (EDR) for 15 energy transmitting cellular metabolites, namely ATP, CTP, GTP, UTP, ITP, NADH, NADPH, $FADH_2$, $FMNH_2$, ubiquinol, menaquinol, 2-demethylmenaquinol, Acetyl-CoA, L-glutamate, and proton. If there is any EGCs, producing any of the mentioned metabolites, by adding an EDR a closed cycle would form with its respective EGR (Figure 20) and the newly added reaction would be active. Therefore, a non-zero flux through any of the EDRs would pinpoint

an EGC in the model. By this method, we found that ATP was available in excess and all ATP-related reactions were curated manually to eliminate any non-biological fluxes.

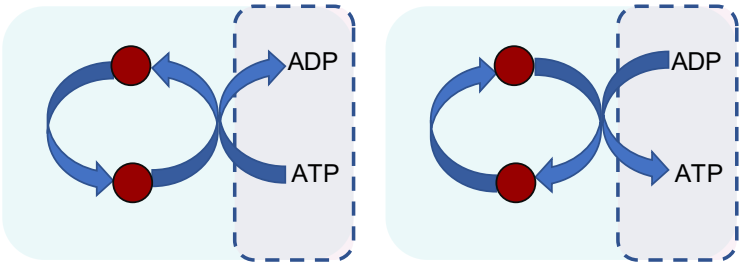


Figure 20. A) Futile cycle which dissipates energy and B) energy generating cycle which produces energy by violating the thermodynamics second law.

All of the modelled organisms were grown on two different media and their growth dynamics and metabolic activities were monitored. The used media consisted of a milk-based and a synthetic complex medium. In each case, bacterial cells enumerations and measurements of concentrations of certain metabolites were performed. Therefore, for each model, exo-metabolomics data and change in cfu (colony-forming unit) or OD (optical density in 600 nanometres of wavelength) were available. The following key metabolites were measured by HPLC during the cultivations: acetic acid, citric acid, fructose, fumaric acid, galactose, glucose, lactic acid, lactose, maltose, orotic acid, pyruvic acid, sucrose, and succinic acid. We used the measured concentrations to estimate consumption or production fluxes for each metabolite as shown below:

$$flux = \frac{\Delta\, concentration\ \left(\frac{g}{ml}\right) \times \frac{1}{molecular\ weight\ \left(\frac{g}{mmol}\right)}}{\Delta\, cfu\ \left(\frac{cell}{ml}\right) \times\ grDW\left(\frac{g}{cell}\right) \times\ \Delta t\ (hr)}$$

Next, fluxes associated with the consumed metabolites were applied as constraints on their respective exchange reactions and flux through the biomass reaction was optimised. The predicted growth rates and flux through exchange reactions of the produced metabolites were used to evaluate our *in silico* predictions. In the case of inconsistencies between the computational predictions and experimental observations, we curated the models to reduce the gap between the two. The predicted growth rates are shown in Figure 21. As depicted in this figure, the predicted growth rates agree well with the experimentally measured ones. For metabolic profiles, we used both FBA and FVA to predict the experimental and computational fluxes semi-quantitively. FAV was performed to validate that the experimentally estimated flux falls into the predicted range of achievable fluxes. As shown in Figure 21B, each organism has been grown on both milk-based and a synthetic medium. For each medium, the concentration changes of several key metabolites have been measured over time. In each case, only the metabolites which had a non-zero change in concentration (therefore a non-zero flux through their respective exchange reactions) are shown in this figure. For all of the strains, glucose and lactose serve as the primary carbon source in synthetic and milk-based media, respectively. Lactate was produced by all of the strains in both media, and with a higher flux in milk-based one. Acetate is also produced by some of the strains, especially *B. lactis* which is known to be heterofermentative.

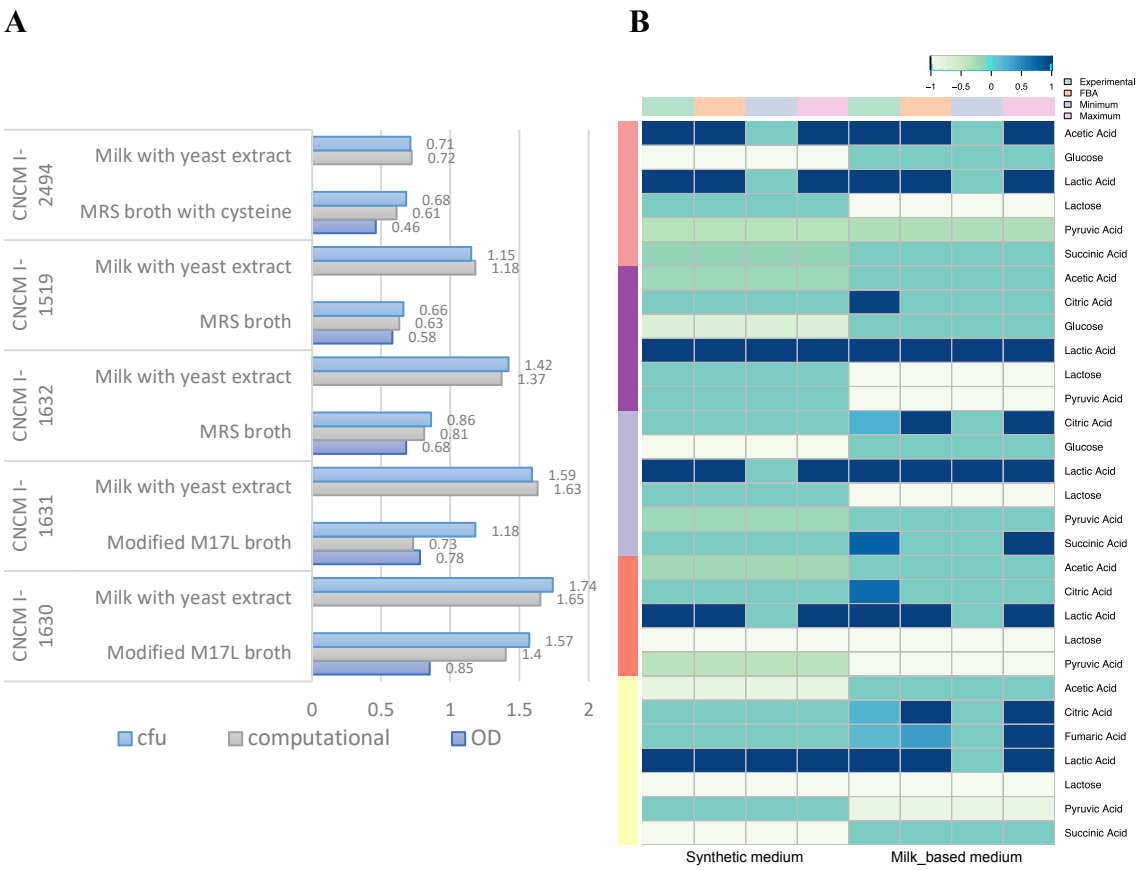


Figure 21. Experimental and computational growth rates and metabolic profiles for the modelled bacteria. A) Growth rates as measured by OD and cfu changes and predicted by the models on two different media. B) metabolic profile of the consumed and produced metabolites by each bacterial strain in the same order as in A.

Yogurt is produced via fermentation of milk by two strains of lactic acid bacteria, namely *L. bulgaricus* and *S. thermophilus*. We used a pairwise growth approach as described in pairwise simulations section in Paper I, to simulate their co-culture. Bi-culture of these two strains in milk was monitored experimentally in terms of changes in bacterial cell numbers, lactose and lactate concentrations. These measurements were used to calculate metabolic fluxes and subsequently, to evaluate modelling predictions. Table 3 shows the results of yogurt co-culture simulation. The calculated flux of lactose consumption was set as the upper bound of the available carbon source. Then, the weighted sum of the biomass reactions based on the experimental growth rates was optimised. Results of this section were in line with the experimental data. Additionally, modelling framework enabled us to calculate the contribution of each of the strains to biotransformation of lactose to lactate which is challenging to measure *in vitro*.

Table 3. Results of yogurt culture simulations.

	Experimental (1/h)	Computational(1/h)	Lactose consumption (mmol/gDW/h)	Lactate production (mmol/gDW/h)
L. bulgaricus	0.97	0.68	10	17.53
S. thermophilus	1.29	1.14	9.04	13.37
Lactate production (mmol/g_{DW}/h)	35.80	30.90	-	-

The yogurt-associated strains contribute to the flavour and aroma of the final product. In addition to lactate production, which gives yogurt its acidic taste, these strains also produce a range of volatiles such as acetaldehyde, diacetyl, acetoin and many more. Among all these volatiles, acetaldehyde is the key contributor to yogurt's unique aroma (Han et al. 2007). There have also been efforts towards maximisation of its production utilizing metabolic engineering tools (Bongers, Hoefnagel, and Kleerebezem 2005; Chaves et al. 2002). Here, we employed our models and the same pairwise growth simulation framework to hypothesize the optimum ratio between the two yogurt strain which would lead to maximum acetaldehyde production. By this means, we first fixed each strain's biomass to its maximum in the pairwise growth mode and changed the objective function to the sum of acetaldehyde production reactions. Next, we tested a large number of random coefficients for the new objective function and evaluated the outcomes; the flux of acetaldehyde production. Results related to ratios capable of producing high amounts of acetaldehyde are shown in Figure 22. As shown below, higher production occurs when the ratio of *L. delbruckeii* to *S. thermophilus* decreases. This observation reflects the actual ratio that is used for yogurt production (Hill et al. 2017), and therefore shows how a computational approach can help optimise microbial communities used in the food industry towards a specific goal.

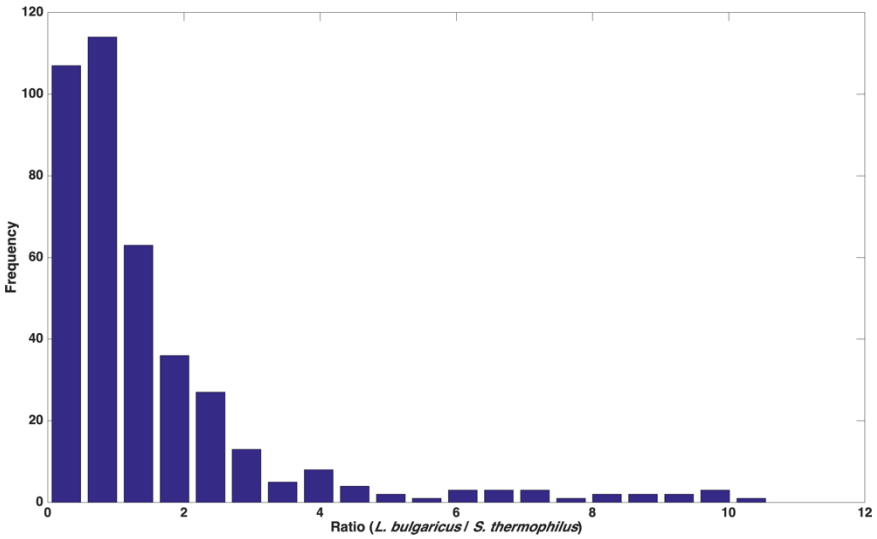


Figure 22. Ratios between the bacterial strains involved in yogurt fermentations leading to high production of acetaldehyde.

We also simulated a more complex microbial community consisted of all the five strains which are used in a fermented milk product containing a probiotic strain. In order to simulate their co-culture, we used a compartmentalization approach in which each strain is regarded as a separate compartment similar to the eukaryotic models, as described in section 2.4. Milk composition was used as the *in silico* medium and similar to the yogurt simulations, total lactose uptake was calculated based on experimental measurements and applied as a constraint. Bacterial cell counts were also measured and used to calculate the growth rates of the strains. The experimentally measured and computationally predicted values for the growth rates, lactose consumption rate ,and lactate production rates are summarized in Table 4. Growth rates and lactate production rate are consistent with the *in vitro* values. Lactose consumption, however, is predicted lower than experimentally observed rate. Measurements of the cellular dry weight might change this value since we have used an estimated value for the cellular dry weights of these strains to calculate metabolic fluxes.

Table 4. growth rates calculated based on bacterial enumerations and the predicted values for each strain in complex symbiosis.

Feature	Strain	Experimental	Computational (1/h)
Growth rates (1/h)	**CNCM I-2494**	0.32	0.37
	CNCM I-1519	0.75	0.65
	CNCM I-1632	0.75	0.79
	CNCM I-1631	0.87	0.68
	CNCM I-1630	0.98	0.93
Metabolite consumption/production ($mmol/g_{DW}/h$)	**Lactose consumption**	20.78	15.35
	Lactate production	25.8	22.36

3.4 PAPER IV: Uniform randomised sampling of microbial communities feasible solution space

GEMs have been employed to simulate microbial communities using different approaches. Enzyme-soup method in which all of the metabolic reactions of a community are assembled into one single model, compartmentalization method in which each community member is regarded as a wall-protected organelle in a eukaryotic-like model, and several other methods rely on optimisation of a pre-assumed objective function, which is usually a combination of the biomass reactions of the community members. Optimisation of growth rates, however, might not be the biological objective of the members of a microbial community. On the other hand, the high number of the variables in a multi-species model hinders the practicality of unbiased methods such as random sampling and elementary flux modes analysis (Øyås and Stelling 2018).

Monte Carlo sampling approaches have been deployed to study small single cell networks such as the human red blood cell (Wiback et al. 2004), cardiac mitochondria (Thiele et al. 2005), and neuron/astrocyte metabolic models (Occhipinti et al. 2007) to thoroughly investigate the possible fluxes under defined physiological constraints. However, as the size of the network increases, this approach requires more time to cover a larger feasible space. Therefore, it is not suitable for a combination of genome-scale networks. It is possible, however, to simplify a microbial community model before applying the randomised sampling method on the feasible space. In paper IV, we have proposed simplification and subsequent randomised sampling of microbial communities. We have implemented the proposed framework on microbial communities comprised of the human gut microbial models of *Bacteroides thetaiotaomicron*, *Faecalibacterium prausnitzii*, *Eubacterium rectale*, *Bifidobacterium adolescentis*, and *Ruminococcus bromii* and have used experimental data to evaluate the results.

In this method, the only variables considered, are the community-level reactions; namely inputs and outputs of each community member and also potential linking reactions; representing a potential cross-feed between the community members. Internal reactions are ignored to reduce the number of variables and keep the resolution at the community level. The first step, therefore, is to determine these community-level reactions. After that, the community members are characterised and divided into two groups based on their growth requirements and the metabolites available in the shared environment: *primary* organisms which are able to grow independently relying solely on the available metabolites in the medium and *non-primary* organisms which require at least one cross-feeding metabolite.

Three conditions should be checked and met at this stage:

1. There should be at least one primary organism in the community.
2. All non-primary organisms should at least have one partner able to produce their growth requirements metabolites that are absent from the medium.
3. There should be at least one linking reaction in the community.

When the community topology is characterised, and all of the necessary conditions are satisfied, the activation of the community members starts in a cascade-like manner. Firstly, the primary organisms are activated, and their metabolic production profile is determined. Then the shared environment is updated in regards to the metabolites produced by the active primary organisms. After that, the community is scanned for the non-primary organisms that can grow with the updated composition of the medium. This iterative activation process continues until all community members are active. At the end of this stage, all community-level variables (consuming, producing and linking reactions) are identified in the context of the shared

environment and the specific community topology. These variables then provide the basis for community formulation in the next steps.

Next, the community matrix is formulated based on the identified variables and a set of equality and inequality constraints are formed. Since the internal reactions of the individual models are not considered, therefore the inputs and outputs are not connected via a stoichiometric matrix. In this approach, we used elemental balance constraints to link each organism's inputs and outputs. These constraints state that the number of carbon, nitrogen, sulphur, phosphorous and hydrogen atoms entering and leaving each system must be equal, or in other words, each model should be elementally balanced in regards to its inputs and outputs.

Inequality constraints are also used to formulate community LP. Firstly, the shared metabolites of the medium are constrained in a way that the sum of consumption fluxes does not exceed the available amount of the respective metabolite. For the linking reactions, the sum of consumption fluxes should be less or equal to the sum of production fluxes. Lower and upper bounds of the community-level reactions are equal to that of their original models. Finally, a hit-and-run sampling is performed on the feasible solution space of the modelled microbial community and the distribution of the samples, and their correlations are analysed.

Since this method is based on the metabolic consumption and production profile, prior to community modelling and sampling, we examined each model in terms of consumed and produced metabolites using experimental data for four of the models; *B. thetaiotaomicron*, *B. adolescentis F. prausnitzii*, *E. rectale*. These strains were grown anaerobically in YCFA medium, and changes in OD and concentration for glucose, succinate, acetate, lactate, formate, butyrate and propionate were measured. For simulations, we performed growth optimisation while minimising the sum of fluxes. Figure 23 shows the comparative results of this section. Initial inconsistencies were manually curated to close the gap between computational predictions and experimental observations. As shown below, in all cases computational and experimental results are consistent, and the models could be used for community simulations.

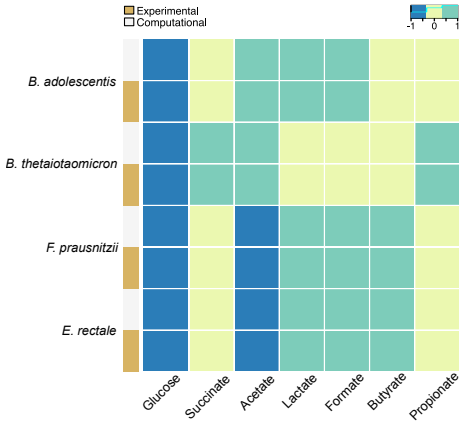
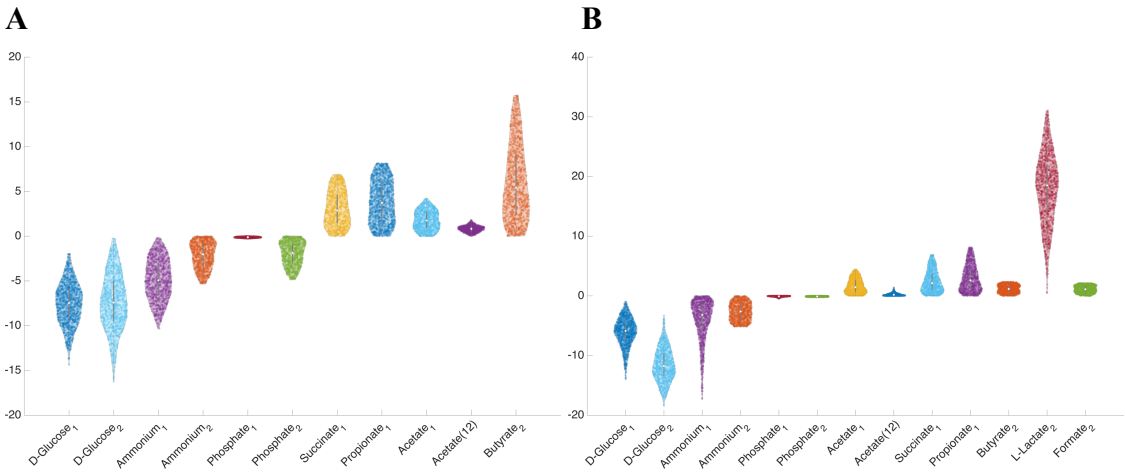


Figure 23. Computational prediction and experimental measurements of metabolic consumption and production profile of four human gut bacterial models. Negative and positive values represent consumption and production respectively. Zero elements represent metabolites without changes in their concentration.

Starting with a simple bi-culture, we modelled the pairwise growth of *B. thetaiotaomicron* and *E. rectale*. In this simple bacterial co-culture, glucose is the sole carbon source shared between the two strains. Acetate is produced by *B. thetaiotaomicron* and consumed by *E. rectale* and subsequently, transformed to butyrate. Figure 24A shows the community resource distribution and metabolite production. As shown below, samples related to the acetate linking reaction, namely Acetate (12), has a lower median compared to the acetate which is released to the micro-environment. Propionate and succinate are also produced by *B. thetaiotaomicron* which is consistent with experimental observations of this co-culture by Mahowald *et al* (Mahowald et al. 2009). It has also been reported in another study (Venturelli et al. 2018) that in this bi-culture, *B. thetaiotaomicron* is the dominant strain. Samples of the biomass reactions in Figure 24C reflect the higher capacity of *B. thetaiotaomicron* to grow within the community constraints. Finally, clustering patterns of the samples shown in Figure 24E highlights the dependency of *E. rectale* on the growth of the other member of the co-culture since the biomass reactions and the acetate linking reaction cluster closely.

We modelled another co-culture comprised of *B. thetaiotaomicron* and *F. prausnitzii* in which the former is an acetate-producer, and the latter is a butyrogenic gut symbiont (Moens, Weckx, and De Vuyst 2016). Distributions of the community-level flux samples are shown in Figure 24B. As depicted below in this bi-culture, it is noticeable that the potential for lactate production is more than that of other produced metabolites. Co-colonization of these two strains has been studied in rats and cfu counts of the cecal content showed the dominance of *B. thetaiotaomicron* over *F. prausnitzii*. As shown in Figure 24D, we can see that *B. thetaiotaomicron* can reach considerably higher growth rates compared to *F. prausnitzii*. Furthermore, the community-level flux samples cluster similar to that of the previous co-culture model where the biomass reactions and the linking acetate reaction cluster closely together highlighting the importance of cross-feeding.



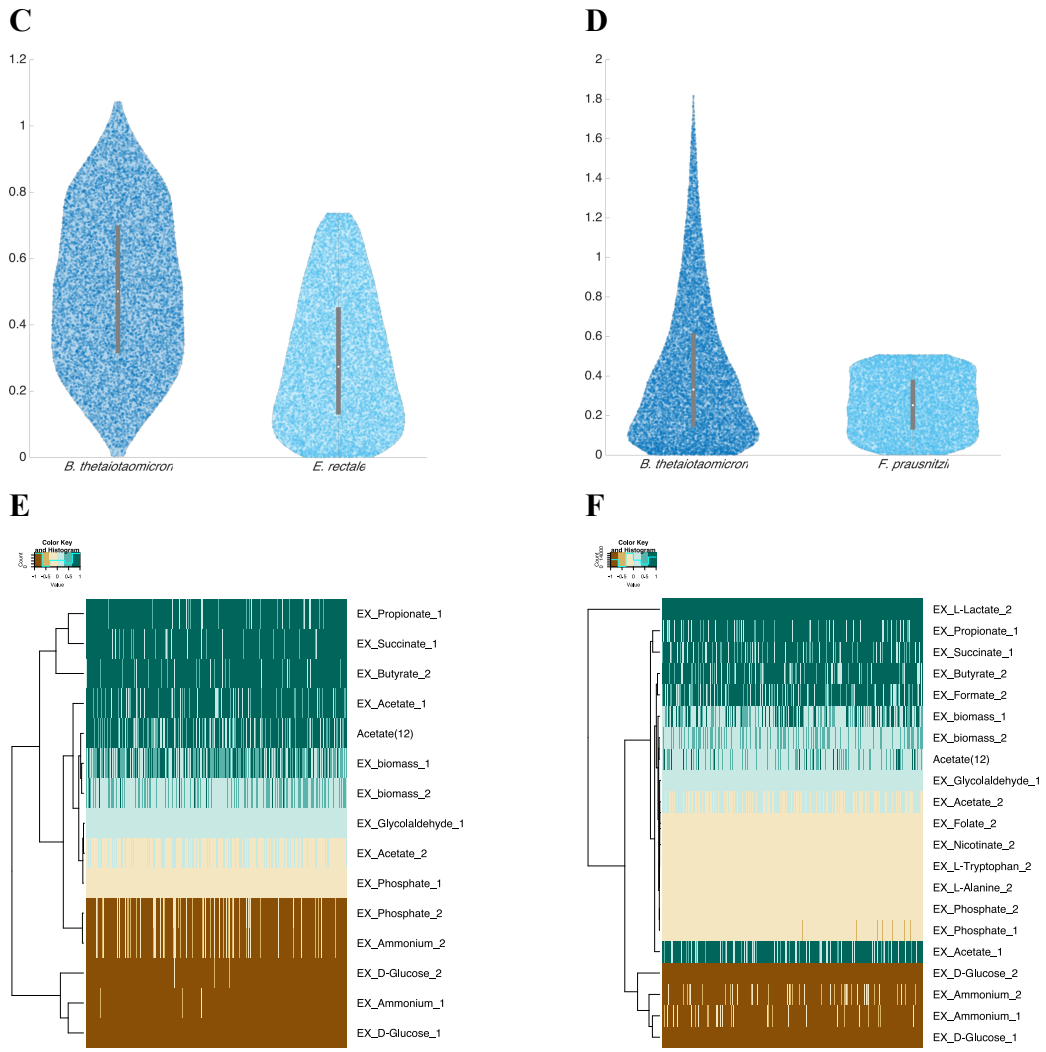


Figure 24. Distribution and clustering of the samples in two bi-member bacterial co-culture. A) Resource distribution and metabolic products in *B. thetaiotaomicron* and *E. rectale* co-culture. B) Resource distribution and metabolic products in *B. thetaiotaomicron* and *F. prausnitzii* co-culture. C) Distribution of the samples of the biomass fluxes in *B. thetaiotaomicron* and *E. rectale* co-culture. D) Distribution of the samples of the biomass fluxes in *B. thetaiotaomicron* and *F. prausnitzii* co-culture. E) Clustergram based on community-level flux samples in *B. thetaiotaomicron* and *E. rectale* co-culture. F) Clustergram based on community-level flux samples in *B. thetaiotaomicron* and *F. prausnitzii* co-culture.

Extending community members to three strains, we modelled and sampled tri-culture of *B. thetaiotaomicron*, *R. bromii* and *E. rectale*. In this community, the first two members are acetate producers, and the last one is a butyrogenic strain. These strains have been grown and studied by Ze et al. (Ze et al. 2012) on glucose and their OD changes have been measured. We used data from that study to calculate the growth rates of these species. The estimated growth rates were 0.27 h^{-1} for *R. bromii*, 0.26 h^{-1} for *B. thetaiotaomicron* and 0.09 h^{-1} for *E. rectale*. As shown in Figure 25, the samples related to the biomass reaction flux of *E. rectale* have a lower median and maximum compared to the two other community members. Furthermore, according to the clustergram, in this tri-culture the biomass of *E. rectale* does not cluster with the linking acetate reactions, namely Acetate(13) provided from *B. thetaiotaomicron* and Acetate(12) produced by *R. bromii*. We plotted the samples of the aforementioned linking

reactions in this community and compared the patterns to *B. thetaiotaomicron- E. rectale* co-culture. Figure 25E shows that in the bi-member community there is an increasing trend for the butyrogenic biomass as the amount of linking acetate rises. However, this pattern does not emerge in the tri-culture (Figure 25C-D) samples probably due the division of labour between the two acetate producers.

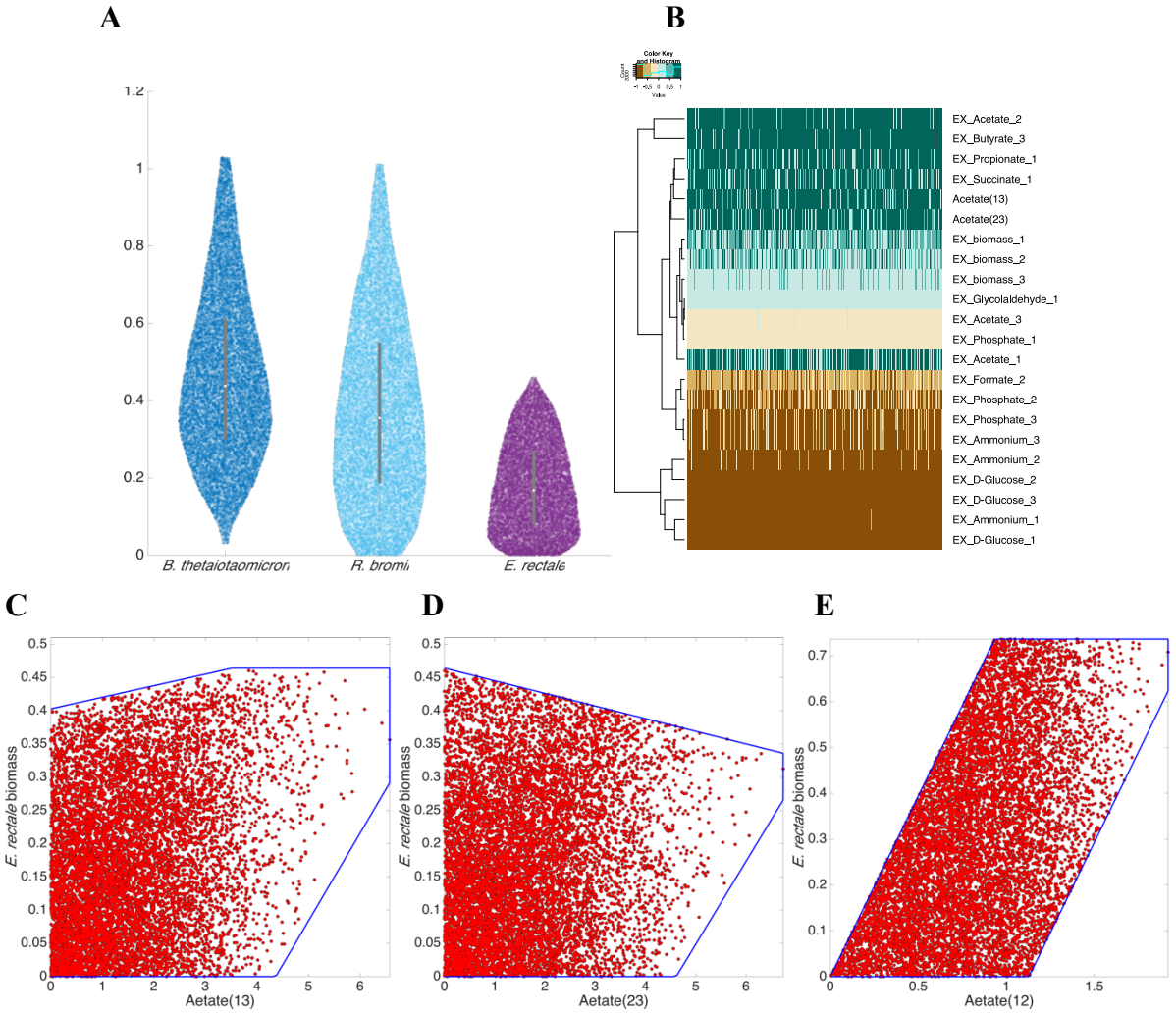
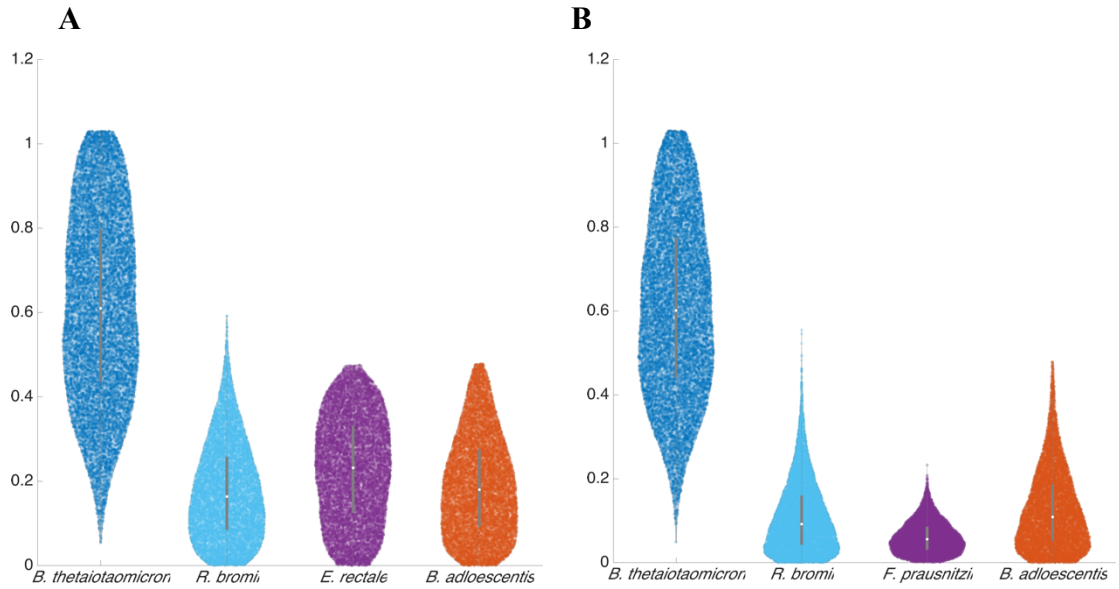


Figure 25. Samples of a tri-member microbial community consisted of *B. thetaiotaomicron*, *R. bromii* and *E. rectale* . A) distribution of the biomass flux samples. B) Clustergram based on community-level flux samples. Samples related to the linking acetate and the biomass flux of the butyrogenic strains concerning C) *B. thetaiotaomicron* and *E. rectale* in tri-culture. D) *R. bromii* and *E. rectale* tri-culture. E) *B. thetaiotaomicron* and *E. rectale* in bi-culture.

Finally, we sampled two gut microbial communities comprised of three acetogens and one butyrogenic strain and compared the results. The two modelled and sampled communities are EBBR consisted of *E. rectale*, *B. thetaiotaomicron*, *B. adolescentis*, and *R. bromii* and FBBR including *F. prausnitzii*, *B. thetaiotaomicron*, *B. adolescentis*, and *R. bromii*. These communities have been studied *in vitro* and we have used the available experimental data to evaluate computational predictions. Samples related to the biomass reactions as depicted in Figure 26A-B show that *B. thetaiotaomicron* has the potential to reach higher growth rates compared to the other community members within the co-cultivation boundaries which is consistent with experimental data (Shoaie et al. 2015). Regarding community metabolic production, experimental measurements for acetate and propionate in the two communities were almost equal with negligible differences. Sampling results also show similar acetate and propionate production potential in EBBR being marginally higher than FBBR. In the case of butyrate, the measurements showed higher production for EBBR. In sampling distributions, although EBBR is able to produce high amounts of butyrate, we noticed that FBBR could reach higher values. This might be due to capability the of *F. prausnitzii* to produce high amounts of butyrate. In regards to clustering patterns, in both cases the biomass samples of the butyrogenic strain clusters closely with that of *B. adolescentis* and *R. bromii* reflecting the importance of these two strains for the growth of the acetate consumer. The biomass reaction of the acetate consumers, however, do not cluster with the linking acetate reactions in any of the communities. The reason might be similar to that of the tri-culture community's clustering pattern where the number of acetogens is more than one.



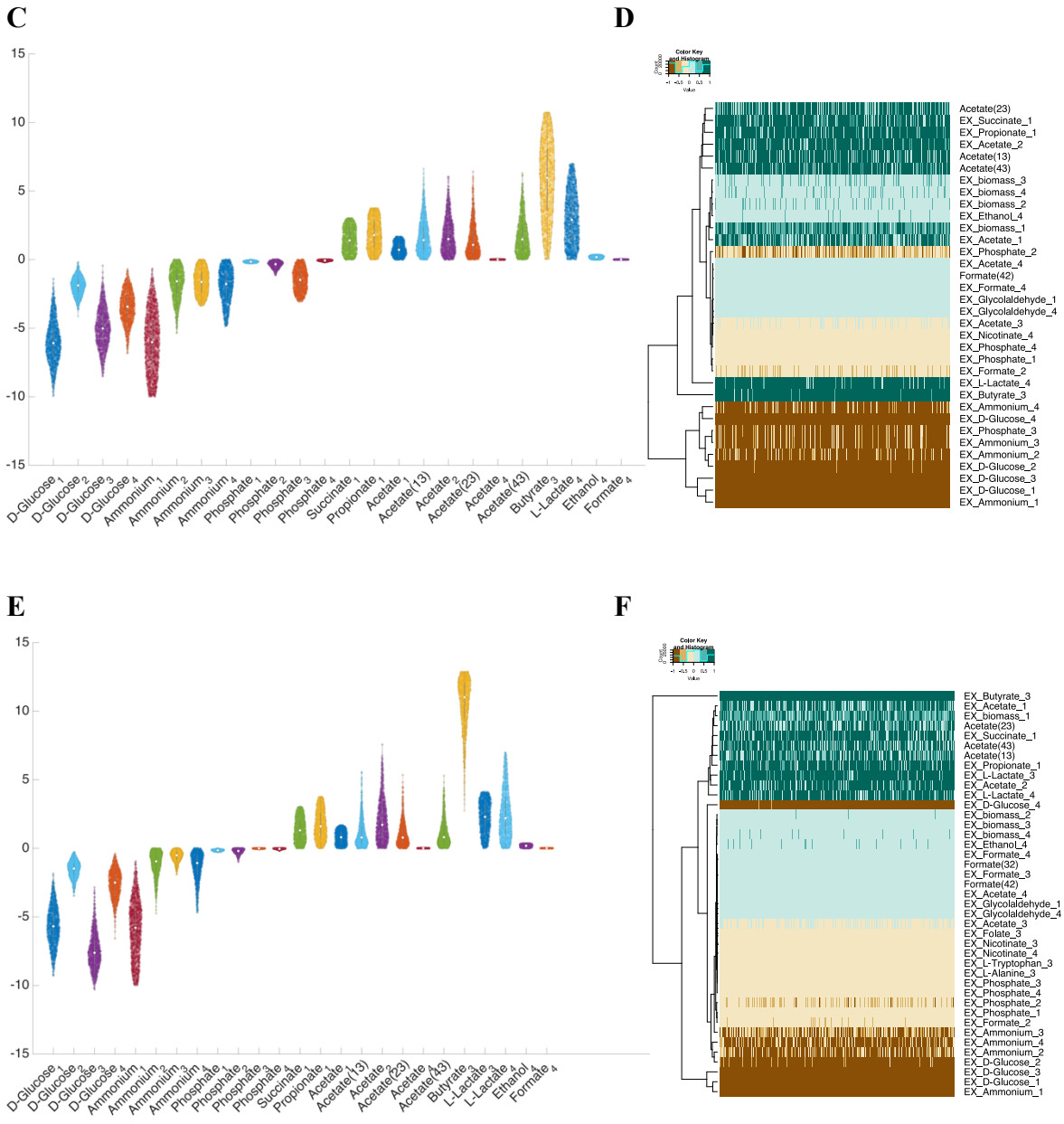


Figure 26. Distribution and clustering of the samples from two four-member bacteria communities. Biomass flux sample distributions in A) EBBR and B) FBBR. C) Distribution of the samples related to the consuming, linking and producing reactions in EBBR. D) Clustergram based on community-level flux samples in EBBR. E) Distribution of the samples related to the consuming, linking and producing reactions in FBBR. F) Clustergram based on community-level flux samples in FBBR. EBBR is comprised of *E. rectale*, *B. thetaiotaomicron*, *B. adolescentis*, and *R. bromii* and FBBR includes *F. prausnitzii*, *B. thetaiotaomicron*, *B. adolescentis*, and *R. bromii*.

4. Conclusion

In paper I and V, starting from metagenomics datasets of various diseased populations and their respective healthy control cohorts, we coupled -omics data analysis with computational modelling tools. In paper I, the studied diseased populations were malnourished children who underwent dietary intervention using ready-to-use therapeutic food (RUTF) from two countries, namely Bangladesh and Malawi. Faecal metagenomics data of a cohort of healthy Swedish infants were also investigated to elucidate diet and country-specific differences between the three groups. We observed an overall higher diversity in the gut microbiota of the Swedish infants compared to that of the other two countries. We then selected the 20 most abundant species in each community and reconstructed genome-scale metabolic models for each in order to compare the gut bacterial communities in a functional scale. By curating and validating these models we ensured the quality of these models and also reliability of their predictive power in terms of growth rates and potential for production of valuable metabolites known to be produced by the human gut bacterial commensals. Subsequently, we simulated the growth of the most abundant gut bacteria on both *in silico* human breast milk (HBM) and RUTF, which showed a higher capacity of growth on HBM compared to RUTF. Metabolic production potential of the gut bacterial communities in the three mentioned countries showed a higher potential in the case of Swedish infants. We also noticed that food therapy significantly improved metabolic production potentials of malnourished children in Bangladeshi cohort, however, for Malawian children the difference between healthy and malnourished children were not significant, highlighting the different response of gut microbiota to dietary interventions depending on compositional varieties. Results of microbial pairwise simulations were also in line with this observation pinpointing health-status and country-specific differences. In paper V, the focus was on vitamin producing reactions in the abundant species identified in diseased and healthy cohorts from four different countries, namely China, Denmark, Spain and USA. We investigated the presence and absence of such reactions and noticed that in most of the cases these pathways are incomplete which has also been observed in previous studies, pointing to possible cooperation for biosynthesis of these macromolecules. In this study, country-specific differences in terms of the prevalence of vitamin-related biosynthesis pathways were noticed.

GEMs are potent tools for omics data contextualization and predicting phenotypic outcomes in various biological scenarios when they have been properly reconstructed and curated. High variety of GEM reconstruction methods raises the question of how each method ensures the GEMs quality. In paper II, a comprehensive quality control was performed on two sets of models, namely AGORA which is a repository of gut bacterial models reconstructed semi-automatically and BiGG which includes a set of manually reconstructed and curated models. We compared the models in terms of ability to predict biologically feasible growth rates, the percentage of blocked reactions and dead-end metabolites, direct biomass metabolites, the metabolic similarity in terms of active reactions and biomass precursors. Results of this comparative approach demonstrated that the automatically generated models tend to contain mathematical artefacts such as thermodynamically infeasible loops and a high degree of similarity irrelevant of biological closeness. Therefore, it is crucial to manually curate automatically generated GEMs using experimental observations and other low- or high-throughput data to ensure the quality and credibility of the biological predictions.

High-quality GEMs have numerous applications amongst which food microbiology is a good example. Complex microbial starter cultures are used in food production and *in silico* tools have the potential to hypothesize different manipulation strategies to ameliorate such bacterial consortia towards higher or more efficient production of valuable nutrients. In paper III, we

reconstructed GEMs for four lactic acid bacterial strains used in the dairy formation and one Bifidobacterium with probiotic potentials. Here, we followed a manual approach to reconstruct and curate the models. Experimental fermentation data were used to evaluate *in silico* predictions. Pairwise growth of two strains used in yogurt fermentation was simulated, and an optimum ratio between the two was found which would result in optimum production of acetaldehyde, the major metabolite contributing to yogurt's unique taste and aroma. Furthermore, the co-culture of all of the five strains in another fermented milk product was simulated and validated using fermentation *in vitro* data. This work shows the potential of computational tools in providing testable hypothesis in case of challenging experiments.

Modelling microbial communities is still a challenge since complicated metabolic interactions, temporal and spatial heterogeneity, and changing dynamics of the community in response to environmental perturbances make it difficult to choose a modelling platform which would encompass all. There has been some effort in combining GEMs to model microbial communities using the regular approach which is optimisation of a pre-assumed biological objective function which is normally the growth rate in case of single strain models. This assumption, however, may not be valid for microbial communities. Therefore, in paper IV, we proposed a pipeline for modelling and uniformly sampling the feasible solution space of microbial communities. Starting from bi-cultures, we extended the microbial communities comprised of models for representatives of the human gut microbial strains to four and in each case sampled and studied the distribution of the samples and their clustering patterns. Experimental data were used to evaluate the outcomes of this method and in the majority of cases the method could make biologically meaningful results. Therefore, it is possible to characterise microbial communities in terms of metabolic and biological interactions using this method without any *a priori* objective functions.

5. Future perspectives

Early microbiome studies were focused on cataloguing the presence and abundance of either microbial species or their respective genes. More recent works, however, have tried to change the centre of attention from compositional descriptive approaches towards mechanistic ones which try to study microbial communities from a dynamic, functional, or evolutionary point of view following a holistic systems biology-based methodology. Even so, most of the microbiome-related studies so far have been favoured towards associative studies comparing the gut configuration in health and disease, mostly describing rises and falls in gut microbial taxa without any apparent causes or consequences.

Elucidating the underlying principles of microbial communities requires mechanistic and phenomenological models which would provide insight into effective manipulation strategies for microbial communities, rational design with synthetic microbes or even synthetic microbial communities, and even pave the way towards precision nutrition (Waldor et al. 2015). Engineering microbiota using phages has been proposed, but currently, the methods or molecules to target specific species are lacking. Employing synthetic bacteriophages has been shown to hold the potential of killing the pathogenic bacteria; therefore, they might be used for modulating microbial communities towards a particular composition (Citorik, Mimee, and Lu 2014). Another exciting potential therapeutic idea concerning microbiome is the possibility of inoculating malnourished infants with healthy microbiota to repair their immature gut community (Sela and Mills 2014). More profound knowledge of any possible short- and long-term effects of such therapeutic proposals needs to be addressed before they can be effectively used to tackle diseases and improve health.

Computational modelling is one of the pillars of systems biology and is an essential element for a holistic approach. In the case of metabolic networks, there are various methods to reconstruct models with different approaches. These differences, part of them non-biological, are usually inherited to the models reconstructed by each method leading to variations of prediction results using different models for the same organism for the same biological context and scenarios. This issue calls for a reconciliation effort towards unified procedures and standards for GEM reconstructions so that models could be studied comparatively in a way that differences would reflect biological distinctions rather than mathematical artefacts (Oberhardt et al. 2011; Babaei, Ghasemi-Kahrizsangi, and Marashi 2014).

Another issue with GEMs is throughput versus accuracy. High-throughput methods have been developed for GEM reconstruction which, in one hand, are very useful in reconciliation terms since obviously, the resulting models are uniform in terms of reconstruction method and furthermore, they are very efficient timewise. On the other hand, however, the accuracy of these models do not compare with manually reconstructed and curated ones since a significant amount of biological literature and experimental data have been embedded in their structure. Additionally, since there is no universal evaluating platform for metabolic models, it makes evaluating these methods even less objective. A community-driven effort for the development of such platform has been recently performed (Lieven et al. 2018), and hopefully, similar future works will continue to contribute to the improvement of the models' quality and accuracy of their predictive power.

It is evident that for modelling microbial communities, reconstruction of any types of models for its constituents is not sufficient, regardless of their quality and predictive power. Features such as metabolic and biological interactions, emergent properties, tempo-spatial heterogeneity, and quorum sensing require accurate modelling platforms so that it could lead to biologically meaningful predictions. Current microbial community simulation methods can

only simulate a small community of microbial species comprising of a few species, however, it is apparent that modelling a complicated bacterial community such as the human gut microbiota consisting of thousands of bacterial strains needs more sophisticated modelling approaches so that it could be deeply understood and its interrelationships with other factors such as diet, health status, antibiotic usage, probiotic administration and other gut community modulators could be unravelled. Finally, modelling any biological processes demands relevant biological data so that the predictions could be compared to the phenomenon happening in nature. In the case of microbial communities, meta-omics data including metatranscriptomics, metaproteomics, and metametabolomics are essential for characterisation of these multifaceted biological assemblages. Research will continue to generate even more multi-dimensional data for different microbial communities from either *in vitro* culturing of microbial communities or controlled animal models and even clinical trials, however, integrating the data avalanche with a proper mechanistic scaffold will be the biggest challenge for the field.

6. Acknowledgements

First and foremost, I am thankful to my supervisor Prof. Jens Nielsen who gave me this unique and challenging opportunity for personal and professional growth. Your patience, efficiency, discipline and vast knowledge allowed me to explore my ideas and shape the research presented in this thesis. If I achieve half of what you have in my lifetime, I will consider myself very successful. I have learned immensely from you and for that, I am extremely delighted and grateful.

None of the work in this thesis would have been possible without the fruitful collaboration and discussions that I have had with many people during my studies. Firstly, I am deeply grateful to my co-supervisor Adil Mardinoglu, your determination and fast progress in research has always amazed me and continues to do so. Dr. Rasmus Ågren, scientific discussions with you in my first year of studies were priceless and extremely helpful for me to shape my critical thinking. I am tremendously thankful to Dr. Manish Kumar. Without your discipline and determination, it would have been impossible to have such a successful collaboration in BMGF project. Dr. Boyang Ji, your critical and accurate scientific vision on many areas of research has influenced me considerably, and I am pleased to have had the chance to work with you on several projects. Dr. Saeed Shoaie, your ambitious ideas and persistence in research has enabled me to overcome many obstacles in our work together. Promi, more than half of what is included in this thesis has been done with you, other than work, you've been an amazing friend to whom I could talk about anything. I have also enjoyed scientific discussions, especially in the Microbial Ecology sub-group, with Sylvain, Pouyan, Amir, Jun, and many other people during the Friday seminars. I should also express my gratitude towards our external collaborators from Danone, Dr. Christian Chervaux and Dr. Jean-Michel Faurie, with whom I had many productive discussions and learned a lot.

Additionally, I would like to express my gratitude towards people in the administration, Anna Bolling, Erica Dahlin, Josefine Jäwert, Anne-Lise Kramer, and Martina Butorac, thank you for answering all my questions and helping me whenever I was confused about administrative procedures.

I would like to thank people with whom I have shared many memorable moments. Aida, you were the first person I met when I arrived in Sweden and you've been just amazing throughout the countless moments we shared together. Demi, thank you for being so supportive and caring. I would also like to express my gratitude towards Azadeh, Shaq, Ling, Rasool, Yasaman, JC, Benjamin, Avlant and many more people in Sysbio who will be an unforgettable part of my PhD studies, for all the friendly gatherings, after works, talks, conferences and trips we went together.

Last but not least, I am deeply grateful to my parents whose unconditional love and support have made this journey possible and pleasant by being there whenever I needed them, and my sister whose energy and positivity has helped my go through many difficulties not only in this work, but throughout all those years that we grew up together.

7. References

Abubucker, Sahar, Nicola Segata, Johannes Goll, Alyxandria M Schubert, Jacques Izard, Brandi L Cantarel, Beltran Rodriguez-Mueller, Jeremy Zucker, Mathangi Thiagarajan, and Bernard Henrissat. 2012. 'Metabolic reconstruction for metagenomic data and its application to the human microbiome', *PLoS computational biology*, 8: e1002358.

Agren, Rasmus, Liming Liu, Saeed Shoaie, Wanwipa Vongsangnak, Intawat Nookaew, and Jens Nielsen. 2013. 'The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum', *PLoS computational biology*, 9: e1002980.

Aguiar-Pulido, Vanessa, Wenrui Huang, Victoria Suarez-Ulloa, Trevor Cickovski, Kalai Mathee, and Giri Narasimhan. 2016. 'Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis: Supplementary Issue: Bioinformatics Methods and Applications for Big Metagenomics Data', *Evolutionary Bioinformatics*, 12: EBO. S36436.

Allison, Steven D, and Jennifer BH Martiny. 2008. 'Resistance, resilience, and redundancy in microbial communities', *Proceedings of the National Academy of Sciences*, 105: 11512-19.

Arakawa, Kazuharu, Yohei Yamada, Kosaku Shinoda, Yoichi Nakayama, and Masaru Tomita. 2006. 'GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes', *BMC bioinformatics*, 7: 168.

Arkin, Adam P, Robert W Cottingham, Christopher S Henry, Nomi L Harris, Rick L Stevens, Sergei Maslov, Paramvir Dehal, Doreen Ware, Fernando Perez, and Shane Canon. 2018. 'KBase: the United States department of energy systems biology knowledgebase', *Nature Biotechnology*, 36.

Babaei, Parizad, Tahereh Ghasemi-Kahrizsangi, and Sayed-Amir Marashi. 2014. 'Modeling the differences in biochemical capabilities of pseudomonas species by flux balance analysis: how good are genome-scale metabolic networks at predicting the differences?', *The Scientific World Journal*, 2014.

Babaei, Parizad, Sayed-Amir Marashi, and Sedigheh Asad. 2015. 'Genome-scale reconstruction of the metabolic network in Pseudomonas stutzeri A1501', *Molecular BioSystems*, 11: 3022-32.

Bachmann, Herwig, Douwe Molenaar, Filipe Branco dos Santos, and Bas Teusink. 2017. 'Experimental evolution and the adjustment of metabolic strategies in lactic acid bacteria', *FEMS Microbiology Ecology*, 41: S201-S19.

Bäckhed, Fredrik. 2011. 'Programming of host metabolism by the gut microbiota', *Annals of Nutrition and Metabolism*, 58: 44-52.

Bäckhed, Fredrik, Ruth E Ley, Justin L Sonnenburg, Daniel A Peterson, and Jeffrey I Gordon. 2005. 'Host-bacterial mutualism in the human intestine', *science*, 307: 1915-20.

Bäckhed, Fredrik, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, Yan Xia, Hailiang Xie, and Huanzi Zhong. 2015. 'Dynamics and stabilization of the human gut microbiome during the first year of life', *Cell host & microbe*, 17: 690-703.

Balagurunathan, Balaji, Sudhakar Jonnalagadda, Lily Tan, and Rajagopalan Srinivasan. 2012. 'Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis', *Microbial cell factories*, 11: 27.

Barton, Larry , and Diana Northup. 2011. *Microbial Ecology* (Wiley-Blackwell).

Bauer, Eugen, Johannes Zimmermann, Federico Baldini, Ines Thiele, and Christoph Kaleta. 2017. 'BacArena: individual-based metabolic modeling of heterogeneous microbes in complex communities', *PLoS computational biology*, 13: e1005544.

Belcheva, Antoaneta, Thergiory Irrazabal, Susan J Robertson, Catherine Streutker, Heather Maughan, Stephen Rubino, Eduardo H Moriyama, Julia K Copeland, Anu Surendra, and Sachin Kumar. 2014. 'Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells', *Cell*, 158: 288-99.

Bellucci, Micol, Irina D Ofiţeru, Luciano Beneduce, David W Graham, Ian M Head, and Thomas P Curtis. 2015. 'A preliminary and qualitative study of resource ratio theory to nitrifying lab-scale bioreactors', *Microbial Biotechnology*, 8: 590-603.

Benedict, Matthew N, Michael B Mundy, Christopher S Henry, Nicholas Chia, and Nathan D Price. 2014. 'Likelihood-based gene annotations for gap filling and quality assessment in genome-scale metabolic models', *PLoS computational biology*, 10: e1003882.

Biggs, Matthew B, Gregory L Medlock, Glynis L Kolling, and Jason A Papin. 2015. 'Metabolic network modeling of microbial communities', *Systems Biology and Medicine*, 7: 317-34.

Biggs, Matthew B, and Jason A Papin. 2013. 'Novel multiscale modeling tool applied to Pseudomonas aeruginosa biofilm formation', *PloS one*, 8: e78011.

Blanton, Laura V, Mark R Charbonneau, Tarek Salih, Michael J Barratt, Siddarth Venkatesh, Olga Ilkaveya, Sathish Subramanian, Mark J Manary, Indi Trehan, and Josh M Jorgensen. 2016. 'Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children', *science*, 351: aad3311.

Bochner, Barry R, Peter Gadzinski, and Eugenia Panomitros. 2001. 'Phenotype microarrays for high-throughput phenotypic testing and assay of gene function', *Genome research*, 11: 1246-55.

Boele, Joost, Brett G Olivier, and Bas Teusink. 2012. 'FAME, the flux analysis and modeling environment', *BMC systems biology*, 6: 8.

Bokulich, Nicholas A, Zachery T Lewis, Kyria Boundy-Mills, and David A Mills. 2016. 'A new perspective on microbial landscapes within food production', *Current Opinion in Biotechnology Journal*, 37: 182-89.

Bongers, Roger S, Marcel HN Hoefnagel, and Michiel Kleerebezem. 2005. 'High-level acetaldehyde production in Lactococcus lactis by metabolic engineering', *Applied and environmental microbiology*, 71: 1109-13.

Bordbar, Aarash, Nathan E Lewis, Jan Schellenberger, Bernhard Ø Palsson, and Neema Jamshidi. 2010. 'Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions', *Molecular systems biology*, 6: 422.

Borre, Yuliya E, Rachel D Moloney, Gerard Clarke, Timothy G Dinan, and John F Cryan. 2014. 'The impact of microbiota on brain and behavior: mechanisms & therapeutic potential.' in, *Microbial Endocrinology: The Microbiota-Gut-Brain Axis in Health and Disease* (Springer).

Brooks, J Paul, William P Burns, Stephen S Fong, Chris M Gowen, and Seth B Roberts. 2012. 'Gap detection for genome-scale constraint-based models', *Advances in Bioinformatics*, 2012.

Brown, Amy C, and Ana Valiere. 2004. 'Probiotics and medical nutrition therapy', *Nutrition in Clinical Care*, 7: 56.

Brüssow, Harald. 2016. 'How stable is the human gut microbiota? And why this question matters', *Environmental Microbiology*, 18: 2779-83.

Burke, Catherine, Peter Steinberg, Doug Rusch, Staffan Kjelleberg, and Torsten Thomas. 2011. 'Bacterial community assembly based on functional genes rather than species', *Proceedings of the National Academy of Sciences*, 108: 14288-93.

Cani, Patrice D, Matthias Van Hul, Charlotte Lefort, Clara Depommier, Marialetizia Rastelli, and Amandine Everard. 2019. 'Microbial regulation of organismal energy homeostasis', *Nature Metabolism*, 1: 34.

Cantrell, Robert Stephen, and Chris Cosner. 2004. *Spatial ecology via reaction-diffusion equations* (John Wiley & Sons).

Cardwell, Christopher R, LC Stene, G Joner, O Cinek, J Svensson, MJ Goldacre, RC Parslow, P Pozzilli, G Brigis, and D Stoyanov. 2008. "Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies." In.: Springer.

Chang, Christopher, and Henry Lin. 2016. 'Dysbiosis in gastrointestinal disorders', *Best Practice & Research Clinical Gastroenterology*, 30: 3-15.

Chaves, ACSD, M Fernandez, ALS Lerayer, I Mierau, M Kleerebezem, and J Hugenholtz. 2002. 'Metabolic engineering of acetaldehyde production by Streptococcus thermophilus', *Applied & Environmental Microbiology*, 68: 5656-62.

Cheng, Hefa. 2010. 'Volatile flavor compounds in yogurt: a review', *Critical reviews in food science and nutrition*, 50: 938-50.

Chiu, Hsuan-Chao, Roie Levy, and Elhanan Borenstein. 2014. 'Emergent biosynthetic capacity in simple microbial communities', *PLoS computational biology*, 10: e1003695.

Citorik, Robert J, Mark Mimee, and Timothy K Lu. 2014. 'Bacteriophage-based synthetic biology for the study of infectious diseases', *Current opinion in microbiology*, 19: 59-69.

Cocolin, Luca, and Danilo Ercolini. 2015. 'Zooming into food-associated microbial consortia: a 'cultural'evolution', *Current Opinion in Food Science*, 2: 43-50.

Cosner, C. 2008. 'Reaction–diffusion equations and ecological modeling.' in, *Tutorials in Mathematical Biosciences IV* (Springer).

Coyte, Katharine Z, Jonas Schluter, and Kevin R Foster. 2015. 'The ecology of the microbiome: networks, competition, and stability', *science*, 350: 663-66.

Cuevas, Daniel A, Janaka Edirisinghe, Chris S Henry, Ross Overbeek, Taylor G O'Connell, and Robert A Edwards. 2016. 'From DNA to FBA: how to build your own genome-scale metabolic model', *Frontiers in microbiology*, 7: 907.

D'Acunto, B, L Frunzo, I Klapper, and MR Mattei. 2015. 'Modeling multispecies biofilms including new bacterial species invasion', *Mathematical Biosciences*, 259: 20-26.

Davidson, Robert T, Andrea L Foley, Jean A Engelke, and John W Suttie. 1998. 'Conversion of dietary phylloquinone to tissue menaquinone-4 in rats is not dependent on gut bacteria', *The Journal of Nutrition*, 128: 220-23.

Dias, Oscar, Miguel Rocha, Eugénio C Ferreira, and Isabel Rocha. 2015. 'Reconstructing genome-scale metabolic models with merlin', *Nucleic acids research*, 43: 3899-910.

Dominguez-Bello, Maria G, Elizabeth K Costello, Monica Contreras, Magda Magris, Glida Hidalgo, Noah Fierer, and Rob Knight. 2010. 'Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns', *Proceedings of the National Academy of Sciences*, 107: 11971-75.

Donaldson, Gregory P, S Melanie Lee, and Sarkis K Mazmanian. 2016. 'Gut biogeography of the bacterial microbiota', *Nature Reviews Microbiology*, 14: 20.

Donohoe, Dallas R, Nikhil Garge, Xinxin Zhang, Wei Sun, Thomas M O'Connell, Maureen K Bunger, and Scott J Bultman. 2011. 'The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon', *Cell Metabolism*, 13: 517-26.

Döring, Frank, Judith Walter, Jutta Will, Melanie Föcking, Michael Boll, Salah Amasheh, Wolfgang Clauss, and Hannelore Daniel. 1998. 'Delta-aminolevulinic acid transport by intestinal and renal peptide transporters and its physiological and clinical implications', *Journal of Clinical Investigation*, 101: 2761-67.

Eisen, Jonathan A. 2007. 'Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes', *PLoS biology*, 5: e82.

El-Semman, Ibrahim E, Fredrik H Karlsson, Saeed Shoaie, Intawat Nookaew, Taysir H Soliman, and Jens Nielsen. 2014. 'Genome-scale metabolic reconstructions of Bifidobacterium adolescentis L2-32 and Faecalibacterium prausnitzii A2-165 and their interaction', *BMC systems biology*, 8: 41.

Faith, Jeremiah J, Janaki L Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L Goodman, Jose C Clemente, Rob Knight, Andrew C Heath, and Rudolph L Leibel. 2013. 'The long-term stability of the human gut microbiota', *science*, 341: 1237439.

Falkowski, Paul G, Tom Fenchel, and Edward F Delong. 2008. 'The microbial engines that drive Earth's biogeochemical cycles', *science*, 320: 1034-39.

Faria, José P, Tahmineh Khazaei, Janaka N Edirisinghe, Pamela Weisenhorn, Samuel MD Seaver, Neal Conrad, Nomi Harris, Matthew DeJongh, and Christopher S Henry. 2016. 'Constructing and analyzing metabolic flux models of microbial communities.' in, *Hydrocarbon and Lipid Microbiology Protocols* (Springer).

Faria, José P, Miguel Rocha, Isabel Rocha, and Christopher S Henry. 2018. 'Methods for automated genome-scale metabolic model reconstruction', *Biochemical Society Transactions*, 46: 931-36.

Feist, Adam M, and Bernhard O Palsson. 2010. 'The biomass objective function', *Current opinion in microbiology*, 13: 344-49.

Fernández, Ana, Suiying Huang, Sherry Seston, Jian Xing, Robert Hickey, Craig Criddle, and James Tiedje. 1999. 'How stable is stable? Function versus community composition', *Applied and environmental microbiology*, 65: 3697-704.

Ferrario, Chiara, Valentina Taverniti, Christian Milani, Walter Fiore, Monica Laureati, Ivano De Noni, Milda Stuknyte, Bessem Chouaia, Patrizia Riso, and Simone Guglielmetti. 2014. 'Modulation of Fecal Clostridiales Bacteria and Butyrate by Probiotic Intervention with Lactobacillus paracasei DG Varies among Healthy Adults–3', *Journal of Nutrition*, 144: 1787-96.

Ferrell, James E. 2009. 'Q&A: systems biology', *Journal of biology*, 8: 2.

Flahaut, Nicolas AL, Anne Wiersma, Bert van de Bunt, Dirk E Martens, Peter J Schaap, Lolke Sijtsma, Vitor A Martins dos Santos, and Willem M de Vos. 2013. 'Genome-scale metabolic model for Lactococcus lactis MG1363 and its application to the analysis of flavor formation', *Applied microbiology and biotechnology*, 97: 8729-39.

Freilich, Shiri, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. 2011. 'Competitive and cooperative metabolic interactions in bacterial communities', *Nature communications*, 2: 589.

Fritzemeier, Claus Jonathan, Daniel Hartleb, Balázs Szappanos, Balázs Papp, and Martin J Lercher. 2017. 'Erroneous energy-generating cycles in published genome scale

metabolic networks: Identification and removal', *PLoS computational biology*, 13: e1005494.

Gaenzle, Michael G. 2015. 'Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage', *Current Opinion in Food Science*, 2: 106-17.

Ganter, Mathias, Hans-Michael Kaltenbach, and Jörg Stelling. 2014. 'Predicting network functions with nested patterns', *Nature communications*, 5: 3006.

Gill, Steven R, Mihai Pop, Robert T DeBoy, Paul B Eckburg, Peter J Turnbaugh, Buck S Samuel, Jeffrey I Gordon, David A Relman, Claire M Fraser-Liggett, and Karen E Nelson. 2006. 'Metagenomic analysis of the human distal gut microbiome', *science*, 312: 1355-59.

Gutleben, Johanna, Maryam Chaib De Mares, Jan Dirk van Elsas, Hauke Smidt, Jörg Overmann, and Detmer Sipkema. 2018. 'The multi-omics promise in context: from sequence to microbial isolate', *Critical Reviews in Microbiology*, 44: 212-29.

Hamilton, Joshua J, and Jennifer L Reed. 2014. 'Software platforms to facilitate reconstructing genome-scale metabolic networks', *Environmental microbiology*, 16: 49-59.

Han, XY, DQ Sun, L Xiang, GC Huo, and YJ Jiang. 2007. 'Gene regulation to lactic acid bacteria for increasing production of flavor metabolite', *Wei Sheng Wu Xue Bao*, 47: 1105-09.

Harcombe, William R, William J Riehl, Ilija Dukovski, Brian R Granger, Alex Betts, Alex H Lang, Gracia Bonilla, Amrita Kar, Nicholas Leiby, and Pankaj Mehta. 2014. 'Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics', *Cell Reports*, 7: 1104-15.

Hartleb, Daniel, Florian Jarre, and Martin J Lercher. 2016. 'Improved Metabolic Models for E. coli and Mycoplasma genitalium from GlobalFit, an Algorithm That Simultaneously Matches Growth and Non-Growth Data Sets', *PLoS computational biology*, 12: e1005036.

Heinken, Almut, Swagatika Sahoo, Ronan MT Fleming, and Ines Thiele. 2013. 'Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut', *Gut Microbes*, 4: 28-40.

Heinken, Almut, and Ines Thiele. 2015. 'Anoxic conditions promote species-specific mutualism between gut microbes in silico', *Applied and environmental microbiology*: AEM. 00101-15.

Hellweger, Ferdi L, Robert J Clegg, James R Clark, Caroline M Plugge, and Jan-Ulrich Kreft. 2016. 'Advancing microbial sciences by individual-based modelling', *Nature Reviews Microbiology*, 14: 461.

Henry, Christopher S, Hans C Bernstein, Pamela Weisenhorn, Ronald C Taylor, Joon-Yong Lee, Jeremy Zucker, and Hyun-Seob Song. 2016. 'Microbial community metabolic modeling: a community data-driven network reconstruction', *Journal of Cellular Physiology*, 231: 2339-45.

Henry, Christopher S, Matthew DeJongh, Aaron A Best, Paul M Frybarger, Ben Linsay, and Rick L Stevens. 2010. 'High-throughput generation, optimization and analysis of genome-scale metabolic models', *Nature Biotechnology*, 28: 977.

Henry, Christopher S, Ella Rotman, Wyndham W Lathem, Keith EJ Tyo, Alan R Hauser, and Mark J Mandel. 2017. 'Generation and Validation of the iKp1289 Metabolic Model for Klebsiella pneumoniae KPPR1', *The Journal of Infectious Diseases*, 215: S37-S43.

Henry, Christopher S, Jenifer F Zinner, Matthew P Cohoon, and Rick L Stevens. 2009. 'i Bsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations', *Genome biology*, 10: R69.

Herrgård, Markus J, Stephen S Fong, and Bernhard Ø Palsson. 2006. 'Identification of genome-scale metabolic network models using experimentally measured flux profiles', *PLoS computational biology*, 2: e72.

Hill, Daragh, Reynolds P Ross, Elke Arendt, and Catherine Stanton. 2017. 'Microbiology of Yogurt and Bio-Yogurts Containing Probiotics and Prebiotics.' in, *Yogurt in Health and Disease Prevention* (Elsevier).

Hooper, Lora V, Dan R Littman, and Andrew J Macpherson. 2012. 'Interactions between the microbiota and the immune system', *science*, 336: 1268-73.

Hosseini, Zhaleh, and Sayed-Amir Marashi. 2017. 'Discovering missing reactions of metabolic networks by using gene co-expression data', *Scientific reports*, 7: 41774.

Hug, Laura A, Brett J Baker, Karthik Anantharaman, Christopher T Brown, Alexander J Probst, Cindy J Castelle, Cristina N Butterfield, Alex W Hernsdorf, Yuki Amano, and Kotaro Ise. 2016. 'A new view of the tree of life', *Nature microbiology*, 1: 16048.

Hughes, Heather K, Destanie Rose, and Paul Ashwood. 2018. 'The gut microbiota and dysbiosis in Autism spectrum disorders', *Current Neurology and Neuroscience Reports*, 18: 81.

Hurlbert, Stuart H. 1971. 'The nonconcept of species diversity: a critique and alternative parameters', *Ecology*, 52: 577-86.

Huttenhower, Curtis, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H Badger, Asif T Chinwalla, Heather H Creasy, Ashlee M Earl, Michael G FitzGerald, and Robert S Fulton. 2012. 'Structure, function and diversity of the healthy human microbiome', *Nature*, 486: 207.

Jandhyala, Sai Manasa, Rupjyoti Talukdar, Chivkula Subramanyam, Harish Vuyyuru, Mitnala Sasikala, and D Nageshwar Reddy. 2015. 'Role of the normal gut microbiota', *World Journal of Gastroenterology*, 21: 8787.

Ji, Boyang, and Jens Nielsen. 2015. 'New insight into the gut microbiome through metagenomics', *Advances in Genomics and Genetics*, 5: 77-91.

Kandler, Otto. 1983. 'Carbohydrate metabolism in lactic acid bacteria', *Antonie Van Leeuwenhoek*, 49: 209-24.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. 'BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences', *Journal of molecular biology*, 428: 726-31.

Karlsen, Emil, Christian Schulz, and Eivind Almaas. 2018. 'Automated generation of genome-scale metabolic draft reconstructions based on KEGG', *BMC bioinformatics*, 19: 467.

Karp, Peter D, Mario Latendresse, Suzanne M Paley, Markus Krummenacker, Quang D Ong, Richard Billington, Anamika Kothari, Daniel Weaver, Thomas Lee, and Pallavi Subhraveti. 2015. 'Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology', *Briefings in Bioinformatics*, 17: 877-90.

Karp, Peter D, Daniel Weaver, and Mario Latendresse. 2018. 'How accurate is automated gap filling of metabolic models?', *BMC systems biology*, 12: 73.

Khandelwal, Ruchir A, Brett G Olivier, Wilfred FM Röling, Bas Teusink, and Frank J Bruggeman. 2013. 'Community flux balance analysis for microbial consortia at balanced growth', *PloS one*, 8: e64567.

Kim, Myunghoo, Carolina Galan, Andrea A Hill, Wan-Jung Wu, Hannah Fehlner-Peach, Hyo Won Song, Deborah Schady, Matthew L Bettini, Kenneth W Simpson, and Randy S Longman. 2018. 'Critical role for the microbiota in CX3CR1+ intestinal mononuclear phagocyte regulation of intestinal T cell responses', *Immunity*, 49: 151-63. e5.

King, Zachary A, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. 2015. 'BiGG Models: A platform for integrating, standardizing and sharing genome-scale models', *Nucleic acids research*, 44: D515-D22.

Kitano, Hiroaki. 2001. *Foundations of systems biology* (MIT press Cambridge).

———. 2002. 'Systems biology: a brief overview', *science*, 295: 1662-64.

Kleiner, Manuel, Xiaoli Dong, Tjorven Hinzke, Juliane Wippler, Erin Thorson, Bernhard Mayer, and Marc Strous. 2018. 'Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities', *Proceedings of the National Academy of Sciences*, 115: E5576-E84.

Klitgord, Niels, and Daniel Segrè. 2010. 'Environments that induce synthetic microbial ecosystems', *PLoS computational biology*, 6: e1001002.

Koh, Ara, Filipe De Vadder, Petia Kovatcheva-Datchary, and Fredrik Bäckhed. 2016. 'From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites', *Cell*, 165: 1332-45.

Kommineni, Sushma, Daniel J Bretl, Vy Lam, Rajrupa Chakraborty, Michael Hayward, Pippa Simpson, Yumei Cao, Pavlos Bousounis, Christopher J Kristich, and Nita H Salzman. 2015. 'Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract', *Nature*, 526: 719.

Konopka, Allan. 2009. 'What is microbial community ecology?', *The ISME journal*, 3: 1223.

Kovatcheva-Datchary, Petia, Valentina Tremaroli, and Fredrik Bäckhed. 2013. 'The gut microbiota.' in, *The prokaryotes* (Springer).

Kreft, Jan-Ulrich, Caroline M Plugge, Volker Grimm, Clara Prats, Johan HJ Leveau, Thomas Banitz, Stephen Baines, James Clark, Alexandra Ros, and Isaac Klapper. 2013. 'Mighty small: observing and modeling individual microbes becomes big science', *Proceedings of the National Academy of Sciences*, 110: 18027-28.

Kristensen, Nadja B, Thomas Bryrup, Kristine H Allin, Trine Nielsen, Tue H Hansen, and Oluf Pedersen. 2016. 'Alterations in fecal microbiota composition by probiotic supplementation in healthy adults: a systematic review of randomized controlled trials', *Genome medicine*, 8: 52.

Kumar, Vinay Satish, Madhukar S Dasika, and Costas D Maranas. 2007. 'Optimization based automated curation of metabolic reconstructions', *BMC bioinformatics*, 8: 212.

Kumar, Vinay Satish, and Costas D Maranas. 2009. 'GrowMatch: an automated method for reconciling in silico/in vivo growth predictions', *PLoS computational biology*, 5: e1000308.

Lafferty, Kevin D, Giulio DeLeo, Cheryl J Briggs, Andrew P Dobson, Thilo Gross, and Armand M Kuris. 2015. 'A general consumer-resource population model', *science*, 349: 854-57.

Lambert, Guillaume, Saurabh Vyawahare, and Robert H Austin. 2014. 'Bacteria and game theory: the rise and fall of cooperation in spatially heterogeneous environments', *Interface Focus*, 4: 20140029.

Lane, David J, Bernadette Pace, Gary J Olsen, David A Stahl, Mitchell L Sogin, and Norman R Pace. 1985. 'Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses', *Proceedings of the National Academy of Sciences*, 82: 6955-59.

Larhlimi, Abdelhalim, Laszlo David, Joachim Selbig, and Alexander Bockmayr. 2012. 'F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks', *BMC bioinformatics*, 13: 57.

Laubereau, B, B Filipiak-Pittroff, A Von Berg, A Grübl, D Reinhardt, HE Wichmann, and S Koletzko. 2004. 'Caesarean section and gastrointestinal symptoms, atopic dermatitis, and sensitisation during the first year of life', *Archives of Disease in Childhood*, 89: 993-97.

Laursen, Martin Frederik, Rikke Pilmann Laursen, Anni Larnkjær, Kim F Michaelsen, Martin Iain Bahl, and Tine Rask Licht. 2017. 'Administration of two probiotic strains during early childhood does not affect the endogenous gut microbiota composition despite probiotic proliferation', *BMC microbiology*, 17: 175.

LeBlanc, Jean Guy, Christian Milani, Graciela Savoy de Giori, Fernando Sesma, Douwe van Sinderen, and Marco Ventura. 2013. 'Bacteria as vitamin suppliers to their host: a gut microbiota perspective', *Current opinion in biotechnology*, 24: 160-68.

Lespinet, Olivier, and Bernard Labedan. 2005. 'Orphan enzymes?', *science*, 307: 42-42.

Lewis, Nathan E, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, and Daniel Lopez-Ferrer. 2010. 'Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models', *Molecular systems biology*, 6: 390.

Li, Junhua, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, Manimozhiyan Arumugam, Jens Roat Kultima, Edi Prifti, and Trine Nielsen. 2014. 'An integrated catalog of reference genes in the human gut microbiome', *Nature Biotechnology*, 32: 834.

Lieven, Christian, Moritz Emanuel Beber, Brett G Olivier, Frank T Bergmann, Meric Ataman, Parizad Babaei, Jennifer A Bartell, Lars M Blank, Siddharth Chauhan, and Kevin Correia. 2018. 'Memote: A community-driven effort towards a standardized genome-scale metabolic model test suite', *bioRxiv*: 350991.

Lin, Lan, and Jianqiong Zhang. 2017. 'Role of intestinal microbiota and metabolites on gut homeostasis and human diseases', *BMC Immunology*, 18: 2.

Louca, Stilianos, and Michael Doebeli. 2015. 'Calibration and analysis of genome-based models for microbial ecology', *eLife*, 4: e08208.

Louca, Stilianos, Saulo MS Jacques, Aliny PF Pires, Juliana S Leal, Diane S Srivastava, Laura Wegener Parfrey, Vinicius F Farjalla, and Michael Doebeli. 2017. 'High taxonomic variability despite stable functional structure across microbial communities', *Nature Ecology and Evolution*, 1: 0015.

Louis, Petra, and Harry J Flint. 2017. 'Formation of propionate and butyrate by the human colonic microbiota', *Environmental microbiology*, 19: 29-41.

Lovley, Derek R. 2017. 'Happy together: microbial communities that hook up to swap electrons', *The ISME journal*, 11: 327.

Luan, Hemi, Xian Wang, and Zongwei Cai. 2019. 'Mass spectrometry-based metabolomics: Targeting the crosstalk between gut microbiota and brain in neurodegenerative disorders', *Mass Spectrometry Reviews*, 38: 22-33.

Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. 'Fast automated reconstruction of genome-scale metabolic models for microbial species and communities', *Nucleic acids research*.

Machado, Daniel, and Markus Herrgård. 2014. 'Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism', *PLoS computational biology*, 10: e1003580.

Magnúsdóttir, Stefanía, Almut Heinken, Laura Kutt, Dmitry A Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, and Paul Wilmes. 2017.

'Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota', *Nature Biotechnology*, 35: 81.

Magnúsdóttir, Stefanía, Dmitry Ravcheev, Valérie de Crécy-Lagard, and Ines Thiele. 2015. 'Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes', *Frontiers in genetics*, 6: 148.

Mahadevan, R, and CH Schilling. 2003. 'The effects of alternate optimal solutions in constraint-based genome-scale metabolic models', *Metabolic engineering*, 5: 264-76.

Mahadevan, Radhakrishnan, Jeremy S Edwards, and Francis J Doyle III. 2002. 'Dynamic flux balance analysis of diauxic growth in Escherichia coli', *Biophysical journal*, 83: 1331-40.

Mahowald, Michael A, Federico E Rey, Henning Seedorf, Peter J Turnbaugh, Robert S Fulton, Aye Wollam, Neha Shah, Chunyan Wang, Vincent Magrini, and Richard K Wilson. 2009. 'Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla', *Proceedings of the National Academy of Sciences*, 106: 5859-64.

Mao, Junwen, Andrew E Blanchard, and Ting Lu. 2014. 'Slow and steady wins the race: a bacterial exploitative competition strategy in fluctuating environments', *ACS Synthetic Biology*, 4: 240-48.

Mariat, Denis, O Firmesse, F Levenez, VD Guimarães, H Sokol, J Doré, G Corthier, and JP Furet. 2009. 'The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age', *BMC microbiology*, 9: 123.

Martínez, Inés, Catherine E Muller, and Jens Walter. 2013. 'Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species', *PloS one*, 8: e69621.

Matteuzzi, D, F Crociani, and O Emaldi. 1978. "Amino acids produced by bifidobacteria and some Clostridia." In *Annales de microbiologie*, 175-81.

McNulty, Nathan P, Tanya Yatsunenko, Ansel Hsiao, Jeremiah J Faith, Brian D Muegge, Andrew L Goodman, Bernard Henrissat, Raish Oozeer, Stéphanie Cools-Portier, and Guillaume Gobert. 2011. 'The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins', *Science translational medicine*, 3: 106ra06-06ra06.

Menni, Cristina, Matthew A Jackson, Tess Pallister, Claire J Steves, Tim D Spector, and Ana M Valdes. 2017. 'Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain', *International Journal of Obesity*, 41: 1099.

Metchnikoff, Ilya Ilyich. 2004. *The prolongation of life: optimistic studies* (Springer Publishing Company).

Metges, Cornelia C. 2000. 'Contribution of microbial amino acids to amino acid homeostasis of the host', *The Journal of Nutrition*, 130: 1857S-64S.

Million, Matthieu, Aldiouma Diallo, and Didier Raoult. 2017. 'Gut microbiota and malnutrition', *Microbial Pathogenesis*, 106: 127-38.

Moens, Frédéric, Stefan Weckx, and Luc De Vuyst. 2016. 'Bifidobacterial inulin-type fructan degradation capacity determines cross-feeding interactions between bifidobacteria and Faecalibacterium prausnitzii', *International Journal of Food Microbiology*, 231: 76-85.

Moran, Mary Ann. 2009. 'Metatranscriptomics: Eavesdropping on Complex Microbial Communities-Large-scale sequencing of mRNAs retrieved from natural communities

provides insights into microbial activities and how they are regulated', *Microbe*, 4: 329.

Nagarajan, Harish, Mallory Embree, Amelia-Elena Rotaru, Pravin M Shrestha, Adam M Feist, Bernhard Ø Palsson, Derek R Lovley, and Karsten Zengler. 2013. 'Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association', *Nature communications*, 4: ncomms3809.

Neidhardt, Frederick Carl, John L Ingraham, and Moselio Schaechter. 1990. *Physiology of the bacterial cell: a molecular approach* (Sinauer Associates Sunderland, MA).

Neis, Evelien PJG, Cornelis HC Dejong, and Sander S Rensen. 2015. 'The role of microbial amino acid metabolism in host metabolism', *Nutrients*, 7: 2930-46.

Nicholson, Jeremy K, Elaine Holmes, James Kinross, Remy Burcelin, Glenn Gibson, Wei Jia, and Sven Pettersson. 2012. 'Host-gut microbiota metabolic interactions', *science*, 336: 1262-67.

Norsigian, Charles J, Erol Kavvas, Yara Seif, Bernhard O Palsson, and Jonathan M Monk. 2018. 'iCN718, an Updated and Improved Genome-Scale Metabolic Network Reconstruction of Acinetobacter baumannii AYE', *Frontiers in genetics*, 9: 121.

O'Brien, Edward J, Jonathan M Monk, and Bernhard O Palsson. 2015. 'Using genome-scale models to predict biological capabilities', *Cell*, 161: 971-87.

Oberhardt, Matthew A, Jacek Puchałka, Vítor AP Martins Dos Santos, and Jason A Papin. 2011. 'Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis', *PLoS computational biology*, 7: e1001116.

Oberhardt, Matthew A, Jacek Puchałka, Kimberly E Fryer, Vítor AP Martins Dos Santos, and Jason A Papin. 2008. 'Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1', *Journal of bacteriology*, 190: 2790-803.

Occhipinti, R, MA Puchowicz, JC LaManna, E Somersalo, and D Calvetti. 2007. 'Statistical analysis of metabolic pathways of brain metabolism at steady state', *Annals of biomedical engineering*, 35: 886-902.

Ogata, Hiroyuki, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. 1999. 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic acids research*, 27: 29-34.

Olsen, Gary J, David J Lane, Stephen J Giovannoni, Norman R Pace, and David A Stahl. 1986. 'Microbial ecology and evolution: a ribosomal RNA approach', *Annual Reviews in Microbiology*, 40: 337-65.

Orth, Jeffrey D, and Bernhard Ø Palsson. 2010. 'Systematizing the generation of missing metabolic knowledge', *Biotechnology and bioengineering*, 107: 403-12.

Orth, Jeffrey D, Ines Thiele, and Bernhard Ø Palsson. 2010. 'What is flux balance analysis?', *Nature Biotechnology*, 28: 245.

Ouwehand, Arthur C, Seppo Salminen, and Erika Isolauri. 2002. 'Probiotics: an overview of beneficial effects.' in, *Lactic Acid Bacteria: Genetics, Metabolism and Applications* (Springer).

Øyås, Ove, and Jörg Stelling. 2018. 'Genome-scale metabolic networks in time and space', *Current Opinion in Systems Biology*, 8: 51-58.

Pan, Shu, and Jennifer L Reed. 2018. 'Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries', *Current opinion in biotechnology*, 51: 103-08.

Pitkänen, Esa, Paula Jouhten, Jian Hou, Muhammad Fahad Syed, Peter Blomberg, Jana Kludas, Merja Oja, Liisa Holm, Merja Penttilä, and Juho Rousu. 2014. 'Comparative genome-

scale reconstruction of gapless metabolic networks for present and ancestral species', *PLoS computational biology*, 10: e1003465.

Pommier, Thomas, Amélie AM Cantarel, Karl Grigulis, Sandra Lavorel, Nicolas Legay, Catherine Baxendale, Richard D Bardgett, Michael Bahn, Franck Poly, and Jean-Christophe Clément. 2018. 'The added value of including key microbial traits to determine nitrogen-related ecosystem services in managed grasslands', *Journal of Applied Ecology*, 55: 49-58.

Pompei, Anna, Lisa Cordisco, Alberto Amaretti, Simona Zanoni, Stefano Raimondi, Diego Matteuzzi, and Maddalena Rossi. 2007. 'Administration of folate-producing bifidobacteria enhances folate status in Wistar rats', *Journal of Nutrition*, 137: 2742-46.

Pouliot, Yannick, and Peter D Karp. 2007. 'A survey of orphan enzyme activities', *BMC bioinformatics*, 8: 244.

Price, Nathan D, Jan Schellenberger, and Bernhard O Palsson. 2004. 'Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies', *Biophysical journal*, 87: 2172-86.

Price, Nathan D, Ines Thiele, and Bernhard Ø Palsson. 2006. 'Candidate states of Helicobacter pylori's genome-scale metabolic network upon application of "loop law" thermodynamic constraints', *Biophysical journal*, 90: 3919-28.

Prigent, Sylvain, Clémence Frioux, Simon M Dittami, Sven Thiele, Abdelhalim Larhlimi, Guillaume Collet, Fabien Gutknecht, Jeanne Got, Damien Eveillard, and Jérémie Bourdon. 2017. 'Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks', *PLoS computational biology*, 13: e1005276.

Prosser, James I, Brendan JM Bohannan, Tom P Curtis, Richard J Ellis, Mary K Firestone, Rob P Freckleton, Jessica L Green, Laura E Green, Ken Killham, and Jack J Lennon. 2007. 'The role of ecological theory in microbial ecology', *Nature Reviews Microbiology*, 5: 384.

Puchałka, Jacek, Matthew A Oberhardt, Miguel Godinho, Agata Bielecka, Daniela Regenhardt, Kenneth N Timmis, Jason A Papin, and Vítor AP Martins dos Santos. 2008. 'Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology', *PLoS computational biology*, 4: e1000210.

Pultz, Nicole J, Usha Stiefel, Suja Subramanyan, Marion S Helfand, and Curtis J Donskey. 2005. 'Mechanisms by which anaerobic microbiota inhibit the establishment in mice of intestinal colonization by vancomycin-resistant Enterococcus', *The Journal of Infectious Diseases*, 191: 949-56.

Raamsdonk, Leonie M, Bas Teusink, David Broadhurst, Nianshu Zhang, Andrew Hayes, Michael C Walsh, Jan A Berden, Kevin M Brindle, Douglas B Kell, and Jem J Rowland. 2001. 'A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations', *Nature Biotechnology*, 19: 45.

Rappé, Michael S, and Stephen J Giovannoni. 2003. 'The uncultured microbial majority', *Annual Reviews in Microbiology*, 57: 369-94.

Reed, Jennifer L, and Bernhard Ø Palsson. 2004. 'Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states', *Genome research*, 14: 1797-805.

Reed, Jennifer L, Trina R Patel, Keri H Chen, Andrew R Joyce, Margaret K Applebee, Christopher D Herring, Olivia T Bui, Eric M Knight, Stephen S Fong, and Bernhard O

Palsson. 2006. 'Systems approach to refining genome annotation', *Proceedings of the National Academy of Sciences*, 103: 17480-84.

Reidy, Shannon P, and Jean-Michel Weber. 2002. 'Accelerated substrate cycling: a new energy-wasting role for leptin in vivo', *AJP Endocrinology and Metabolism*, 282: E312-E17.

Renz-Polster, H, MR David, A Sonia Buist, WM Vollmer, EA O'connor, EA Frazier, and MA Wall. 2005. 'Caesarean section delivery and the risk of allergic disorders in childhood', *Clinical & Experimental Allergy*, 35: 1466-72.

Riesenfeld, Christian S, Patrick D Schloss, and Jo Handelsman. 2004. 'Metagenomics: genomic analysis of microbial communities', *Annual Review of Genetics*, 38: 525-52.

Rodríguez, Juan Miguel, Kiera Murphy, Catherine Stanton, R Paul Ross, Olivia I Kober, Nathalie Juge, Ekaterina Avershina, Knut Rudi, Arjan Narbad, and Maria C Jenmalm. 2015. 'The composition of the gut microbiota throughout life, with an emphasis on early life', *Microbial Ecology in Health and Disease*, 26: 26050.

Rosenbaum, Michael, Rob Knight, and Rudolph L Leibel. 2015. 'The gut microbiota in human energy homeostasis and obesity', *Trends in Endocrinology & Metabolism*, 26: 493-501.

Rowland, Ian, Glenn Gibson, Almut Heinken, Karen Scott, Jonathan Swann, Ines Thiele, and Kieran Tuohy. 2018. 'Gut microbiota functions: metabolism of nutrients and other food components', *European Journal of Nutrition*, 57: 1-24.

Russell, James B. 2007. 'The energy spilling reactions of bacteria and other organisms', *Journal of molecular microbiology and biotechnology*, 13: 1-11.

Said, Hamid M, and Zainab M Mohammed. 2006. 'Intestinal absorption of water-soluble vitamins: an update', *Current Opinion in Gastroenterology*, 22: 140-46.

Savage, Dwayne C. 1977. 'Microbial ecology of the gastrointestinal tract', *Annual Reviews in Microbiology*, 31: 107-33.

Schaechter, Moselio. 2009. *Encyclopedia of microbiology* (Academic Press).

Schluter, Jonas, and Kevin R Foster. 2012. 'The evolution of mutualism in gut microbiota via host epithelial selection', *PLoS biology*, 10: e1001424.

Schrezenmeir, Jürgen, and Michael de Vrese. 2001. 'Probiotics, prebiotics, and synbiotics—approaching a definition', *The American journal of clinical nutrition*, 73: 361s-64s.

Schuetz, Robert, Lars Kuepfer, and Uwe Sauer. 2007. 'Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli', *Molecular systems biology*, 3: 119.

Schwarzer, Martin. 2018. 'Gut microbiota: puppeteer of the host juvenile growth', *Current Opinion in Clinical Nutrition and Metabolic Care*, 21: 179-83.

Sela, David A, and David A Mills. 2014. 'The marriage of nutrigenomics with the microbiome: the case of infant-associated bifidobacteria and milk–', *The American journal of clinical nutrition*, 99: 697S-703S.

Sender, Ron, Shai Fuchs, and Ron Milo. 2016. 'Revised estimates for the number of human and bacteria cells in the body', *PLoS biology*, 14: e1002533.

Shearer, Alexander G, Tomer Altman, and Christine D Rhee. 2014. 'Finding sequences for over 270 orphan enzymes', *PloS one*, 9: e97250.

Shoaie, Saeed, Pouyan Ghaffari, Petia Kovatcheva-Datchary, Adil Mardinoglu, Partho Sen, Estelle Pujos-Guillot, Tomas de Wouters, Catherine Juste, Salwa Rizkalla, and Julien Chilloux. 2015. 'Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome', *Cell Metabolism*, 22: 320-31.

Shoaie, Saeed, Fredrik Karlsson, Adil Mardinoglu, Intawat Nookaew, Sergio Bordel, and Jens Nielsen. 2013. 'Understanding the interactions between bacteria in the human gut through metabolic modeling', *Scientific reports*, 3: 2532.

Shoaie, Saeed, and Jens Nielsen. 2014. 'Elucidating the interactions between the human gut microbiota and its host through metabolic modeling', *Frontiers in genetics*, 5: 86.

Smith, Michelle I, Tanya Yatsunenko, Mark J Manary, Indi Trehan, Rajhab Mkakosya, Jiye Cheng, Andrew L Kau, Stephen S Rich, Patrick Concannon, and Josyf C Mychaleckyj. 2013. 'Gut microbiomes of Malawian twin pairs discordant for kwashiorkor', *science*, 339: 548-54.

Sokol, H, P Seksik, JP Furet, O Firmesse, I Nion-Larmurier, L Beaugerie, J Cosnes, G Corthier, P Marteau, and J Doré. 2009. 'Low counts of Faecalibacterium prausnitzii in colitis microbiota', *Inflammatory bowel diseases*, 15: 1183-89.

Sonnenburg, Erica D, Samuel A Smits, Mikhail Tikhonov, Steven K Higginbottom, Ned S Wingreen, and Justin L Sonnenburg. 2016. 'Diet-induced extinctions in the gut microbiota compound over generations', *Nature*, 529: 212.

Stauffer, Robert C. 1957. 'Haeckel, Darwin, and ecology', *The Quarterly Review of Biology*, 32: 138-44.

Stein, Richard R, Vanni Bucci, Nora C Toussaint, Charlie G Buffie, Gunnar Rätsch, Eric G Pamer, Chris Sander, and João B Xavier. 2013. 'Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota', *PLoS computational biology*, 9: e1003388.

Stolyar, Sergey, Steve Van Dien, Kristina Linnea Hillesland, Nicolas Pinel, Thomas J Lie, John A Leigh, and David A Stahl. 2007. 'Metabolic modeling of a mutualistic microbial community', *Molecular systems biology*, 3: 92.

Strozzi, G Paolo, and Luca Mogna. 2008. 'Quantification of folic acid in human feces after administration of Bifidobacterium probiotic strains', *Journal of Clinical Gastroenterology*, 42: S179-S84.

Subramanian, Sathish, Laura V Blanton, Steven A Frese, Mark Charbonneau, David A Mills, and Jeffrey I Gordon. 2015. 'Cultivating healthy growth and nutrition through the gut microbiota', *Cell*, 161: 36-48.

Subramanian, Sathish, Sayeeda Huq, Tanya Yatsunenko, Rashidul Haque, Mustafa Mahfuz, Mohammed A Alam, Amber Benezra, Joseph DeStefano, Martin F Meier, and Brian D Muegge. 2014. 'Persistent gut microbiota immaturity in malnourished Bangladeshi children', *Nature*, 510: 417.

Suttie, JW. 1995. 'The importance of menaquinones in human nutrition', *Annual Review of Nutrition*, 15: 399-417.

Swainston, Neil, Kieran Smallbone, Pedro Mendes, Douglas B Kell, and Norman W Paton. 2011. 'The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks', *Journal of Integrative Bioinformatics*, 8: 187-203.

Taffs, Reed, John E Aston, Kristen Brileya, Zackary Jay, Christian G Klatt, Shawn McGlynn, Natasha Mallette, Scott Montross, Robin Gerlach, and William P Inskeep. 2009. 'In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study', *BMC systems biology*, 3: 114.

Terzer, Marco, Nathaniel D Maynard, Markus W Covert, and Jörg Stelling. 2009. 'Genome-scale metabolic networks', *Systems Biology and Medicine*, 1: 285-97.

Teusink, Bas, Anne Wiersma, Douwe Molenaar, Christof Francke, Willem M de Vos, Roland J Siezen, and Eddy J Smid. 2006. 'Analysis of growth of Lactobacillus plantarum WCFS1

on a complex medium using a genome-scale metabolic model', *Journal of Biological Chemistry*, 281: 40041-48.

Thiele, Ines, and Bernhard Ø Palsson. 2010. 'A protocol for generating a high-quality genome-scale metabolic reconstruction', *Nature protocols*, 5: 93.

Thiele, Ines, Nathan D Price, Thuy D Vo, and Bernhard Ø Palsson. 2005. 'Candidate metabolic network states in human mitochondria impact of diabetes, ischemia, and diet', *Journal of Biological Chemistry*, 280: 11683-95.

Thiele, Ines, Nikos Vlassis, and Ronan MT Fleming. 2014. 'fastGapFill: efficient gap filling in metabolic networks', *Bioinformatics*, 30: 2529-31.

Thursby, Elizabeth, and Nathalie Juge. 2017. 'Introduction to the human gut microbiota', *Biochemical Journal*, 474: 1823-36.

Tidjani Alou, Maryam, Matthieu Million, Sory I Traore, Donia Mouelhi, Saber Khelaifia, Dipankar Bachar, Aurelia Caputo, Jeremy Delerce, Souleymane Brah, and Daouda Alhousseini. 2017. 'Gut bacteria missing in severe acute malnutrition, can we identify potential probiotics by culturomics?', *Frontiers in microbiology*, 8: 899.

Tilg, Herbert, Timon E Adolph, Romana R Gerner, and Alexander R Moschen. 2018. 'The intestinal microbiota in colorectal cancer', *Cancer Cell*.

Tobalina, Luis, Rafael Bargiela, Jon Pey, Florian-Alexander Herbst, Iván Lores, David Rojo, Coral Barbas, Ana I Peláez, Jesús Sánchez, and Martin von Bergen. 2015. 'Context-specific metabolic network reconstruction of a naphthalene-degrading bacterial community guided by metaproteomic data', *Bioinformatics*, 31: 1771-79.

Torrallardona, D, CI Harris, Eric Milne, and MF Fuller. 1994. 'The contribution of intestinal microflora to amino acid requirements in pigs', *European Association for Animal Production Publication*, 80: 245-48.

Torrallardona, David, C Ian Harris, Marie E Coates, and Malcolm F Fuller. 1996. 'Microbial amino acid synthesis and utilization in rats: incorporation of 15 N from 15 NH 4 Cl into lysine in the tissues of germ-free and conventional rats', *British journal of nutrition*, 76: 689-700.

Tremaroli, Valentina, and Fredrik Bäckhed. 2012. 'Functional interactions between the gut microbiota and host metabolism', *Nature*, 489: 242.

Tringe, Susannah Green, Christian Von Mering, Arthur Kobayashi, Asaf A Salamov, Kevin Chen, Hwai W Chang, Mircea Podar, Jay M Short, Eric J Mathur, and John C Detter. 2005. 'Comparative metagenomics of microbial communities', *science*, 308: 554-57.

Turnbaugh, Peter J, Micah Hamady, Tanya Yatsunenko, Brandi L Cantarel, Alexis Duncan, Ruth E Ley, Mitchell L Sogin, William J Jones, Bruce A Roe, and Jason P Affourtit. 2009. 'A core gut microbiome in obese and lean twins', *Nature*, 457: 480.

Turnbaugh, Peter J, Ruth E Ley, Michael A Mahowald, Vincent Magrini, Elaine R Mardis, and Jeffrey I Gordon. 2006. 'An obesity-associated gut microbiome with increased capacity for energy harvest', *Nature*, 444: 1027.

Tzamali, Eleftheria, Panayiota Poirazi, Ioannis G Tollis, and Martin Reczko. 2011. 'A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities', *BMC systems biology*, 5: 167.

Ubeda, Carles, Ana Djukovic, and Sandrine Isaac. 2017. 'Roles of the intestinal microbiota in pathogen protection', *Clinical & Translational Immunology*, 6: e128.

UniProt Consortium. 2018. 'UniProt: a worldwide hub of protein knowledge', *Nucleic acids research*, 47: D506-D15.

Vanwonterghem, Inka, Paul D Jensen, Paul G Dennis, Philip Hugenholtz, Korneel Rabaey, and Gene W Tyson. 2014. 'Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters', *The ISME journal*, 8: 2015.

Varma, Amit, and Bernhard O Palsson. 1993. 'Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors', *Journal of theoretical biology*, 165: 477-502.

Venter, J Craig, Karin Remington, John F Heidelberg, Aaron L Halpern, Doug Rusch, Jonathan A Eisen, Dongying Wu, Ian Paulsen, Karen E Nelson, and William Nelson. 2004. 'Environmental genome shotgun sequencing of the Sargasso Sea', *science*, 304: 66-74.

Venturelli, Ophelia S, Alex V Carr, Garth Fisher, Ryan H Hsu, Rebecca Lau, Benjamin P Bowen, Susan Hromada, Trent Northen, and Adam P Arkin. 2018. 'Deciphering microbial interactions in synthetic human gut microbiome communities', *Molecular systems biology*, 14: e8157.

Voit, Eberhard O. 2016. *The inner workings of life: Vignettes in systems biology* (Cambridge University Press).

Wakeham, Stuart G, Rudi Amann, Katherine H Freeman, Ellen C Hopmans, Bo Barker Jørgensen, Isabell F Putnam, Stefan Schouten, Jaap S Sinninghe Damsté, Helen M Talbot, and Dagmar Woebken. 2007. 'Microbial ecology of the stratified water column of the Black Sea as revealed by a comprehensive biomarker study', *Organic Geochemistry*, 38: 2070-97.

Waldor, Matthew K, Gene Tyson, Elhanan Borenstein, Howard Ochman, Andrew Moeller, B Brett Finlay, Heidi H Kong, Jeffrey I Gordon, Karen E Nelson, and Karim Dabbagh. 2015. 'Where next for microbiome research?', *PLoS biology*, 13: e1002050.

Wang, Chongxin, Satoru Nagata, Takashi Asahara, Norikatsu Yuki, Kazunori Matsuda, Hirokazu Tsuji, Takuya Takahashi, Koji Nomoto, and Yuichiro Yamashiro. 2015. 'Intestinal microbiota profiles of healthy pre-school and school-age children and effects of probiotic supplementation', *Annals of Nutrition and Metabolism*, 67: 257-66.

Wang, Leli, Vinothkannan Ravichandran, Yulong Yin, Jia Yin, and Youming Zhang. 2018. 'Natural Products from Mammalian Gut Microbiota', *Trends in biotechnology*.

Wang, Xiaohui, Xianghua Wen, Hengjing Yan, Kun Ding, Fang Zhao, and Man Hu. 2011. 'Bacterial community dynamics in a functionally stable pilot-scale wastewater treatment plant', *Bioresource Technology*, 102: 2352-57.

Wangersky, Peter J. 1978. 'Lotka-Volterra population models', *Annual Review of Ecology, Evolution, and Systematics*, 9: 189-218.

Wiback, Sharon J, Iman Famili, Harvey J Greenberg, and Bernhard Ø Palsson. 2004. 'Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space', *Journal of theoretical biology*, 228: 437-47.

Wilmes, Paul, and Philip L Bond. 2006. 'Metaproteomics: studying functional gene expression in microbial ecosystems', *Trends in Microbiology*, 14: 92-97.

Wittebolle, Lieven, Han Vervaeren, Willy Verstraete, and Nico Boon. 2008. 'Quantifying community dynamics of nitrifiers in functionally stable reactors', *Applied & Environmental Microbiology*, 74: 286-93.

Woese, Carl R, and George E Fox. 1977. 'Phylogenetic structure of the prokaryotic domain: the primary kingdoms', *Proceedings of the National Academy of Sciences*, 74: 5088-90.

World Health Organization. 2016. 'What is malnutrition?'. https://www.who.int/features/qa/malnutrition/en/.

Xu, Jian, Michael A Mahowald, Ruth E Ley, Catherine A Lozupone, Micah Hamady, Eric C Martens, Bernard Henrissat, Pedro M Coutinho, Patrick Minx, and Philippe Latreille. 2007. 'Evolution of symbiotic bacteria in the distal human intestine', *PLoS biology*, 5: e156.

Yatsunenko, Tanya, Federico E Rey, Mark J Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N Baldassano, and Andrey P Anokhin. 2012. 'Human gut microbiome viewed across age and geography', *Nature*, 486: 222.

Ye, Chao, Wei Zou, Nan Xu, and Liming Liu. 2014. 'Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production', *Journal of Biotechnology*, 182: 61-67.

Young, Iain M, John W Crawford, Naoise Nunan, Wilfred Otten, and Andrew Spiers. 2008. 'Microbial distribution in soils: physics and scaling', *Advances in Agronomy*, 100: 81-121.

Youssef, Noha H, MB Couger, Alexandra L McCully, Andres Eduardo Guerrero Criado, and Mostafa S Elshahed. 2015. 'Assessing the global phylum level diversity within the bacterial domain: a review', *Journal of Advanced Research*, 6: 269-82.

Ze, Xiaolei, Sylvia H Duncan, Petra Louis, and Harry J Flint. 2012. 'Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon', *The ISME journal*, 6: 1535.

Zhuang, Kai, Mounir Izallalen, Paula Mouser, Hanno Richter, Carla Risso, Radhakrishnan Mahadevan, and Derek R Lovley. 2011. 'Genome-scale dynamic modeling of the competition between Rhodoferax and Geobacter in anoxic subsurface environments', *The ISME journal*, 5: 305.

Zmora, Niv, Gili Zilberman-Schapira, Jotham Suez, Uria Mor, Mally Dori-Bachash, Stavros Bashiardes, Eran Kotler, Maya Zur, Dana Regev-Lehavi, and Rotem Ben-Zeev Brik. 2018. 'Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features', *Cell*, 174: 1388-405. e21.

Zomorrodi, Ali R, Mohammad Mazharul Islam, and Costas D Maranas. 2014. 'd-OptCom: dynamic multi-level and multi-objective metabolic modeling of microbial communities', *ACS Synthetic Biology*, 3: 247-57.

Zomorrodi, Ali R, and Costas D Maranas. 2012. 'OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities', *PLoS computational biology*, 8: e1002363.

Zomorrodi, Ali R, and Daniel Segre. 2016. 'Synthetic ecology of microbes: mathematical models and applications', *Journal of molecular biology*, 428: 837-61.